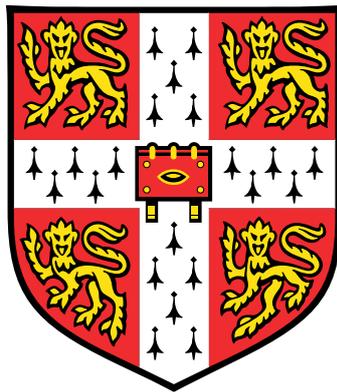


GRAVITATIONAL LENSING IN THE SOLAR NEIGHBOURHOOD
AND TOWARDS THE MILKY WAY BULGE

PETER MCGILL



Darwin College
University of Cambridge

August 2021

This dissertation is submitted for the degree of Doctor of Philosophy

Peter McGill: *Gravitational lensing in the Solar Neighbourhood and towards the Milky Way bulge*, , © August 2021

SUMMARY

This thesis is concerned with looking for and characterizing microlensing events in two places that are relatively unexplored. Leveraging astrometry from the Gaia satellite, I search for predicted close stellar alignments by lenses in the solar neighbourhood that will give rise to a microlensing event. Using Near-Infrared (NIR) photometry from the Vista Variables in the Via Lactea (VVV) survey, I extract microlensing events towards highly-extinct regions of the Galactic bulge. In both cases, I develop a Bayesian methodology to characterize the microlensing signals. In Chapter 1, I review the history of finding microlensing events both by predicting stellar alignments and by monitoring millions of stars. I describe two uses for these types of events; lens mass determination and probing structure of the Galactic bulge. In Chapter 2, I detail the microlensing signals and the methods which underpin the results presented in the rest of the thesis. In Chapter 3, I find a predicted microlensing event where the lens is a nearby white dwarf. Analysis of this event permitted a direct mass determination of the white dwarf which in turn allowed a test of the white dwarf mass-radius relationship. In Chapter 4, I present a search for predicted photometric microlensing events. For these events, I investigate combining prior astrometric information from Gaia with photometric follow-up data to extract the lens mass. In Chapter 5, I extend predicted microlensing searches using Gaia in combination with astrometry from the VVV. In Chapter 6, I critically examine the reliability of predicted microlensing events found with Gaia. I find that the majority of high-quality events expected to occur over Gaia's life time are in fact spurious. Finally, in Chapter 7, I use machine learning to extract 1959 microlensing events from the VVV and I develop a Bayesian methodology to characterize their sparsely sampled signals.

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed 60,000 word limit.

The majority of the work in this thesis has been published in the following papers and was completed in collaboration with the corresponding co-authors:

- Chapter 3 is partly based on: McGill, P., Smith, L. C., Evans, N. W., Belokurov, V., Smart, R. L. 2018. A predicted astrometric microlensing event by a nearby white dwarf. *MNRAS*, 478, L29–L33.
- Chapter 4 is based on: McGill, P., Smith, L. C., Evans, N. W., Belokurov, V., Zhang, Z. H. 2019. Microlens mass determination for Gaia’s predicted photometric events. *MNRAS*, 483, 4210–4220.
- Chapter 5 is based on: McGill, P., Smith, L. C., Evans, N. W., Belokurov, V., Lucas, P. W. 2019. Ongoing astrometric microlensing events from VVV and Gaia. *MNRAS*, 487, L7–L12.
- Chapter 6 is based on: McGill, P., Everall, A., Boubert, D., Smith, L. C. 2020. Predictions of Gaia’s prize microlensing events are flawed. *MNRAS*, 498, L6–L10.
- Chapter 7 is based on: Husseiniova, A., McGill, P., Smith, L.C., Evans, N.W. 2021. A microlensing search of 700 million VVV light curves. *MNRAS*, 506, 2482–2502.

Cambridge, August 2021

Peter McGill

ACKNOWLEDGEMENTS

Completing a Ph.D. in the midst of the COVID-19 pandemic has been difficult, and I could not have done it without the support of so many kind people. First and foremost I would like to thank my supervisor Wyn Evans. Wyn gave me complete freedom to work on what I found interesting over the past four years, whilst masterfully keeping me on fruitful paths. Wyn provided an exciting, stimulating, and productive research environment and one which I will miss.

I would like to thank all members of the Cambridge Streams group, for a reliable stream of tricky questions that always improved my work. I thank Vasily Belokurov for a critical eye, always having interesting project ideas, and two memorable research trips to Santiago and New York. I'm grateful to Leigh Smith for the constant support, invaluable help, endless patience, and his exquisite reduction of the VVV data on which some of the work in the thesis is based. I am indebted to Sergey Koposov for maintenance and access to the Whole Sky Database. I'm grateful to Andrew Everall and Douglas Boubert for a fun and fast collaboration. I would like to thank Andrea Husseiniova for great work on a productive summer internship.

I would like to thank my collaborators based outside of Cambridge. I would like to thank Ricky Smart, Kailash Sahu, Stefano Casertano, Jay Anderson, and Pierre Bergeron for invaluable help with Chapter 3. I would like to thank Daniel Bramich for organising the followup for the events in Chapter 4.

I would like to thank Nick Massey, Abi Steadman, and my Mum for proof reading parts of this thesis at incredibly short notice. I am grateful to Luis Welbanks and Jake Taylor for a critical read of Chapter 2. I would like to thank my uncle, Andrew Colman, for meticulously reading every single word in this thesis, and suggesting valuable alterations to the text. I could not have finished this thesis without your help.

I would like to thank my mentors Alan Saul, James Hensman, and Sarah Jarvis during my industrial placement. I thoroughly enjoyed my time at Secondmind and I learned so much. I would also like my undergraduate summer internship mentors Chris Lintott and Thomas Greve whose great mentorship confirmed I wanted to pursue a Ph.D. in astronomy. I would also like to thank my physics and

maths teachers, Martin Welch, James Pearson, and Clair Collins. You gave me the push-start I needed on this journey, and were always enthusiastic to answer my tricky questions. For that, I am ever grateful.

Where would I be without my friends? You all mean the world to me. In Cambridge: Luis Welbanks, James Grady, GyuChul Myeong, Chris Desira, and Fruzsina Agocs. And beyond: Jake Taylor, Nick Massey, Sam Nursey, Alex Rossi, Harry Read, and Karen Luong. I could not have managed my final year without the support of Nick in Cambridge. I would like to thank Abigail Williamson for constant support, kindness, and for keeping me sane during the writing process.

I would like to thank my Grandad for igniting my scientific curiosity, I wish you were here to see this thesis. Finally, I would like to thank my parents. Their endless support, encouragement, and sacrifice are the only reasons I've made it this far. Thank you.

SCIENTIFIC ACKNOWLEDGEMENTS

I would like to acknowledge the following financial, technical, and database support that made the work in this thesis possible.

I acknowledge studentship funding support from the United Kingdom Science and Technology Facilities Council (STFC) and the Cambridge Centre for Doctoral Training in Data Intensive Science (CDT-DIS).

This work presents results from the European Space Agency (ESA) space mission Gaia. Gaia data are being processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the Gaia Multi-Lateral Agreement (MLA). The Gaia mission website is <https://www.cosmos.esa.int/gaia>. The Gaia Archive website is <http://archives.esac.esa.int/gaia>.

This work has made use of the VISTA Variables in the Via Lactea Survey (VVV) survey. The survey is based on data products from observations made with ESO Telescopes at the La Silla or Paranal Observatories under ESO programme ID 179.B-2002.

This work used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the DOE and NSF (USA), MISE (Spain), STFC (UK), HEFCE (UK), NCSA (UIUC), KICP (U. Chicago), CCAPP (Ohio State), MIFPA (Texas A&M), CNPQ, FAPERJ, FINEP (Brazil), MINECO (Spain), DFG (Germany) and the collaborating institutions in the Dark Energy Survey, which are Argonne Lab, UC Santa Cruz, University of Cambridge, CIEMAT-Madrid, University of Chicago, University College London, DES-Brazil Consortium, University of Edinburgh, ETH Zürich, Fermilab, University of Illinois, ICE (IEEC-CSIC), IFAE Barcelona, Lawrence Berkeley Lab, LMU München and the associated Excellence Cluster Universe, University of Michigan, NOAO, University of Nottingham, Ohio State University, OzDES Membership Consortium, University of Pennsylvania, University of Portsmouth, SLAC

National Lab, Stanford University, University of Sussex, and Texas A&M University.

This work has made use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

The work makes use of the Digitized Sky Survey (DSS). The images are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain. The Palomar Observatory Sky Survey was funded by the National Geographic Society. The Oschin Schmidt Telescope is operated by the California Institute of Technology and Palomar Observatory. The plates were processed into the present compressed digital form with their permission. The Digitized Sky Survey was produced at the Space Telescope Science Institute (ST ScI) under U. S. Government grant NAG W-2166.

This work has made use of VOSA, developed under the Spanish Virtual Observatory project supported by the Spanish MINECO through grant AyA2017-84089.

This work has made use of the following python packages: Matplotlib (Hunter, 2007), Numpy Harris et al. (2020), PyMC3 Salvatier et al. (2016), emcee Foreman-Mackey et al. (2013), dynesty Speagle (2020), and Astropy (Astropy Collaboration et al., 2013, 2018).

This work has made use of the Whole Sky Database (WSDB) created by Sergey Koposov and maintained at the Institute of Astronomy, University of Cambridge by Sergey Koposov, Vasily Belokurov and Wyn Evans with financial support from the Science and Technology Facilities Council (STFC) and the European Research Council (ERC) (Koposov & Bartunov, 2006).

CONTENTS

1	INTRODUCTION	1
1.1	Galactic microlensing	3
1.1.1	What are the chances?	5
1.1.2	The prediction channel	8
1.1.3	The monitoring channel	11
1.2	The utility of microlensing events	15
1.2.1	Lens mass determination	15
1.2.2	Structure towards the Galactic bulge	20
1.3	Data	24
1.3.1	Gaia	24
1.3.2	The VVV survey	27
1.4	Thesis outline	28
2	MICROLENSING MATHEMATICS AND METHODS	31
2.1	Image positions and amplifications	31
2.2	Lens-source relative motion	33
2.3	The lensing signals	35
2.3.1	The unresolved regime	35
2.3.2	The partially-resolved regime	40
2.3.3	Following an event through both regimes	42
2.3.4	What can we infer from the lensing signals?	43
2.4	Predicting microlensing events	45
2.4.1	Future projections of celestial positions	45
2.4.2	Finding the closest approach	47
2.4.3	Estimating microlensing signals	47
2.4.4	Finding candidate lens-source pairs	48
2.5	Bayesian statistics	49
3	A PREDICTED ASTROMETRIC LENSING EVENT BY A NEARBY WHITE DWARF	53
3.1	A predictive microlensing search of GDR1	53
3.1.1	Candidate event search	53
3.1.2	The candidate	54
3.1.3	Observational outlook with Gaia	56
3.1.4	Observational outlook with HST	59
3.1.5	Conclusion	60
3.2	Analysis of the LAWD 37 astrometric event	60
3.2.1	Outline	61

3.2.2	Data	61
3.2.3	Models	66
3.2.4	Model comparison and criticism	74
3.2.5	Results	78
3.2.6	Discussion and conclusions	88
3.3	The astrophysics of LAWD 37	89
4	MICROLENS MASS DETERMINATION FOR GAIA'S PREDICTED EVENTS	93
4.1	Introduction	93
4.2	A model of the photometric microlensing signal	95
4.3	Candidate search	97
4.4	Candidate 1	98
4.5	Candidate 2	104
4.6	Observational outlook	105
4.6.1	Astrometric signal	105
4.6.2	Photometric signal	107
4.7	Simulated mass determination	108
4.7.1	Candidate 1 mass inference	109
4.7.2	Candidate 2 mass inference	111
4.7.3	Caveats	112
4.8	Discussion and conclusions	114
4.9	Post-conclusion note	117
5	A PREDICTIVE MICROLENSING SEARCH BEYOND GAIA	119
5.1	Introduction	119
5.2	Search for predicted events	120
5.2.1	Selection and lens mass estimates	120
5.2.2	Search method	121
5.3	Candidate events	122
5.3.1	L 338-152	122
5.3.2	NLTT 45128	123
5.3.3	Observational outlook	123
5.4	Conclusions	127
5.5	Post-conclusion note	128
6	IS THAT REALLY A PREDICTED MICROLENSING EVENT?	129
6.1	Introduction	129
6.2	The predicted events	130
6.3	The intriguing case of G123-61	130
6.4	Diagnostics for microlensing candidates	133
6.5	Conclusions	140

7	A MICROLENSING SEARCH OF 700 MILLION VVV LIGHT CURVES	143
7.1	Introduction	144
7.2	Data	146
7.2.1	Simulated light curves	146
7.3	Search algorithm	149
7.3.1	Maximum likelihood estimates	150
7.3.2	Model comparison	152
7.3.3	Model checking	152
7.3.4	The von Neumann ratio	155
7.3.5	Decision tree classifier	157
7.3.6	Validation	163
7.4	Results and completeness analysis	164
7.4.1	Results	164
7.4.2	Visual inspection	167
7.4.3	Spatial variation	168
7.5	Comparison with previous VVV microlensing searches	172
7.5.1	Known event recovery	172
7.5.2	Bayesian inference of PSPL parameters	176
7.6	Inferences on the Einstein timescale	180
7.7	Conclusions	183
8	CONCLUSIONS	189
8.1	Contributions of this thesis	189
8.2	Conclusions and future work	192
8.2.1	Follow-up of predicted microlensing events can be optimized	192
8.2.2	Gaia is incomplete for predictable microlensing events	193
8.2.3	A Bayesian analysis is the natural choice for predicted microlensing events and sparsely sampled photometric events	194
8.2.4	Historical imaging data is a vital diagnostic for predicted microlensing events	195
8.2.5	Future work for the photometric VVV events	195
8.3	Final remarks	196
A	APPENDIX: LAWD 37 ADDITIONAL MATERIAL	197
A.1	Leave-One-Out cross-validation approximation	197
A.2	Posterior approximations	201
A.2.1	The Laplace approximation	201
A.2.2	Monte Carlo MAP estimate for correlated noise	202

A.3	Consistency with Gaia	205
-----	-----------------------	-----

LIST OF FIGURES

Figure 1	The 1919 solar eclipse	2
Figure 2	Microlensing event sketch	5
Figure 3	Astrometric versus photometric optical depth	8
Figure 4	First predicted microlensing event	10
Figure 5	Photometric microlensing towards the bulge	13
Figure 6	White dwarf mass-radius relationships	20
Figure 7	Spatial t_E distribution	22
Figure 8	Geometry of the lens equation	32
Figure 9	Microlensing image distortion	34
Figure 10	Microlensing event diagram	36
Figure 11	The photometric microlensing signal	37
Figure 12	The astrometric microlensing signal	39
Figure 13	Lensing signals through the resolution regimes	44
Figure 14	Lens and source matching strategies	49
Figure 15	LAWD 37 Stellar Field	55
Figure 16	LAWD 37 astrometric signal	57
Figure 17	LAWD 37 on the Gaia focal plane	59
Figure 18	LAWD 37 HST image cutouts	62
Figure 19	LAWD 37 correlated noise realization	64
Figure 20	HST follow-up data of the LAWD 37 event	67
Figure 21	LAWD 37 probabilistic graphical model	73
Figure 22	LAWD 37 posterior realizations over data	79
Figure 23	LAWD 37 model comparison	80
Figure 24	Epoch model comparison	81
Figure 25	Prior versus posterior on noise parameters	82
Figure 26	LAWD 37 source astrometry posterior	84
Figure 27	LAWD 37 angular Einstein radius posterior	85
Figure 28	Prior sensitivity	86
Figure 29	Mass-radius relationship for LAWD 37	91
Figure 30	LAWD 37 Hertzsprung-Russell diagram	92
Figure 31	Stellar field of photometric candidates	98
Figure 32	Known subdwarf spectrum	101
Figure 33	Spectrum of candidate 1	103
Figure 34	Posterior distribution of candidate 1	110
Figure 35	Expected constraints on candidate 2	113
Figure 36	Candidate 2 versus observing strategy	115

Figure 37	Stellar field around L 338-152 and NLTT 45128	122
Figure 38	L 338-152 and NLTT 45128 events	124
Figure 39	G 123-61 events	131
Figure 40	Predicted event diagnostics	134
Figure 41	AL precision for predicted events	137
Figure 42	Distribution of peak signal times	139
Figure 43	Simulated light curve examples	148
Figure 44	Maximum likelihood cuts	154
Figure 45	Feature distribution	159
Figure 46	Decision tree tuning	162
Figure 47	Projected decision surface	165
Figure 48	Visual inspection purity	166
Figure 49	VVV spatial observation distribution	169
Figure 50	Detection efficiency maps	171
Figure 51	Efficiency versus observation strategy	172
Figure 52	Decision tree diagnostic	175
Figure 53	Einstein timescale inference comparison	179
Figure 54	Complex t_E posteriors	181
Figure 55	Einstein timescale posterior summaries	182
Figure 56	Example microlensing events	187
Figure 57	PSIS Pareto k values	200
Figure 58	PSIS approximation accuracy	201
Figure 59	Posterior approximation comparisons	203
Figure 60	Gaia source prior versus posterior	205
Figure 61	Gaia lens prior versus posterior	206

LIST OF TABLES

Table 1	LAWD 37 astrometric data	56
Table 2	PSF correlated noise estimates	65
Table 3	LAWD 37 model summaries	71
Table 4	LAWD 37 parameter priors	72
Table 5	LAWD 37 posterior summaries	78
Table 6	Gaia priors for predicted photometric events	97
Table 7	Event data for two photometric events	100
Table 8	VVV astrometric microlensing events	126
Table 9	Predicted microlensing event diagnostics	141

Table 10	VVV quality cuts	156
Table 11	Decision tree confusion matrix	164
Table 12	VVV Priors	177
Table 13	VVV photometric microlensing events	188

COMMON ACRONYMS

AC	ACross Gaia's scan direction
AL	ALong Gaia's scan direction
GDR ₁	Gaia Data Release 1
GDR ₂	Gaia Data Release 2
GEDR ₃	Gaia Early Data Release 3
CCD	Charged-Coupled Device
HST	Hubble Space Telescope
LOO	Leave-One-Out
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimate
MRR	Mass-Radius Relationship
OGLE	Optical Gravitational Lensing Experiment
PSPL	Point Source Point Lens
TGAS	Tycho-Gaia Astrometric Solution
VLT	Very Large Telescope
VVV	VISTA Variables in the Via Lactea
VIRAC	VVV Infrared Astrometric Catalogue
WFC ₃	The Wide Field Camera 3

COMMON UNITS AND CONSTANTS

arcsecond	arcsecond, angle, 1 arcsecond $\approx 2.8 \times 10^{-4}$ degrees
mas	milliarcsecond, angle, 1 mas $\approx 2.8 \times 10^{-7}$ degrees
μas	microarcsecond, angle, 1 $\mu\text{as} \approx 2.8 \times 10^{-10}$ degrees
pc	parsec, distance, 1 pc $\approx 3.086 \times 10^{16}\text{m}$
au	astronomical unit, distance, 1 au $\approx 1.496 \times 10^{11}\text{m}$
K	kelvin, temperature
J	julian year, time, 1 julian year $\approx 3.156 \times 10^7\text{s}$
M_{\odot}	solar mass, mass, 1 $M_{\odot} \approx 1.989 \times 10^{30}\text{kg}$
R_{\odot}	solar radius, distance, 1 $R_{\odot} \approx 6.957 \times 10^8\text{m}$
mmag	millimagnitude, 1 mmag = 10^{-3} mag
G_{N}	gravitational constant, $G_{\text{N}} \approx 6.674 \times 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2}$
c	speed of light, $c \approx 2.998 \times 10^8\text{ms}^{-1}$

COMMON MATHEMATICAL NOTATION

\vec{x}	vector quantity, $x = \vec{x} $ denotes its magnitude
\hat{x}	unit vector \vec{x}
\vec{x}^{T}	transpose of vector \vec{x}
\mathbf{X}	matrix quantity
\mathbf{X}^{T}	transpose of matrix \mathbf{X}
\mathbf{X}^{-1}	inverse of matrix \mathbf{X}
$ \mathbf{X} $	determinant of matrix \mathbf{X}
\mathcal{D}	data set
\mathcal{M}	model
\ln	natural logarithm, \log_e

INTRODUCTION

Gravity lies at the heart of our understanding of the cosmos over a multitude of large distance scales. Gravity governs the ocean tides on Earth, orbits of the planets in the solar system, the formation of stars from the collapse of molecular clouds, the dynamics of stars within galaxies, and even the configuration of galaxy superclusters; the largest known structures in the universe.

The most accurate and successful current theory of gravity is Einstein's General Theory of Relativity (Einstein, 1916). One of the key predictions of general relativity is that light passing close to a massive object is deflected because of the curvature of spacetime. This means that massive objects can act like lenses and focus light from background sources. The predicted angle of deflection from general relativity is,

$$\tilde{\alpha} = \frac{4G_{\text{N}}M_{\text{L}}}{c^2b}. \quad (1)$$

Here b is the impact parameter between the light and the massive object, G_{N} is Newton's gravitational constant, and c is the speed of light. This is twice the value predicted by the classic Newtonian theory of gravity (Soldner, 1804).

Faced with two competing theories of gravity making two different predictions, Frank Dyson and Arthur Eddington sought an observation test. They measured the deflection of background starlight passing close to the Sun. Due to the dazzling brightness of the Sun, they had to wait for a solar eclipse in order to measure the apparent shift in the position of a background star close to the solar limb. The chance came at the 1919 solar eclipse (Fig. 1), and Dyson et al. (1920) measured deflections of 1.98 ± 0.18 and 1.60 ± 0.31 arcseconds. This was in agreement with the prediction of general relativity of 1.75 arcseconds, providing popular confirmation of Einstein's theory of gravity and catapulted him to fame.

Since the 1919 solar eclipse, gravitational lensing has matured into a useful tool for understanding the universe. Many types of astrophysical objects can act as gravitational lenses over different distance scales, from galaxies (Zwicky, 1937), to clusters of galaxies (e.g. Lynds

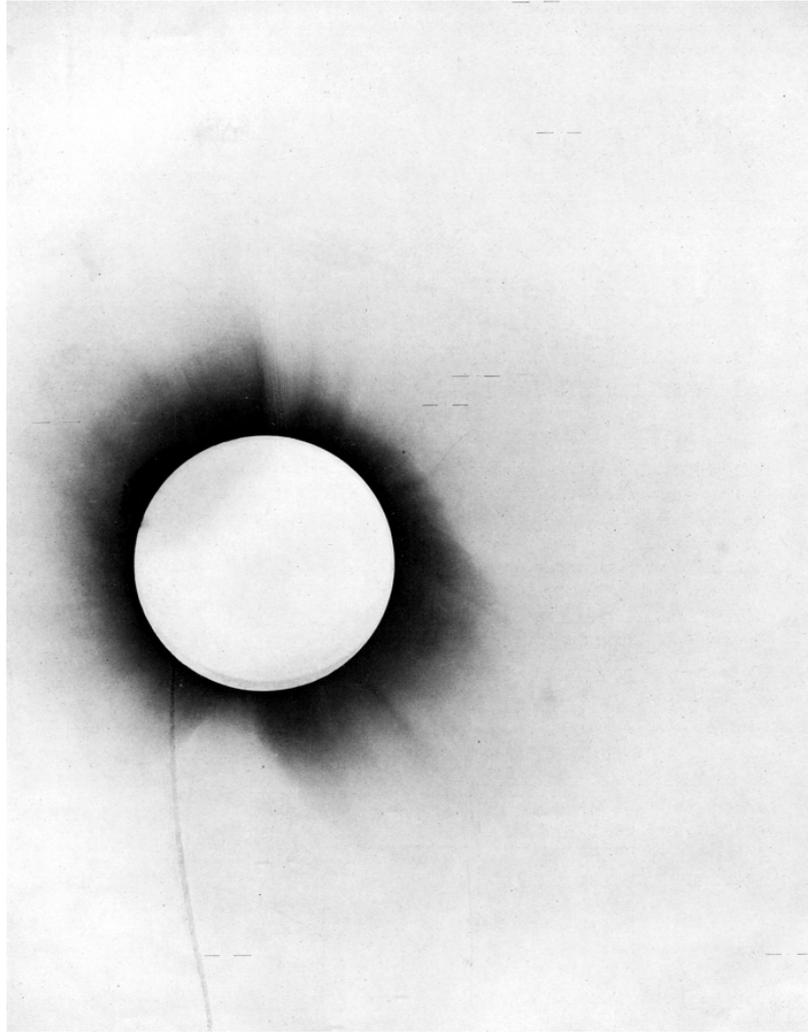


Figure 1: The 1919 solar eclipse. Dyson and Eddington measured the position of stars during the solar eclipse when they were close to the solar limb and their light was being deflected. The deflected positions were then compared to their usual positions, allowing the light deflection caused by the Sun to be measured.

& Petrosian, 1989), to single stars (Chwolson, 1924), and even exoplanets (e.g. Mróz et al., 2020b). Consequently, the lensing signals of these objects are detectable over different angular scales and can be grouped into three main categories; strong lensing, weak lensing and microlensing.

In strong lensing, the lensing object splits the light of a more distant object into multiple resolved images. This effect typically requires galactic mass objects and close alignments in order to produce images with resolvable separations. The first example of strong lensing was discovered by Walsh et al. (1979) where two separate images of a quasar were observed due to it being lensed by an interven-

ing galaxy. These systems have many applications. To name a couple, strongly lensed systems can act as telescopes, magnifying the light of the most distant galaxies in the universe allowing us to study them (e.g. [Marrone et al., 2018](#)). Moreover, they can constrain the dark mass components of galaxies and galaxy clusters (e.g. [Oguri et al., 2014](#)).

Weak lensing occurs at wide alignments where the lensing object subtly alters the shapes and orientations of background galaxies ([Tyson et al., 1990](#); [Kaiser & Squires, 1993](#)). This effect requires galaxy cluster mass lenses and the perturbation of many background galaxies by the same lens to be measured. In addition to probing the structure of particular lensing objects ([Kneib & Natarajan, 2011](#)), these systems can be used to study the large scale structure of the universe between the observer and background source, constraining cosmological parameters (e.g. [Erben et al., 2013](#)).

Microlensing is a version of strong lensing but on small (classically unresolved) angular scales caused by compact stellar mass lenses. In this scenario, typically only a blend of the images is seen.¹ Prompted by the discovery of the strongly lensed system in [Wales & Doye \(1997\)](#), microlensing of one of the resolved images by stars in the intervening lensing galaxy was suggested (e.g. [Chang & Refsdal, 1979](#); [Kayser et al., 1986](#); [Paczynski, 1986a](#)) and then eventually detected in a different system ([Irwin et al., 1989](#); [Vanderriest et al., 1989](#)). This thesis is concerned with Galactic microlensing where the background source is a star and the lens is also a star (or stellar remnant) located in our home galaxy, the Milky Way.

1.1 GALACTIC MICROLENSING

Consider a foreground star (the lens, with mass M_L and at distance D_L from an observer) lensing the light of a more distant background star (the source, at distance D_S from an observer). The first to consider this scenario, with a lens star other than the Sun, were [Chwolson \(1924\)](#) and [Einstein \(1936\)](#). In the case of a perfect lens-source alignment, gravitational lensing causes an image of the source in a

¹ A notable exception to this was presented by [Dong et al. \(2019\)](#), using interferometry on the Very Large Telescope two images separated by ≈ 3 mas were detected during a lensing event between two stars.

circular Einstein ring around the lens to be formed with angular radius (Einstein, 1936),

$$\Theta_E = \sqrt{\frac{4G_N M_L (D_S - D_L)}{c^2 D_S D_L}}. \quad (2)$$

In the case of imperfect lens-source alignment, two images of the source are formed. As the lens passes between the observer and the source, the images change position and brightness which in principle gives rise to two effects. Despite these two effects being manifestations of the same underlying physical phenomenon, the characteristics of their signals are different. Fig. 2 shows this lensing scenario and the two types of effects.

The first effect is photometric microlensing (e.g. Paczynski, 1986b). As the lens passes in front of the source, a symmetrical transient brightening takes place (Fig.2 bottom left panel). This is due to the gravitational field of the lens deflecting light from the source towards the observer that wouldn't usually be seen. This effect typically requires close lens-source alignments $\approx \Theta_E$. The size of the brightening (or amplification) depends primarily on the ratio of the lens-source angular separation and Θ_E , or the normalized lens-source separation. For arbitrarily small normalized lens-source separations (approaching perfect lens-source alignments), the size of the amplification can become extremely (almost arbitrarily) large.

The second effect is astrometric microlensing (e.g. Walker, 1995; Hog et al., 1995; Miyamoto & Yoshii, 1995). As the lens passes in front of the source, there is an apparent elliptical excursion of the source from its unlensed position (Fig.2 bottom right panel). This is due to the source images changing position and brightness as light is deflected. A snapshot of this effect was measured at the 1919 solar eclipse. This effect is typically detectable at large lens-source alignments $> \Theta_E$ and has a longer duration than the photometric effect. This size of the astrometric effect depends on the normalized lens-source separation, but is scaled by Θ_E . Lensing configurations with larger Θ_E (more massive and close by lenses) have larger astrometric effects at a given normalized lens-source separation. For arbitrarily small normalized lens-source separations, the size of the astrometric signal does not increase arbitrarily, and is fundamentally limited by the size of Θ_E .

Both of these effects require the very close chance alignments of two stars $\approx \Theta_E \approx 10^{-8}$ degrees or sub milliarcsecond (mas) scales for

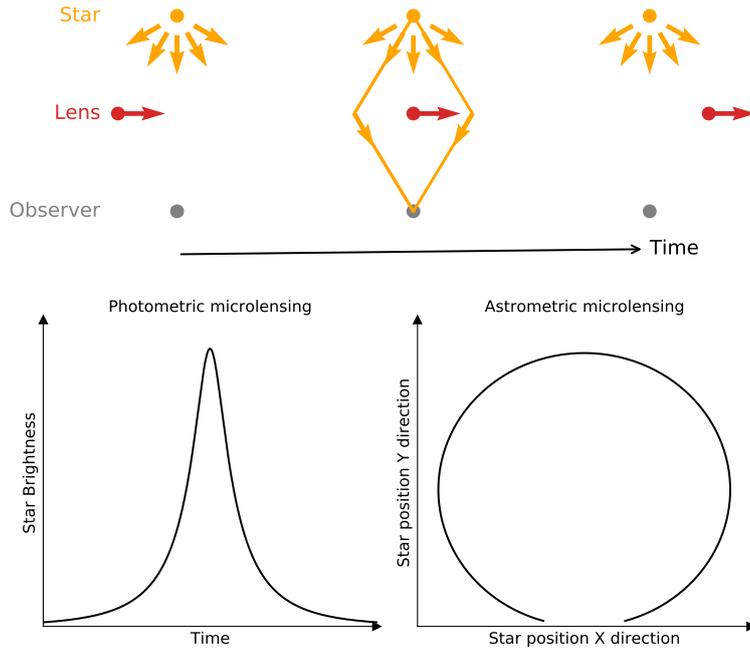


Figure 2: A massive lens intervenes between an observer and a distant background star. The gravitational field of the lens focuses the light of the background star when they are closely aligned. **Top:** Three snapshots of a microlensing event. Red and orange arrows mark the direction of the lens velocity and light rays, respectively. The middle panel shows the close alignment where the gravitational field of the lens deflects light rays from the star. **Bottom:** The two detectable effects. The left panel shows the apparent brightening of the background star due to lensing (photometric microlensing). The right panel shows the apparent change in source position due to lensing (astrometric microlensing). The positive X and Y directions would be to the right and in/out of the plane in the top panel, respectively.

typical stars in the Galaxy. For this reason, [Einstein \(1936\)](#) concluded, for the photometric effect at least, that there was "no great chance" of observing this phenomenon.

1.1.1 *What are the chances?*

The probability of any star being lensed at a given time can be calculated using the idea of optical depth ([Vietri & Ostriker, 1983](#); [Paczynski, 1986b](#)). Consider an observer looking out along a particular line of sight to some patch of sky with angular area, $d\Omega$. Let us consider all possible lenses out to distance D_L . Let A be an area around the lens in the lens plane for which a background source position, projected into the lens plane, would cause a threshold microlensing signal. The

microlensing optical depth is the product of A with the number volume density of lenses integrated over the observed volume of lenses and the distribution of lens masses f_{M_L} . This is expressed as a fraction of the observed area of sky, $d\Omega$. The optical depth gives the probability that a given star at a given time is being lensed with a threshold microlensing signal and is given by (e.g. [Dominik & Sahu, 2000](#)) as,

$$\tau = \frac{1}{d\Omega} \int \int \int_S dV \int_0^\infty dM_L \frac{\rho(D_L)}{M_L D_L^2} f_{M_L}(M_L) A(D_L, D_S, M_L). \quad (3)$$

$\rho(D_L)$ is the mass volume density of the lenses out to distance D_L , the volume element $dV = d\Omega D_L^2 dD_L$, S is the observed spherical volume of radius of D_S and we are integrating over all possible lens masses. Canceling area elements and substituting $x = D_L/D_S$, the optical depth simplifies to an integral along the line of sight (e.g. [Dominik & Sahu, 2000](#)),

$$\tau = D_S \int_0^1 dx \int_0^\infty dM_L \frac{\rho(x)}{M_L} A(x, M_L) f_{M_L}(M_L). \quad (4)$$

For a detectable photometric signal, the source is typically taken to be within Θ_E of the lens so $A = \pi D_L^2 \Theta_E^2$ is a circle, and the optical depth simplifies to ([Paczynski, 1986b](#)),

$$\tau_{\text{phot}} = \frac{4\pi G}{c^2} D_S^2 \int_0^1 dx \rho(x) x(1-x). \quad (5)$$

Notice here that τ_{phot} is independent of the lens masses.

The astrometric signal is detectable at much larger angular separations and falls off more slowly with separation compared with the photometric signal. For reasonable observing timescales $T_{\text{obs}} \approx 10$ years, it's common for the astrometric signals to be detectable throughout. This means that, unlike the photometric signal, the baseline, or the unlensed position of the source, is unlikely to be measured. This suggests that, instead of considering the A that gives rise to a peak signal, it is more sensible to consider the A that gives rise a variation of the astrometric signal within T_{obs} . This area is circular with radius in the lens plane given by ([Dominik & Sahu, 2000](#)) as,

$$b_T = \Theta_E \sqrt{\frac{D_L T_{\text{obs}} v_{\text{rel}}}{\sigma_{\text{ast}}}}. \quad (6)$$

Here v_{rel} is the relative lens-source transverse velocity. If we require an astrometric event to have both a peak signal and variation $> \sigma_{\text{ast}}$

within T_{obs} , then $A = 2b_T T_{\text{obs}} v_{\text{rel}}$ is a rectangle and the astrometric optical depth is (e.g. [Dominik & Sahu, 2000](#)),

$$\tau_{\text{ast}} = 4 \sqrt{\frac{G^2 v_{\text{rel}}^3 T_{\text{obs}}^3}{c^2 \sigma_{\text{ast}}}} D_S \langle M_L^{1/2} \rangle \int_0^1 dx \rho(x) \sqrt{(1-x)}. \quad (7)$$

In contrast to the photometric optical depth, τ_{ast} is dependent on the average square root mass of the population of lenses, $\langle M_L^{1/2} \rangle = \int_0^\infty M_L^{1/2} f_{M_L}(M_L) dM_L$. Most strikingly, however, the difference between the photometric and astrometric signals is the likely distance of the lens (D_L). This is due to the different functional forms of the integrand terms in Eqs. (5) and (7).

For any detailed optical depth estimation, the functional form of the mass density of lenses along a particular line of sight, $\rho(x)$, has to be carefully constructed (e.g. [Belokurov & Evans, 2002](#); [Specht et al., 2020](#)). However, for the purposes of comparing the photometric and astrometric signals, it is informative to assume a constant mass density of lenses, $\rho(x) = \rho_0$. Assuming $p(x) = p_0$, both τ_{ast} and τ_{phot} can be written in terms of the integrals of a probability density function, $f(x)$; $f_{\text{phot}}(x) \propto x(1-x)$, and $f_{\text{ast}}(x) \propto \sqrt{1-x}$.

[Fig. 3](#) shows the two probability density functions. We can see for the photometric optical depth, lenses at intermediate distances to the source (half way to the source $D_L/D_S = 0.5$), are most likely to yield detectable signals. Conversely, for the astrometric optical depth, close-by lenses are favoured. The consequence of this is that the majority of events with detectable astrometric signatures are caused by nearby lenses in the solar neighborhood ($D_L < 100$ pc; [Belokurov & Evans, 2002](#)), regardless of where the source star is located ([Dominik & Sahu, 2000](#)).

In addition to the different lens distance sensitivities, the magnitudes of the photometric and astrometric optical depths also differ. Looking towards the Galactic bulge, the most recent estimates for the photometric optical depth from the Optical Gravitational Lensing Experiment (OGLE; [Udalski et al., 2015](#)) measure τ_{phot} between 0.5×10^{-6} and 2.5×10^{-6} depending on the exact line of sight ([Mróz et al., 2019](#)). τ_{ast} has not been measured but is expected to be higher than τ_{phot} . Using models of the Galaxy, the assumption of $T_{\text{obs}} = 10$ years and σ_{ast} ranging from $\approx 0.03 - 1$ mas (the parameters of the Gaia astrometric space mission; [Gaia Collaboration et al., 2018](#)), [Belokurov & Evans \(2002\)](#) estimated the all-sky averaged $\tau_{\text{ast}} \approx 2.5 \times 10^{-5}$. Overall, with current astrometric and photometric measuring capability and

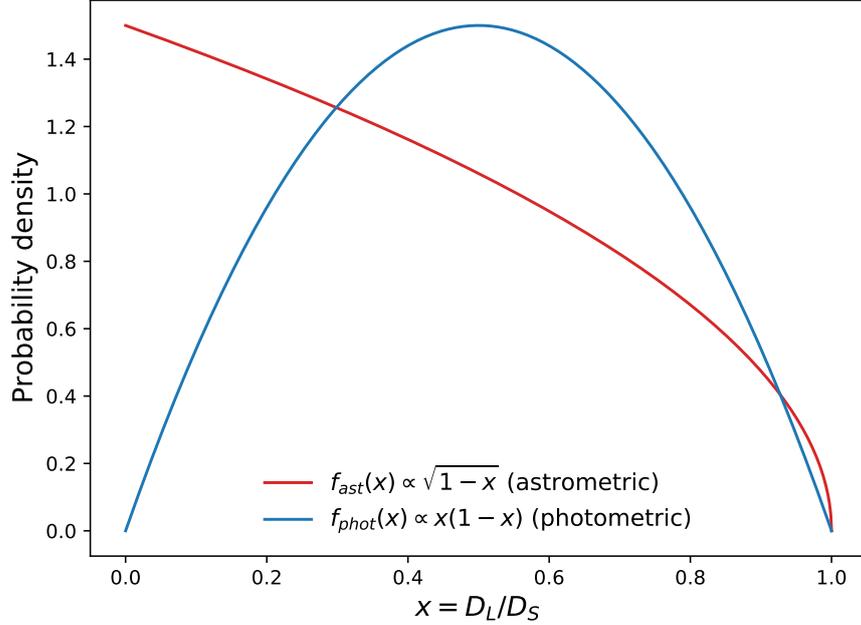


Figure 3: Astrometric versus photometric optical depth. Intergrands of Eqs. (5) and (7) under the assumption of a constant lens mass density. Both functions have been normalized so the area underneath them is unity. Lenses halfway between the observer and source are preferred for the photometric effect whereas close-by lenses are favoured by the astrometric effect. Figure after (Dominik & Sahu, 2000).

given that the Galaxy contains \sim a billion detectable stars, the sizes of τ_{phot} and τ_{ast} suggest that microlensing events with detectable astrometric or photometric signals should be numerous.

Given a number of monitored stars, both τ_{ast} and τ_{phot} translate into expected numbers of events along a given line of sight. That means if significant numbers of events can be detected, then the optical depth can be measured which will contain information about the distribution of lenses along a given line of sight (e.g. Paczynski, 1986b).

1.1.2 The prediction channel

It was first realized that the photometric signals of microlensing events by stellar lenses and sources within the Galaxy should be detectable and occur frequently by both Refsdal (1964) and Liebes (1964). The major hurdle to observing these signals, however, was figuring out when and where to look to catch the chance close alignment of two stars. Whilst Liebes (1964) appeared defeated by the

sheer enormity of figuring which of the \sim billion detectable stars in the Galaxy would be lensed at a given time, Refsdal (1964) came up with a practical suggestion. The final passage of Refsdal (1964) reads:

"It seems safe to conclude that passages observable from the Earth occur rather frequently. The problem is to find where and when the passages take place. By comparing photographs of the sky taken at different times, the angular velocity of a great number of stars can be determined, and passages may be predicted."

In other words, Refsdal suggested that, if the positions and proper motions of stars could be measured, their future trajectories across the sky could be predicted. For a given star, it would then be possible to check if it would align sufficiently with a distant background star in the future to produce a detectable lensing signal. This solved the problem of when and where to look.

Following Refsdal's suggestion, Feibelman (1966) initiated the first search for future close stellar alignments. Feibelman (1966) found that the nearby star, 40 Eridani-A, would align closely with a background source and produce a detectable photometric signal with the predicted event maximum happening ≈ 20 years in the future. Fig. 4 shows 40 Eridani-A's future predicted path and Feibelman's lensing event prediction. Unfortunately, it turned out that upon further refinement of 40 Eridani-A's proper motion, the alignment would not be close enough to produce any detectable signal (Feibelman, 1986). Reeling from his "20-year exercise in frustration", Feibelman encouraged others to conduct more systematic searches for predicted future stellar alignments using larger databases of stars.

The prediction channel lay dormant until Paczynski (1995, 1996b) reignited interest. Motivated by the realization that the photometric signals of these events could be used to measure the mass of nearby stars, Paczynski (1995) highlighted the growing astrometric databases of high-proper motion stars that could be used to predict events. With the same motivations and the prospect of future space-based high-precision astrometry, Miralda-Escude (1996) pointed out that the astrometric signals of predicted events could also be detected and used to determine lens masses. The key insight by Miralda-Escude (1996) was that astrometric signals are detectable at much larger lens-source separations (or looser alignments) compared with photometric signals. This meant that astrometric events were easier to predict and more likely to happen in comparison to events with just photometric signals.

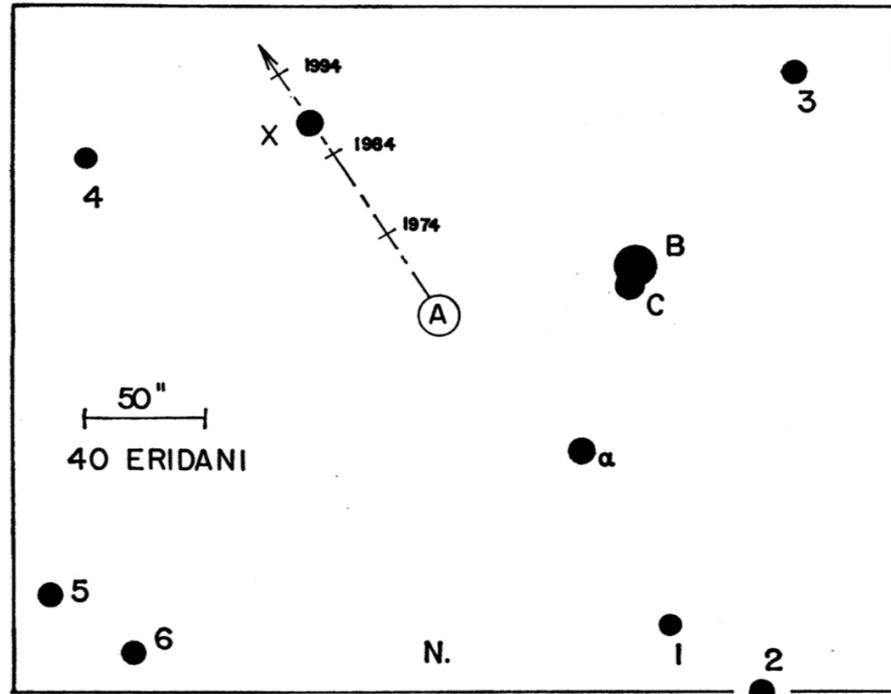


Figure 4: Diagram of the region around 40 Eridani-A. A marks the position of 40 Eridani-A in the year 1966 and the arrow marks its predicted trajectory and proper motion. X marks the position of the background source which 40 Eridani-A was predicted to align with in the year 1986. Circles mark the positions of other stars. The Scale bar shows 50 arcseconds, $\approx 10^{-2}$ degrees. From Feibelman, Walter A. "Gravitational Lens Effect: An Observational Test." *Science* 151.3706 (1966): 73-74. Reprinted with permission from AAAS.

Miralda-Escude (1996)'s insight triggered a detailed analysis on the prospects of detecting predicted astrometric microlensing events with future astrometric space missions (Gould, 2000), and the first systematic searches for predicted astrometric microlensing events, yielded 178 candidates (Sahu et al., 1998; Salim & Gould, 2000). With the advent of deep proper motion catalogues, Gould (2000) warned that the number of spurious candidate events would rise, and highlighted the need for additional effort to sort the "wheat from the chaff".

As astrometric catalogues of stars grew in size and precision (e.g. LSPM-North; Lépine & Shara 2005, PPMXL; Roeser et al. 2010, OGLEB; Sumi et al. 2003), further systematic searches for predicted astrometric events were performed (Proft et al., 2011; Lépine & DiStefano, 2012; Sahu et al., 2014; Proft, 2016) yielding hundreds of event predictions. These searches focused on eliminating spurious high-proper motions stars, which plagued early astrometric catalogues, by visual inspection of imaging data and a reduced proper motion

analysis (e.g. [Proft et al., 2011](#)). For the surviving candidates, the combination of imprecise lens proper motions and no distance information (no parallax) on the lens stars, often hindered these searches. This resulted in the majority of event predictions being uncertain (see e.g. [Proft et al., 2011](#), where all the predicted astrometric signals have $\approx 100\%$ uncertainty). Additionally, [Harding et al. \(2018\)](#) examined the predictable event rates of different types of stellar remnants (white dwarfs, black holes, and neutron stars).

Only two of the predicted astrometric microlensing events had been successfully followed up before the work presented in this thesis. Using the Hubble Space Telescope (HST), [Sahu et al. \(2017\)](#) successfully detected the astrometric signal of a microlensing event caused by the nearby white dwarf Stein 2051B and originally predicted by [Proft et al. \(2011\)](#). The astrometric microlensing signal permitted [Sahu et al. \(2017\)](#) to measure a gravitational mass of Stein 2051 B, of $0.675 \pm 0.051 M_{\odot}$. This marked the first ever detection of the astrometric microlensing effect outside the solar system and provided a stringent test of white dwarf evolutionary theory. Next, [Zurlo et al. \(2018\)](#) detected an astrometric microlensing event using the Very Large Telescope (VLT). This event was caused by our nearest stellar neighbour, Proxima Centauri, and was originally predicted by [Sahu et al. \(2014\)](#). Using the astrometric signal, [Zurlo et al. \(2018\)](#) determined the mass of Proxima Centauri to be $0.150_{-0.051}^{+0.062} M_{\odot}$. This marked the first direct gravitational mass measurement of Proxima Centauri, and provided the only current opportunity for a direct mass determination.

1.1.3 *The monitoring channel*

The sheer enormity and technical challenge of monitoring enough stars to detect gravitational lensing events by chance kept the monitoring channel dormant for two decades after the works of [Refsdal \(1964\)](#) and [Liebes \(1964\)](#). This changed when [Paczynski \(1986b\)](#) suggested monitoring millions of stars in the Large and Small Magellanic Clouds to detect photometric microlensing events with the motivation of shedding light on the nature of dark matter. [Paczynski \(1986b\)](#) estimated that if dark matter was in the form of massive compact objects in the Galactic halo (Massive Compact Halo Objects; MACHOs), the photometric optical depth for such events would be $\approx 10^{-6}$. Luckily, technological advances in observation capabilities soon after [Paczynski \(1986b\)](#)'s suggestions made it a practical

idea (Paczynski, 1996a) and it triggered several photometric monitoring campaigns and a trickle of tens of detected events; e.g. EROS (Aubourg et al., 1993), MACHO (Alcock et al., 1993), OGLE (Udalski et al., 1992). Initial analysis of the events implied that 20% of the mass in the halo consisted of sub-stellar MACHOs with an average mass of $0.4M_{\odot}$ (Alcock et al., 2000a; Bennett, 2005). However, subsequent studies (Tisserand et al., 2007; Wyrzykowski et al., 2011a,b) found much lower sub-stellar MACHO fraction contents and concluded that nearly all events could be explained by normal stars.

Motivated again by detecting MACHOs, Crotts (1992) and Baillon et al. (1993) suggested the idea of pixel lensing where, instead of monitoring single resolved stars, the light of many unresolved stars within an area, or pixel, are monitored. The unresolved stellar population in the Andromeda galaxy was proposed as a target due to its high inclination. This geometry would result in different event rates between the near and far side of Andromeda’s disc only if events were being caused by MACHOs in Andromeda’s halo. This triggered several monitoring campaigns (e.g. Paulin-Henriksson et al., 2002, 2003; Calchi Novati et al., 2005a), resulting in the detection of a handful of candidate events.

Griest et al. (1991) and Paczynski (1991) first proposed monitoring the Milky Way bulge for microlensing events primarily as a reliability check on the results towards the Magellanic clouds (Alard et al., 1995; Evans, 2003). The measured photometric optical depth towards the bulge turned out to be higher than expected, $(3.3 \pm 1.2) \times 10^{-6}$ (Udalski et al., 1994). Leveraging this high optical depth, modern microlensing monitoring surveys of the Milky Way bulge provides the largest source of microlensing events and detect about 2000 photometric events per year (Udalski et al., 2015; Kim et al., 2016). The events have proved useful for a range of different astrophysical applications (Gould, 2001; Mao, 2012). Most notably, short timescale deviations, caused by planets around the lens, have been detected for 108^2 events to date (e.g. Bond et al., 2004; Udalski et al., 2005; Beaulieu et al., 2006). These events have proved a probe of low-mass planets (down to Earth-mass) at intermediate separations from their hosts (≈ 1 au) which are typically off-limits to other planet detection techniques (Tsapras, 2018).

While current dedicated monitoring surveys of the Galactic bulge have yielded thousands of photometric microlensing events, they are

² https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html

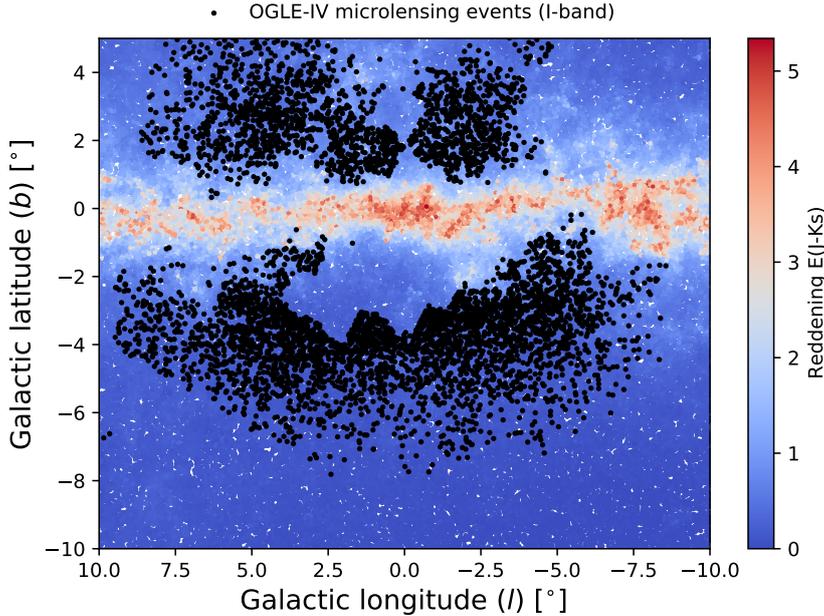


Figure 5: Locations of photometric microlensing events found in the optical by OGLE-IV and presented in (Mróz et al., 2019). Plotted in the background is the reddening map from Surot et al. (2020) which traces extinction. No events in the optical are found towards the inner regions of the Galactic bulge where extinction is high. It is noted that this plot does not contain the OGLE events found by the OGLE early warning system that are located between $-4 < b < -2$ which are not included in the Mróz et al. (2019) sample.

only conducted in the optical wavelengths (Udalski et al., 2015; Kim et al., 2016). This constrains them to high galactic latitudes ($|b| > 2^\circ$) around the bulge in windows where extinction from interstellar dust is low. Fig. 5 shows the locations of the ≈ 8000 photometric events detected by OGLE between J2010–J2017 and presented in Mróz et al. (2019). The reddening map of Surot et al. (2020) is plotted in the background. Fig. 5 shows the stark lack of events in inner regions of the Galactic bulge in the optical, where the extinction and reddening is high. Of course, this does not mean there are no microlensing events in the inner bulge regions. In fact, event rates in the innermost regions of the bulge are expected to be high (Gould, 1994), they are just not bright enough to be detected at optical wavelengths due to extinction.

Gould (1994) noted that this issue could be overcome by monitoring the inner bulge regions in the Near-Infrared (NIR) wavelengths. NIR photometry is less affected by extinction and provides a means to look through the interstellar dust. The United Kingdom Infrared Telescope (UKIRT) microlensing survey Shvartzvald et al. (2017), was the first

to survey the inner regions of the bulge, examining ≈ 10 square degree patch between J2015-J2016 close to the Galactic plane ($b \approx 0$). In addition to supporting space-based microlensing campaigns by both the Kepler and Spitzer space telescopes (e.g. [Henderson et al., 2016](#)), the UKIRT microlensing survey reported five highly extinct events missed by optical surveys demonstrating the power of observing in the NIR.

Following [Shvartzvald et al. \(2017\)](#), efforts to search regions of the inner bulge in the NIR with the Vista Variables in the Via Lactea (VVV) survey ([Minniti et al., 2010](#)) were undertaken ([Navarro et al., 2017, 2018, 2020b,a](#)). The VVV survey has been monitoring ≈ 560 squared degrees of sky centered on the Galactic bulge and inner disc in the NIR K_s -band for nearly a decade. While the VVV may seem ideal for finding microlensing events towards the inner region of the bulge, in terms of wavelength and spatial coverage, it is not a dedicated microlensing survey. Consequently, the adopted sparse and spatially varying cadences makes detecting and characterizing microlensing events difficult.

Despite the challenges, [Navarro et al. \(2017\)](#) quickly realized the importance of this data set for microlensing. They reported the discovery of 182 new microlensing events based on observations acquired between J2010 and J2015 in the three innermost tiles of the survey. This covers an area $1.68^\circ \geq \ell \geq -2.68^\circ$ and $0.65^\circ \geq b \geq -0.46^\circ$. Later, [Navarro et al. \(2020a,b\)](#) extended the microlensing search to cover 14 tiles, encompassing all of the Galactic plane in VVV. They presented a catalogue of the 630 microlensing events covering the region within $10.44^\circ \geq \ell \geq -10^\circ$ and $0.65^\circ \geq b \geq -0.46^\circ$. Even so, this is only 14 of the 348 tiles in the entire VVV survey. Searching through the whole VVV data set remained an intimidating task. The majority of events found by Navarro et al. searches were dubbed ‘forsaken’ and not used in subsequent analysis. This was due to the sparse coverage of the events resulting in ambiguous interpretation of the microlensing signals. Specifically, two or more distinct sets of parameters were consistent with the microlensing signal ([Navarro et al., 2020b](#)). These types of events are common in the VVV, so therefore interpreting them is required if the VVV data is to be fully utilized.

There have also been photometric events found by space telescopes (the Kepler K2 Campaign 9; [McDonald et al. 2021](#) and the Gaia photometric science alerts; [Hodgkin et al. 2021](#), [Wyrzykowski et al. 2020](#)). Additionally, there have been dedicated space-based follow-up cam-

paings of photometric events found by ground-based telescopes (e.g. the Spitzer space telescope; Zhang et al., 2021).

To date, there has been no definitive detection of the astrometric microlensing signal for events found via the monitoring channel. This is mainly due to the absence of large-scale dedicated space-based astrometric monitoring surveys that can achieve required sub-mas astrometric precision, although smaller-scale dedicated monitoring campaigns with HST have been attempted (Kains et al., 2016, 2017, 2018). Kains et al. (2018) followed up several bright events found photometrically by OGLE towards the bulge. Combining photometric data from OGLE with astrometric data from HST, Kains et al. (2018) was able to weakly constrain the size of the astrometric signal for six events (although, the 99.7% confidence interval contained zero astrometric signal for five of the events). This permitted weak (uncertainties $> 100\%$) or upper-limit constraints on the lens masses.

1.2 THE UTILITY OF MICROLENSING EVENTS

1.2.1 *Lens mass determination*

In the case of predictable microlensing events, it is possible to extract a gravitational mass for the lens (Paczynski, 1995; Miralda-Escude, 1996). In some scenarios where parallax effects are detected in photometric events found by the monitoring channel, the mass of the lens can also be estimated (Wyrzykowski et al., 2016). This makes these events unique opportunities to directly probe the mass of stars and stellar remnants. This is important because mass is the most influential fundamental parameter which drives a star's structure, evolution, and determines its ultimate fate. Moreover, accurate stellar masses are relied upon in nearly all areas of astrophysics. Most directly, accurate stellar masses are required for the understanding of stellar astrophysics where the mass of a star controls its internal physical processes, which ultimately determine its external observable properties. For example, in the case of stars at the tip of the Red Giant branch, being able to predict their intrinsic brightness permits them to be used as distance indicators over galactic and extragalactic scales (Serenelli et al., 2017).

Another example requiring accurate stellar masses is understanding the properties of exoplanets. The derived properties of an exoplanet are often dependent on the properties of its stellar host. For

instance, for transiting exoplanets (e.g. [Borucki et al., 2010](#)) where the planetary orbit Doppler shift in the host's spectrum has been measured (e.g. [Hojjatpanah et al., 2019](#)), the planetary radius and mass can only be determined accurately if the mass and radius of the stellar host are known ([Seager & Mallén-Ornelas, 2003](#)). In turn, accurate planetary properties are fundamental for characterizing the bulk properties and atmospheres of exoplanets ([Seager & Deming, 2010](#)).

Despite mass being a fundamental stellar property, it is not an easy quantity to measure directly. For the vast majority of stars, mass is determined indirectly, relying on stellar atmospheric modeling and evolution assumptions reconciled with photometric or spectroscopic observations of a star's surface. Most simply, the mass of a main sequence star can be determined using an empirical mass-luminosity relationship ([Hertzsprung, 1923](#); [Eker et al., 2018](#)). In either case, direct and accurate masses are a necessary ingredient and calibrator of these methods ([Serenelli et al., 2021](#)).

The main source of direct mass determinations of stars are from dynamical measurements of double-lined eclipsing (e.g. [Torres et al., 2010](#)) or visual binary systems (e.g. [Bond et al., 2020](#)). In both cases, orbital motion changes the line-of-sight velocity of the binary components which causes spectral lines to be shifted due to the Doppler effect. Measurement of the Doppler shift for both components permits their mass to be directly determined when combined with either light curve (eclipsing systems) or astrometric (visual systems) information.

However, from the standpoint of understanding stellar evolution of single stars, binary systems can potentially complicate matters. For close-separation semi-detached or contact binaries, evolution of the components of the system are driven by interactions between them. These interactions can be via mass transfer between the two components, or even the two components evolving in a common envelope of gas (e.g. [Hurley et al., 2002](#)). In these cases, binary stars have likely evolved differently to single isolated stars. In wide-orbit detached binaries, the components are separated such that they do not interact with one another, and can be treated as if each component evolved in isolation ([Hurley et al., 2002](#)). Therefore, detached binaries provide vital checks of stellar evolution models of single stars ([Torres et al., 2010](#)).

Direct masses of isolated stars which are not in binary systems would also clearly be important probes of single star stellar evolution

models. However, due to no dynamical signal from a companion, direct mass determinations of single stars are difficult. Other than microlensing, asteroseismology is the only currently used method to measure the mass of isolated objects. In asteroseismology, internal oscillations of a star are detected as high frequency signals in its light curve (e.g. [Chaplin et al., 2011](#)). Asteroseismic mass determinations are however not direct, and rely on models of the interior structure of stars (e.g. [Serenelli et al., 2021](#)). On the other hand, microlensing does not rely on any structural or evolutionary models and is the only currently used method to directly determine the mass of isolated objects (e.g. [Sahu et al., 2017](#); [Zurlo et al., 2018](#)).

The need for direct masses is particularly acute in the case of white dwarfs and studies of their Mass-Radius Relationships (MRRs). Carbon-Oxygen (CO) core white dwarfs are the final evolutionary stage for the vast majority of stars ($\approx \leq 8M_{\odot}$). They are expected to consist mainly of an electron degenerate core, surrounded by a thin envelope of non-degenerate hydrogen and helium (e.g. [Tremblay & Bergeron, 2008](#)). Their mainly degenerate nature means they are expected to follow a MRR as they evolve and cool. The white dwarf MRR is important to many areas of astrophysics. In the first instance, it is typically relied upon to calculate the mass of white dwarfs from photometric or spectroscopic measurements (e.g. [Falcon et al., 2010](#)). Additionally, the MRR's upper limit ($\approx 1.4M_{\odot}$) underpins our understanding of the progenitors of type Ia supernovae. Moreover, it is vital when using cooling white dwarfs to date stellar populations in global clusters (e.g. [Hansen et al., 2002](#)).

To a first-order approximation, in a white dwarf, the inward force of gravity is balanced by the outward pressure of the electron-degenerate gas, resulting in a white dwarf's radius being inversely proportional to the cube root of its mass ([Chandrasekhar, 1935](#)). Today, detailed evolutionary cooling models including specific degenerate core compositions, the mass of the non-degenerate interior hydrogen layers, and the effects of finite temperature are used to calculate theoretical MRRs (e.g. [Bédard et al., 2020](#)). Despite their sophistication, theoretical MRRs have to make assumptions about the interior structure of the white dwarf. This is because the mass of the gravitational stratified non-degenerate interior hydrogen, helium, and CO layers are poorly constrained by observations of the white dwarf's surface. This is particularly important in the case of the mass fraction (q_{H}) of the interior hydrogen layer. Depending on tempera-

ture and whether a ‘thick’ or ‘thin’ hydrogen is assumed, theoretical MRRs can vary between 1% and 15% (Tremblay et al., 2017).

For hydrogen-rich white dwarfs (DA type), a thick hydrogen layer is assumed $q_{\text{H}} = 10^{-4}$ which is the estimated maximum hydrogen mass post-asymptotic-giant-branch evolution models predict for a $0.6M_{\odot}$ white dwarf after residual nuclear burning (Iben & Tutukov, 1984). Helium-rich white dwarfs (non-DA type) are either created with hydrogen deficient atmospheres or their hydrogen is hidden beneath their surface (Tremblay & Bergeron, 2008). For these white dwarfs, a thin hydrogen layer is assumed $q_{\text{H}} = 10^{-10}$ and represents only trace amounts of hydrogen.

Despite the white dwarf MRR’s integral importance, observational tests of the relationship are challenging. Truly direct tests of the MRR are rare. This is only possible in cases where both the white dwarf’s mass (M) and radius (R) can be determined independent of any atmospheric models. Such a direct test has been performed in the case of eclipsing binary systems (e.g. Parsons et al., 2016). However, the configuration of these systems implies the white dwarf is always post-common envelope and has therefore evolved differently to an isolated white dwarf. Otherwise semi-empirical tests of the MRR which depend on white dwarf atmospheric models, are possible.

By fitting atmospheric models to broad-band photometry and spectroscopy (e.g. Giammichele et al., 2012), white dwarfs’ atmospheric parameters (effective temperature $T_{\text{eff}}[\text{K}]$ and surface gravity $\log g$ where g has units $[\text{cm}/\text{s}^2]$), and its solid angle can be measured. Combining the solid angle with distance information (D, from parallax), allows the photometric radius of the white dwarf to be measured (e.g. Kilic et al., 2020). The photometric radius can then be combined with $\log g$ to infer the mass of the white dwarf, and the MRR can be tested (Schmidt, 1996). The problem with this approach, however, is that both the mass and radius are entirely derived from the atmospheric models and it is often difficult to disentangle signatures of the MRR from systematic effects and degeneracies in the atmospheric models (Tremblay et al., 2017). For robust semi-empirical tests of the MRR, the photometric radius needs to be combined with a mass determination independent of atmospheric models.

For white dwarfs, it is also possible to obtain direct mass information from gravitational redshift measurements. Gravitational redshift is a general relativistic effect (Einstein, 1916) where photons climbing out of a gravitational potential well (i.e from the surface of a white

dwarf) are shifted to longer (redder) wavelengths. The effect can be measured by comparing the apparent position of absorption lines in a white dwarf's spectrum, with their expected unshifted laboratory positions, λ_0 . The shift of the absorption lines due to gravitational redshift is,

$$\Delta\lambda = \frac{G_N M \lambda_0}{R c^2}. \quad (8)$$

The difficulty with this technique is that, for white dwarfs, the shift in absorption lines due to gravitational redshift is a similar size to, and is degenerate with, the Doppler shift caused by radial motion. This means that the observed shift in spectral lines is a superposition of both effects and the gravitational redshift signal can only be isolated if the radial velocity of the white dwarf is known. Determining the radial velocity of a white dwarf is possible when it is in a binary system by taking measurements of its companion (e.g. [Joyce et al., 2018](#)). The MRR can also be studied using gravitational redshift for groups of white dwarfs that are co-moving (e.g. [Pasquini et al., 2019](#)), or by averaging over random radial motions (e.g. [Falcon et al., 2010](#); [Chandra et al., 2020](#)).

The most precise direct masses (and semi-empirical tests of the MRR) for white dwarfs that are not post-common envelope, come from astrometric measurements of visual binary systems. Fig. 6 shows the MRR for nearby white dwarfs in visual binaries with direct mass determinations; 40 Eri B ([Bond et al., 2015](#)), Procyon B ([Bond et al., 2017a](#)), and Sirius B ([Bond et al., 2017b](#)). Fig. 6 also shows Stein 2051 B, which also happens to be in a visual binary system but its mass has been determined by astrometric microlensing ([Sahu et al., 2017](#)). All of these objects have photometric radius determinations, and are in agreement with the theoretical MRRs ([Bédard et al., 2020](#)). A semi-empirical test of the MRR for an isolated white dwarf has yet to be performed.

Also shown in Fig. 6 for comparison is the early theoretical MRR for a zero-temperature iron (Fe) degenerate core from [Hamada & Salpeter \(1961\)](#). This is usually indicated on MRR plots because it was originally invoked by [Provencal et al. \(1997\)](#) as a challenge to stellar evolution theory (iron white dwarf cores are not expected) to explain early (imprecise and ultimately inaccurate e.g. [Bond et al., 2015](#)) measurements of Procyon B's mass and radius.

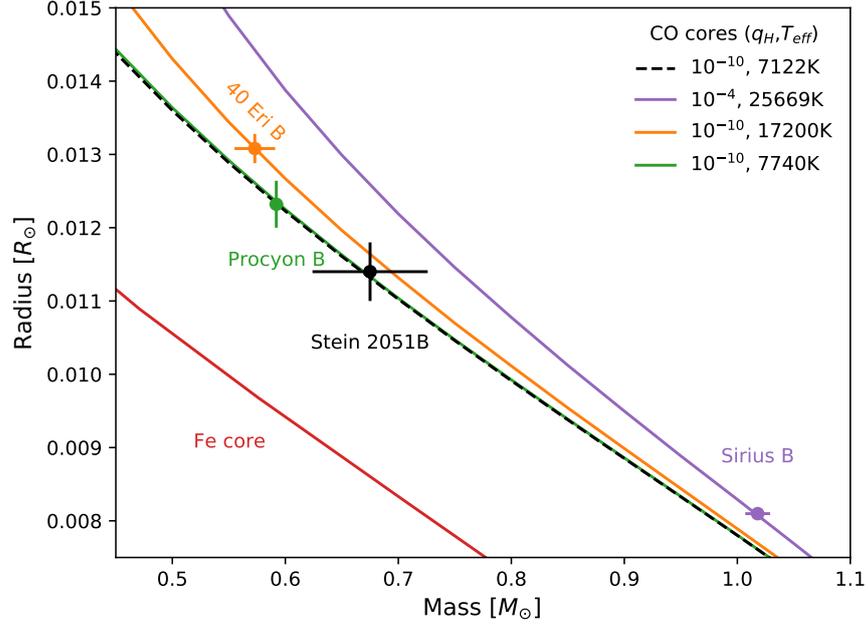


Figure 6: MRRs for nearby white dwarfs in visual binary systems. The masses for 40 Eri B, Procyon B, and Sirius B were determined from astrometric measurement of their orbits (Bond et al., 2015, 2017a,b). The error bars on Procyon B’s mass and Sirius B’s radius are too small to be seen. The mass for Stein 2051B was obtained via astrometric microlensing (Sahu et al., 2017). For comparison, the red curve shows the theoretical MRR for zero-temperature white dwarfs with an iron (Fe) core (Hamada & Salpeter, 1961). Theoretical MRRs for CO core white dwarfs were obtained from the cooling models of Bédard et al. (2020). Figure after Fig. 4 in Bond et al. (2017b).

1.2.2 Structure towards the Galactic bulge

For the majority of photometric events found via the monitoring channel, and in the absence of any higher-order effects, the only parameter that can be extracted containing physical information is the Einstein timescale (Paczynski, 1986b),

$$t_E = \frac{\Theta_E}{\mu_{\text{rel}}} = \frac{1}{v_{\text{rel}}} \sqrt{\frac{4GM_L D_L (D_S - D_L)}{c^2 D_S}}. \quad (9)$$

This is the time taken for the source to traverse Θ_E . μ_{rel} , and v_{rel} are the lens-source relative proper motions and velocities, respectively. While t_E contains information about the lens-mass and lens-source relative distance and velocities, it is completely degenerate. This means that for a single event, t_E contains little or no information about the lens. However, measurements of t_E for large samples of

events can tell us about the characteristics of the population of lenses causing them and hence probe the structure of the Galaxy (e.g. [Han & Gould, 2003](#); [Kerins et al., 2009](#)). For a sample of photometric events there are three main observables.

Firstly, the photometric optical depth mentioned in Section 1.1.1, can be calculated using t_E for a sample of N microlensing events. Since the photometric optical depth can be interpreted as the fraction of sky covered by the Einstein ring of all lenses, it can be calculated by comparing the average time a source spends in the Einstein ring ($\pi t_E/2$) to the total survey observation time ([Udalski et al., 1994](#)),

$$\tau_{\text{phot}} = \frac{\pi}{2N_S T_{\text{obs}}} \sum_{i=1}^N \frac{t_{E,i}}{\epsilon(t_{E,i})}. \quad (10)$$

N_S is the total number of stars monitored, $t_{E,i}$ is the t_E of the i th microlensing events, and $\epsilon(t_{E,i})$ is the detection efficiency at that timescale. This quantity only depends on the total mass density distributions of the lenses. It is independent of the kinematics and mass distribution function of the lenses ([Paczynski, 1986b](#)).

Secondly, the rate of events is τ_{phot} divided by the average timescale ([Griest, 1991](#)),

$$\Gamma_{\text{phot}} = \frac{1}{N_S T_{\text{obs}}} \sum_{i=1}^N \frac{1}{\epsilon(t_{E,i})}. \quad (11)$$

Γ_{phot} not only tells us about the distribution of lenses, but also depends on the lens and source kinematics and the mass distribution function of the lenses ([Griest, 1991](#)). Finally, there is also the global distribution of t_E . The shape of the global t_E distribution (after detection efficiency correction) contains information on the mass distribution function of the lenses ([Mao & Paczynski, 1996](#); [Awiphan et al., 2016](#); [Mróz et al., 2017](#)).

Each of these quantities can be spatially resolved over different lines of sight towards the Galactic bulge. However, an important point is that the measurement of all these quantities requires a large sample of events with a quantified detection efficiency as a function of t_E . Fig. 7 shows the spatially resolved average of the global distribution t_E in the NIR (K_S -band) distribution and towards the Galactic bulge predicted by the second-generation Manchester–Besançon Microlensing Simulator (MaB μ LS-2; [Specht et al., 2020](#)). Fig. 7 shows a spatially resolved signature in $\langle t_E \rangle$ resulting from a given model

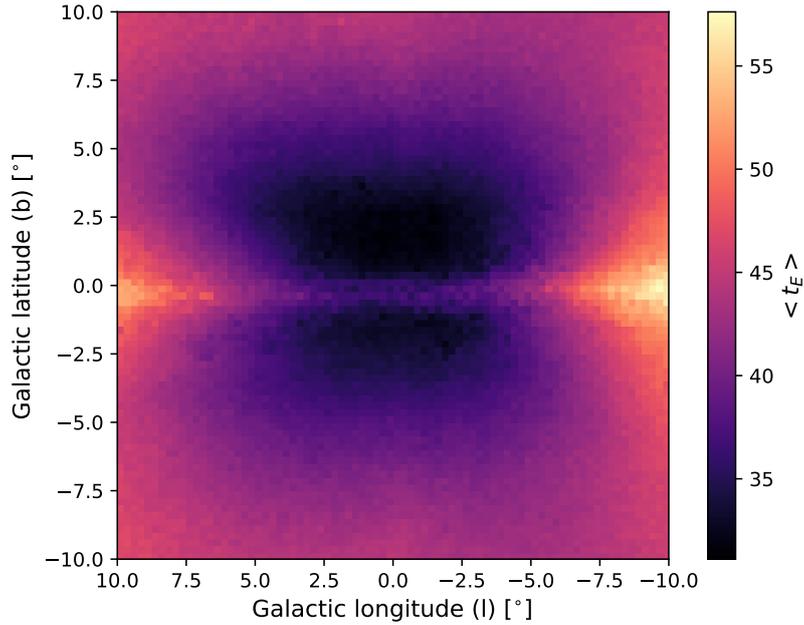


Figure 7: Spatial distribution of the expected t_E in the NIR (K_s -band) for different lines of sight towards the Galactic bulge. This plot was made using the second-generation Manchester–Besançon Microlensing Simulator (MaB μ LS-2; Specht et al., 2020). A particular Galactic model with assumed kinematics, lens mass and distance distributions, and structure, produces a particular spatial signal of t_E . This was obtained from the MaB μ LS-2 online simulator (www.mabuls.net).

of the Galaxy which can be compared against a detected sample of microlensing events (e.g. Wyrzykowski et al., 2015; Mróz et al., 2019).

The first measurements of the optical depth towards the bulge ($\tau_{\text{phot}} \approx 3.3 \times 10^{-6}$; Alcock et al., 1993; Udalski et al., 1994) were significantly higher than predicted by theoretical models including only disc stars as lenses ($\tau_{\text{phot}} \approx 5 \times 10^{-7}$; Griest et al., 1991; Paczynski, 1991). This prompted consideration of bulge as well as disc stars as lenses (Kiraga & Paczynski, 1994). This revealed that bulge star lenses contribute the majority of the optical depth ($\approx 60\%$), revising the theoretical prediction to $\tau_{\text{phot}} \approx 8.5 \times 10^{-7}$. To try and close the remaining theory-observation shortfall, consideration of a Galactic bar, elongated along the line of sight (Paczynski et al., 1994; Evans, 1994) was proposed. These models of the Galactic bar had varying masses and orientations (e.g. Han & Gould, 2003; Evans & Belokurov, 2002) but resulted in values of $\tau_{\text{phot}} \approx 1 \times 10^{-6}$, somewhat closer to the observed values.

All models thus far had underestimated the effect of blending. Blended light is unresolved light from another star, and is common in crowded areas of the Galactic bulge. This light makes it hard to robustly estimate t_E (Alcock et al., 1997), and ultimately results in biased optical depth estimations (e.g. Woźniak & Paczyński, 1997). To overcome this, Popowski et al. (2001) suggested using only bright Red Clump Giant stars (RCGs) to calculate the optical depth. This has the advantage that bias inducing blending effects would be minimized, but still could not be completely ignored (e.g. Smith et al., 2007). Optical depth calculations using only RCGs followed (e.g. OGLE-II, $\tau_{\text{phot}} \approx 2.55 \times 10^{-6}$; Sumi et al. 2006, EROS 2, $\tau_{\text{phot}} \approx 0.94 \times 10^{-6}$; Afonso et al. 2003) and resulted in lower values than calculations based on all stars. Using larger samples of events with RCGs (≈ 400 events) MOA-II (Sumi & Penny, 2016), after correcting for underestimated star counts (Sumi et al., 2013), found good agreement between RCGs and all star calculations of the optical depth.

The most recent and extensive microlensing maps towards the Galactic bulge are based on samples of $> 10^3$ events from OGLE (Wyrzykowski et al., 2015, 2016; Mróz et al., 2019). Using 3718 events found during 8 years of the OGLE-III campaign, Wyrzykowski et al. (2015) computed spatially resolved maps of $\langle t_E \rangle$ which were generally in good agreement with the Besançon galactic models of the bulge (Kerins et al., 2009). Wyrzykowski et al. (2015) noted the asymmetry in the $\langle t_E \rangle$ distribution around $l = 0$, attributed to the structure and viewing angle of the Galactic bar, was slightly more pronounced than expected. Using a larger sample of 8000 events found during 8 years from the OGLE-IV campaign (Udalski et al., 2015), Mróz et al. (2019) investigated $\langle t_E \rangle$, optical depth, and event rate compared with the MaBulS-2 (Specht et al., 2020) Galactic models. In general, Mróz et al. (2019) found good agreement with Wyrzykowski et al. (2015) regarding $\langle t_E \rangle$ and good agreement with MaBulS-2 for predicted $\langle t_E \rangle$, optical depth, and event rates.

All current tests of the MaBulS-2 galactic microlensing models have been in the optical wavelengths (Sumi & Penny, 2016; Mróz et al., 2019) and therefore restricted to low-extinction windows at high galactic latitudes ($1^\circ < |b| < 6^\circ$ Specht et al., 2020). Even though Navarro et al. (2018) and Navarro et al. (2020a) extracted microlensing events along strips of longitude and latitude in the NIR and covered regions of $|b| < 1^\circ$, detailed comparisons with Galactic models were not carried out. Regarding the latitudinal strip, Navarro et al. (2020a)

simply reported an observed decrease in events with increasing latitude and a shorter mean timescale to events found in the Galactic plane. Regarding the longitudinal strip, Navarro et al. (2018) noted a smooth increase in the number of events towards the Galactic centre, with a slight asymmetry to more events at negative latitudes. While the work of Navarro et al. (2018, 2020a) has demonstrated the power of the VVV for microlensing, work utilizing the full survey is the clear next step.

1.3 DATA

The results in this thesis are based on two major sources of data, from the Gaia satellite and the VVV survey. In this section, I outline their properties.

1.3.1 *Gaia*

The Gaia satellite was launched on the 19th of December 2013, with the primary objective of creating the most detailed and precise three-dimensional map of our home galaxy, the Milky Way (Gaia Collaboration et al., 2016a). From Gaia’s vantage position at the second Lagrangian point, it rotates and precesses continuously, scanning the sky and taking multiple measurements of $\sim 10^9$ stars in the optical G-band (330 – 1050nm). During Gaia’s nominal five-year mission, it will see each star on average ~ 80 times. However, extensions to the mission have already been announced and Gaia will continue gathering data until at least the end of 2022.³ By the end of its lifetime, Gaia will have produced a rich data set of unprecedented quality and size. In this section, both the Gaia science instruments and how data is collected by the satellite are briefly outlined, focusing on astrometric measurements. The contents of the current incremental data releases, on which the work in this thesis is based, are described.

The Gaia satellite consists of two identical telescopes separated by a constant angle of 106.5° along the scanning direction. As Gaia rotates, objects scan across the shared focal plane of 106 Charge-Coupled Devices (CCDs). On an object’s ≈ 10 second journey across the focal plane, it first transits across the two columns of Sky Mapper CCDs. Here, objects are automatically detected and information about their projected transit is communicated to the rest of the focal plane. Next,

³ <https://www.cosmos.esa.int/web/gaia/news>

the object transits the nine columns totaling 62 astrometric field CCDs. This yields one astrometric positional measurement of the object. For objects brighter than $G = 13$, a two-dimensional observation window is assigned. This means that both along scan (AL) and the orthogonal across scan (AC) positions are recorded. However, for the vast majority of objects which have $G > 13$, only the AL positions are kept.

For $G > 13$ objects, because only AL positions are recorded, single epoch positions in the AL direction are much better constrained (\approx by a factor of 100; Rybicki et al., 2018) than in the AC direction, rendering the measurement effectively one-dimensional (e.g. Fabricius et al., 2016). However, because Gaia sees an object many times over its mission and at different scan angles, this allows precise two-dimensional astrometric information to be derived. The single epoch AL astrometric precision (σ_{AL}) decreases sharply with increasing magnitude. For a $G = 13$ object, $\sigma_{AL} \approx 0.06$ mas whereas for a $G=18$ object, $\sigma_{AL} \approx 1.0$ mas (e.g. Rybicki et al., 2018; Bramich, 2018; Everall et al., 2021). Ultimately this astrometric data allows a fit to the standard five parameter astrometric solution, providing reference positions ($\alpha_{ref}, \delta_{ref}$), tangent plane projected proper motions ($\mu_{\alpha*}, \mu_{\delta}$) and parallax ϖ for an object. In addition to collecting astrometric measurements and G-band photometry, blue-band G_{BP} (330 – 680 nm) and red-band G_{RP} (630 – 1050nm) photometry are measured by the blue and red low-resolution spectroscopy CCDs, respectively (Evans et al., 2018).

At the end of Gaia’s mission lifetime, it will produce time-series astrometry, microarcsecond (μas) astrometric solutions⁴, G , G_{BP} , G_{RP} , and low resolution spectra for ≈ 2 billion objects. Until then, incremental data products have been released which were used in this thesis; Gaia Data Release 1 (GDR1; Gaia Collaboration et al., 2016b), Gaia Data Release 2 (GDR2; Gaia Collaboration et al., 2018), and Gaia Early Data Release 3 (GEDR3; Gaia Collaboration et al., 2021a).

GDR1 was released in September 2016 and is based on data collected between 25th July 2014 and 16th September 2015. It contains two parameter astrometric solutions (α, δ) and G-band magnitudes for ≈ 1.1 billion objects with a limiting magnitude of $G \approx 20$. Positional uncertainties are in the range 0.35 – 16.3 mas, with the larger uncertainties belonging to faint $G > 18$ objects. G-band photometric precision ranges from $\approx 1 - 10$ mmag. Additionally, for a small subset of bright $G < 13$ (≈ 2 million objects), five-parameter astrometric solutions are provided which include proper motions and parallaxes.

⁴ <https://www.cosmos.esa.int/web/gaia/sp-table1>

This subset of sources is the Tycho-Gaia Astrometric Solution catalogue (TGAS; Michalik et al., 2015) which was constructed using auxiliary data from the Tycho-2 catalogue produced by the Hipparcos satellite (van Leeuwen, 2007). Positional, proper motion, and parallax uncertainties for TGAS objects range from $0.2 - 0.75$ mas, $0.04 - 0.14$ mas/year, and $0.24 - 0.32$ mas for $G = 9 - 12$ objects, respectively⁵.

GDR2 is the successor to GDR1 and was released on April 25th 2016, and is based on data collected between July 25th 2014 and May 23rd 2016. GDR2 provides a five-parameter astrometric solution for ≈ 1.3 billion objects with a limiting magnitude of $G \approx 21$. Uncertainties in parallax and proper motion range from $0.04 - 0.7$ mas and $0.06 - 1.2$ mas/year for $G = 15 - 20$ objects, respectively. A two-parameter (α, δ) astrometric solution for an additional ≈ 300 million sources is also provided with positional uncertainties < 2 mas. G-band magnitudes for the total ≈ 1.6 billion sources are provided with uncertainties ranging from $1 - 20$ mmag. G_{BP} and G_{RP} magnitudes derived from the blue and red plate photometers are also provided for ≈ 1.3 billion objects with uncertainties ranging from $2 - 200$ mmag.⁶

GEDR3 is the first instalment of the third Gaia data release and was made available on December 3rd 2020. GEDR3 is based on data collected between July 25th 2014 and May 28th 2017 and extends the time baseline from GDR2. Five-parameter astrometric solutions are provided for ≈ 1.468 billion sources with a limiting magnitude of $G \approx 21$. GEDR3 also contains G, G_{RP}, G_{BP} magnitudes for $\approx 1.8, 1.5,$ and 1.6 billion sources, respectively.⁷

Gaia is the successor mission to the Hipparcos satellite. The Hipparcos mission took place between 1989 – 1993 and was the first space telescope dedicated to measuring stellar positions. Hipparcos produced position, proper motion, and parallax measurements for 117,955 objects at the ≈ 1 mas precision level.⁸ GDR2 produced astrometric data $\approx 1 - 2$ orders of magnitude more precise and for ≈ 4 orders of magnitude more objects than Hipparcos, representing a huge leap forward in astrometry.

⁵ <https://www.cosmos.esa.int/web/gaia/dr1>

⁶ <https://www.cosmos.esa.int/web/gaia/dr2>

⁷ <https://www.cosmos.esa.int/web/gaia/earlydr3>

⁸ <https://www.cosmos.esa.int/web/hipparcos/catalogue-summary>

1.3.2 The VVV survey

The VVV survey, and its temporal and spatial extension, the VVV eXtended (VVVX; Minniti 2018) survey, sample $> 10^9$ sources in the NIR in the southern Galactic disc and bulge over hundreds of epochs across from J2010 to J2019. These surveys utilized the VISTA Infrared Camera (VIRCAM; Sutherland et al., 2015) and covered a total area of approximately 1700 square degrees, although only the 560 square degrees of the VVV survey contains data covering the full time baseline.

Chapters 5 and 7 use the astrometric and photometric data from version two of the VVV Infrared Astrometric Catalogue (VIRAC; Smith et al. in preparation), respectively. Version two of the VVV Infrared Astrometric Catalogue (see Smith et al. 2018 for details of Version one) was produced by Leigh Smith and has yielded a time-series photometry database for $\approx 10^9$ stars in the VVV area (≈ 560 square degrees) using VVV and VVVX data. Hereafter we will refer to the VVV survey plus the VVVX survey data covering the VVV area as simply the VVV survey.

For the photometric time-series catalogue, VIRCAM stacked paw-print images from the VVV survey were processed with a modified version of the DoPHOT profile fit image reduction software (Schechter et al., 1993; Alonso-García et al., 2012). Photometric measurements were coarsely calibrated using a globally optimized zero-point plus illumination map model. A further fine calibration stage was developed and deployed. This reduced high spatial frequency coherent structure in the maps of photometric residuals to the survey average, which reduced scatter in the light curves. In tandem with the latter calibration stage, we also scale the pipeline photometric uncertainties to better account for residual scatter across each observation. This resulted in ≈ 669 million K_s -band source light curves, with average cadences ranging from $\approx 0.05 - 1.2$ per day. Magnitude of the sources ranged from K_s -band 10-18 mag, with single epoch photometric uncertainties of $\approx 3 - 100$ mmag.

The astrometric catalogue was constructed as follows. First, astrometric solutions contained in the VVV image headers were applied and then crossmatched with the source catalogue to projected GDR2 source positions at the image epoch. At each point on a 10×10 grid placed over each VIRCAM (VISTA InfraRed CAMera) array, plate constants using local Gaia reference sources were fitted. Linear in-

terpolation of the plate constants between grid points in effect produces a piecewise linear function transforming array coordinates onto the GDR2 coordinate system. Next, the transformation is applied and sources are matched between overlapping observations. Least-squares fitting implemented by the SciPy Python package (Jones et al., 2001) is used to fit either a four-parameter proper motion and position (at epoch J2014.0) solution, when a source has 5 to 19 epochs inclusive, or a five-parameter solution that also includes parallax when the source has ≥ 20 epochs. If a source has two to four epochs inclusive, then a simple average position (at the average epoch) is recorded. Crucially for the work in this thesis, all positions, proper motions, and parallaxes are anchored to the GDR2 astrometric reference frame. This results in reliable five-parameter astrometric solutions for ≈ 669 million sources with errors on the proper motion and parallax ranging from ($K_s \approx 12 - 20$ mag) $\approx 0.25 - 2.0$ mas $\approx 1.0 - 50.0$ mas, respectively. This catalogue maintains a high completeness, approximately 1 to 1.5 K_s magnitudes deeper than its predecessor.

1.4 THESIS OUTLINE

Leveraging data from both Gaia and the VVV, this thesis is concerned with looking for and characterizing microlensing events in two places that are relatively unexplored. Firstly, with astrometric data from Gaia, I search for predicted microlensing events which are caused by nearby lenses in the solar neighbourhood. Secondly, I search for photometric microlensing events in the NIR and towards the Galactic bulge with the VVV survey. Both these searches present their own challenges, opportunities, and scientific goals and ultimately guide the questions this thesis seeks to answer.

In Chapter 2, I review the microlensing signals that are required for the rest of this thesis. Then I outline the predicted events search methods common to Chapters 3, 4, and 5. Finally, I outline the Bayesian statistical methodology which underpins the work in all Chapters. This thesis progresses towards answering the following questions:

Can Gaia data be used to make microlensing event predictions?

GDR1 and GDR2 contain astrometric data orders of magnitude more precise and numerous than their predecessors which were used in previous predicted event searches. It is therefore natural to reconsider

the suggestions of Refsdal (1964) to search for predicted events. In Chapter 3, I present a search for predicted microlensing events in GDR1. I present one promising candidate in which the lens is a white dwarf. I then present the analysis of the astrometric HST follow-up of this event and extract a direct mass for the white dwarf. In Chapter 4, I present a search of GDR2 for upcoming predicted microlensing events with significant photometric signatures.

How can predicted microlensing events be modeled?

With the goal of extracting precise lens masses, analysis of predicted microlensing events involves combining prior information on the lens and source with follow-up data. I investigate the application of a Bayesian analysis framework which naturally allows this combination of information. In Chapter 3, I investigate the application of a Bayesian analysis framework to the astrometric microlensing found with GDR1. In Chapter 4, I investigate the prospects of combining prior astrometric information from Gaia with photometric follow-up data to extract lens masses for the two predicted photometric events found with GDR2.

Can Gaia data be used in combination with VIRAC to find predicted microlensing events?

Astrometric data from Gaia is incomplete in some aspects critical for finding predicted microlensing events. One of these aspects is that Gaia is incomplete for background sources towards heavily-extinct high-source-density regions of the Galactic bulge, a likely key place to find events. In Chapter 5, I investigate using the combination of GDR2 with VIRAC to find upcoming predicted events which were missed by GDR2 searches.

How reliable are microlensing event predictions with the Gaia data?

The advent of GDR2 caused a flurry of interest in predicting microlensing events. This resulted in thousands of predicted events by numerous different studies. In Chapter 6, I investigate some of the puzzling characteristics of the GDR2 predicted microlensing events and ultimately seek to answer the question of how to sort the "wheat

from the chaff" proposed by [Gould \(2000\)](#).

How can photometric microlensing events be extracted and characterized over the whole VVV survey?

Only two small strips of the VVV survey have been searched for photometric microlensing events ([Navarro et al., 2017, 2018, 2020a,b](#)). These searches primarily relied on visual inspection to find events which cannot be scaled over the ≈ 700 million sources across the entire VVV survey. Mainly due to the sparse and noisy VVV light curve data, a search of the full VVV data set is a challenging task. In [Chapter 7](#) I develop a scaleable machine learning classification algorithm which can extract microlensing events across the entire VVV survey. I investigate the use of a Bayesian analysis combined with nested sampling to characterize sparsely-sampled photometric events of the "forsaken" type previously discarded by [Navarro et al. \(2020b\)](#).

In this Chapter, I provide an explanation of the microlensing signals required for the rest of this thesis. This Chapter largely follows the literature derivations (e.g. [Dominik & Sahu, 2000](#); [Congdon & Keeton, 2018](#); [Bramich, 2018](#)), but also aims to clarify some apparent confusion in the literature about the astrometric signals. I then detail the methods used to find predicted microlensing events common to Chapters 3, 4, and 5. Finally, I outline Bayesian statistical methodology which underpins the work in Chapters 3, 4, and 7.

2.1 IMAGE POSITIONS AND AMPLIFICATIONS

Consider a point lens with mass M_L , at distance from an observer D_L , lensing light from a distance background point source (hereafter the source) at distance D_S where $D_S > D_L$. Under the thin lens approximation, we assume that a light ray travels in a straight line from the source, is deflected instantaneously at the lens plane, and then travels in a straight line to the observer. This scenario is shown in Fig. 8. The light ray deflection angle, for a ray with impact parameter b , is given by [Einstein \(1916\)](#) as,

$$\tilde{\alpha}(b) = \frac{4G_N M_L}{c^2 b}. \quad (12)$$

Our task is to figure out, for a given deflection angle $\tilde{\alpha}(b)$, where the images of the source θ are, as seen by the observer. This is achieved using the geometric relations between angles in Fig. 8. We have,

$$D_S \tan \beta = D_S \tan \theta - D_{LS} \tan \tilde{\alpha} \rightarrow D_S \beta = D_S \theta - D_{LS} \tilde{\alpha}. \quad (13)$$

Here, $b = D_L \theta$, β is the angular impact parameter, and we have used the small angle approximation ($\sin \bullet \approx \tan \bullet \approx \bullet$).¹ Next, it is useful to define the reduced deflection angle, $\alpha(\theta)$, in terms of the angular Einstein radius Θ_E ,

$$\alpha(\theta) = \frac{D_{LS} \tilde{\alpha}(D_L \theta)}{D_S} = \frac{\Theta_E^2}{\theta}. \quad (14)$$

¹ All lensing configurations considered in this thesis have angular scales $\approx 10^{-7}$ radians.

the lens-source separation line. For the derivation of lensing signals in the coming sections, it is useful to write Eq. (16) in vector form,

$$\vec{\theta}_{\pm}(\vec{u}) = \left(u \pm \sqrt{u^2 + 4} \right) \frac{\Theta_E}{2} \hat{u}. \quad (17)$$

Here, $\vec{u} = \vec{\beta}/\Theta_E$ is the normalized lens-source angular separation vector in the direction pointing away from the lens and towards the source. In Eq. (17) and all equations that follow, $\vec{\bullet}$ denotes a vector quantity, $\hat{\bullet} = \vec{\bullet}/|\vec{\bullet}|$, denotes the unit vector, and $\bullet = |\vec{\bullet}|$.

With the source image positions determined, the next task is to compute their amplifications (or magnifications). Since lensing conserves surface brightness (no photons from the source are created or destroyed), the amplification of the source is found by calculating the surface area ratio between the source image and unlensed source. Specifically, the amplification of an image with position θ , is given by,

$$|A| = \left| \frac{d\Omega_I}{d\Omega_S} \right| = \left| \frac{\theta d\theta}{\beta d\beta} \right|. \quad (18)$$

Here, $d\Omega_I$ and $d\Omega_S$ are elements of solid angle in the image and source planes, respectively. Fig. 9 shows the source image distortions responsible for the amplification. Using Eq. (18) with θ_{\pm} in Eq. (16), the image amplifications are (Paczynski, 1986b),

$$A_{\pm}(u) = \frac{u^2 + 2}{2u\sqrt{u^2 + 4}} \pm \frac{1}{2}. \quad (19)$$

2.2 LENS-SOURCE RELATIVE MOTION

With θ_{\pm} and A_{\pm} determined, the final piece to the microlensing signal is the relative lens-source trajectory. Both θ_{\pm} and A_{\pm} are functions of the normalized lens-source separation vector, \vec{u} , (or just u in the case of A_{\pm}). The simplest model for \vec{u} is a rectilinear trajectory which is commonly parameterized using (Paczynski, 1986b, 1996a) as,

$$\vec{u}(t) = \vec{u}_0 + \frac{t - t_0}{t_E} \hat{\mu}_{\text{rel}}. \quad (20)$$

Here, $\vec{u}_0 = \vec{\beta}_0/\Theta_E$ is the normalized minimum lens-source separation which occurs at time t_0 . t_E is Einstein timescale and is the time

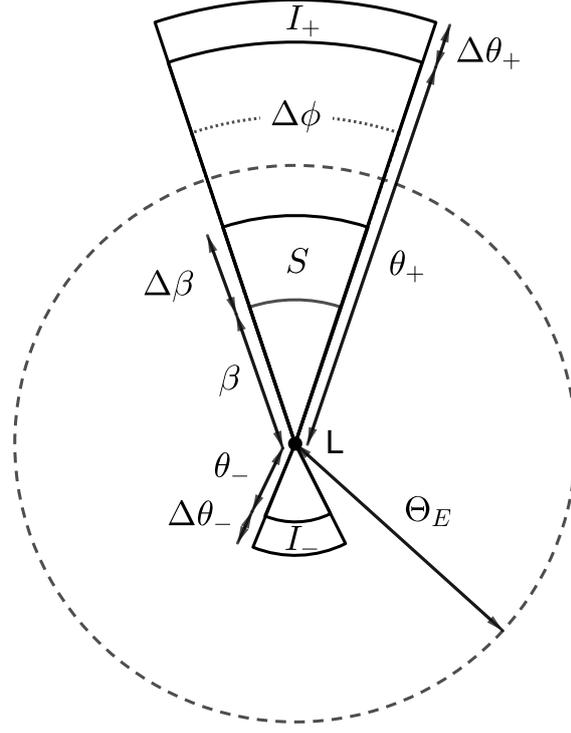


Figure 9: Amplification by source surface area distortion. For a point source and point lens the images undergo radial distortion while their tangential angular size, $\Delta\phi$, is conserved. The infinitesimal source area is $\beta\Delta\beta\Delta\phi$ while the image, I_{\pm} , areas are $\theta_{\pm}\Delta\theta_{\pm}\Delta\phi$ and $\theta_{\pm}\Delta\theta_{\pm}\Delta\phi$, respectively. Taking the ratio of the image to source areas and $\Delta \rightarrow 0$ so that $\Delta\theta_{\pm}/\Delta\beta \rightarrow d\theta_{\pm}/d\beta$, the expression in Eq. (23) is recovered. At $\beta = 0$, a circular ring of light with angular radius Θ_E (an Einstein ring), is formed. This figure was inspired by Fig. 2.3 in Congdon & Keeton (2018).

taken for the source to cross Θ_E . $\hat{\mu}_{\text{rel}}$ is the direction of relative lens source proper motion and is perpendicular to \vec{u}_0 .

The rectilinear trajectory is sufficient to describe the majority of detected photometric microlensing events to date (e.g. Mróz et al., 2019). However, this model neglects the motion of a terrestrial observer in orbit around the Sun. For events with timescales ≈ 1 year, and with similar projected lens velocities to the Earth’s orbital velocity around the Sun, the imprint of relative lens-source parallax motion (annual parallax) can be detected during an event (Alcock et al., 1995; Smith et al., 2002, 2005; Wyrzykowski et al., 2016). The relative lens-source parallax motion causes deviations of \vec{u} from the rectilinear model which manifest as asymmetrical distortions of the lensing signals. In the case of unresolved photometric events found via the monitoring

channel, the annual parallax signal is commonly parameterized by [Gould \(2000\)](#) as,

$$\vec{u}(t) = \vec{u}_0 + \frac{t - t_0}{t_E} \hat{\mu}_{\text{rel}} + \vec{\pi}(t). \quad (21)$$

Here, $\vec{\pi}(t)$ is the deviation due to relative lens-source parallax motion and is parameterized by the microlensing parallax vector $\vec{\pi}_E$. The size of $\vec{\pi}(t)$'s components are functions of time and the on-sky location of the microlensing event. Parallax effects can also be measured in the case of the same event being observed by two spatially separated observers. The observers could be Earth-based and space-based (i.e. space parallax; [Refsdal, 1966](#); [Calchi Novati et al., 2015](#)), or even two observers in different locations on the Earth (i.e. terrestrial parallax; [Holz & Wald, 1996](#); [Gould & Yee, 2013](#); [Shan et al., 2021](#)).

In more complicated lensing systems involving binary lenses or sources, the orbital motion of the lens or source can imprint onto the lensing signals ([Wyrzykowski et al., 2020](#); [Miyazaki et al., 2021](#)), but they are not important to the work in this thesis so we do not consider them further.

2.3 THE LENSING SIGNALS

2.3.1 The unresolved regime

Events found by the photometric monitoring channel are typically in the unresolved regime and the lenses are typically located at intermediate distances between the observer and the background source. In this case, the observer sees a blend of the lens, source, and possibly some unresolved stellar background. Let $g = F_b/F_s$ be the ratio of blended flux (in a given photometric band) to the source flux, the observed unresolved amplification is,

$$A_{\text{lum}}(u) = \frac{A_+ + A_- + g}{1 + g} = \frac{u^2 + 2 + gu\sqrt{u^2 + 4}}{(1 + g)u\sqrt{u^2 + 4}}. \quad (22)$$

In the case of no blended light ($g = 0$), we recover the familiar [Paczynski \(1986b\)](#) amplification,

$$A(u) = \frac{u^2 + 2}{u\sqrt{u^2 + 4}}. \quad (23)$$

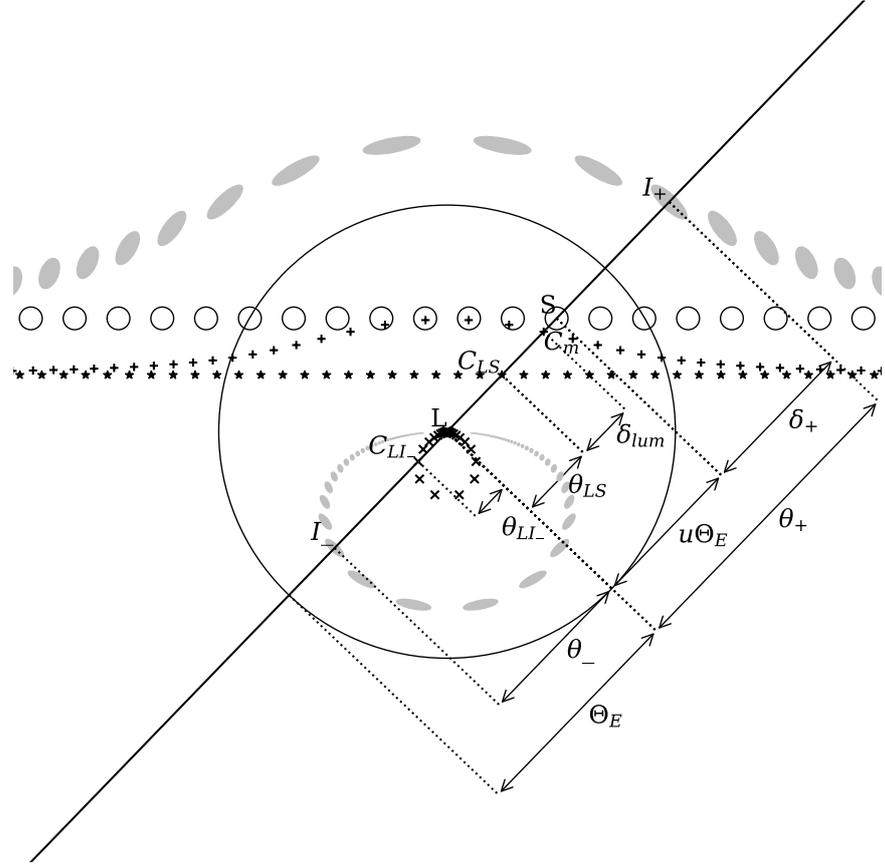


Figure 10: Configuration of the astrometric effects during a microlensing event in the lens rest frame. L marks the position of the lens at the centre. S marks the position of the source, and I_{\pm} marks the positions of the major and minor images at evenly spaced time intervals. \circ is the true position of the source, grey ellipses are the source images, + is the position of the microlensing centroid C_M , * is the position of the lens-source centroid C_{LS} , \times is the position of the lens-minor image centroid C_{LI-} . All other symbols are on sky scalar angular distances and are defined in the text. Layout of this figure was inspired by Fig. 1 in [Bramich \(2018\)](#)

This amplification is symmetric about u_0 , always > 1 , and is larger for decreasing u_0 . A is also only a function u . Assuming a rectilinear trajectory, and taking the magnitude of Eq. (20) we recover the familiar [Paczynski \(1986b\)](#) trajectory parameterisation,

$$u(t) = \sqrt{u_0^2 + \frac{(t - t_0)^2}{t_E^2}}. \quad (24)$$

Fig. 11 shows various photometric microlensing amplifications signals assuming a rectilinear lens-source trajectory and $g = 0$, or no blended light. For $u \gg 1$, we can Taylor expand the Eq. (22) in pow-

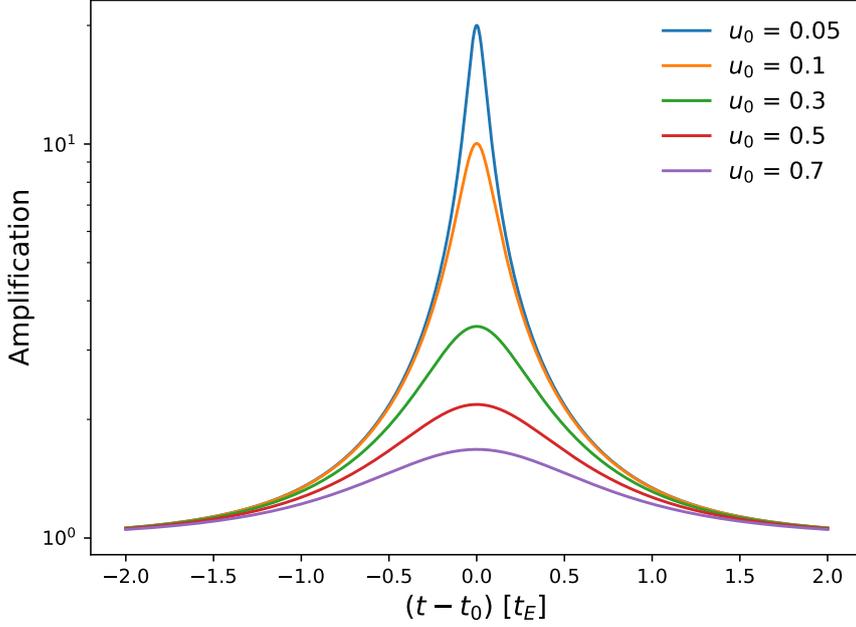


Figure 11: Example of the photometric amplification signal for a range of u_0 values, and a rectilinear lens-source trajectory. Plotted is the unresolved amplification with no blended flux $g = 0$ or A from Eq. (23)

ers of u to determine how quickly the signal strength dies off with lens-source separation (Dominik & Sahu, 2000),

$$A(u) \stackrel{u \gg 1}{\approx} 1 + \frac{2}{(1+g)u^4}. \quad (25)$$

The photometric signal falls off sharply ($\propto 1/u^4$) with lens-source separation and blended light always acts to suppress the signal. All of the above expressions suggest that the amplification can become arbitrarily large (or even infinite) as $u_0 \rightarrow 0$. However, in reality, infinite amplification is never achieved due to finite source effects and the breakdown of the point source approximation used to derive the amplification expressions (Witt & Mao, 1994; Nemiroff & Wickramasinghe, 1994; Gould, 1994). For very close alignments, the peak of the amplification is smeared out due to the finite source size (e.g. Li et al., 2019).

We can also derive the astrometric lensing signal in the unresolved regime. To proceed in this case, we first have to assume blended light only comes from the lens, ($g = F_L/F_S$), where F_L is the flux of the

lens.¹ Fig. 8 shows the astrometric quantities we are about to derive and serves as a useful guide when navigating the coming equations.

In the unresolved regime, the observer sees a blend of the lens and source images. We start by calculating the image and lens or microlensing centroid by taking the amplification-weighted average of their positions,

$$\vec{\theta}_m(\vec{u}) = \frac{A_+ \vec{\theta}_+ + A_- \vec{\theta}_-}{A_+ + A_- + g} = \Theta_E \left[\frac{u^2 + 3}{u^2 + 2 + gu\sqrt{u^2 + 4}} \right] \vec{u}. \quad (26)$$

In the absence of any lensing effects, the position of the lens-source centroid, C_{LS} , would be,

$$\vec{\theta}_{LS}(\vec{u}) = \frac{\Theta_E}{1 + g} \vec{u}. \quad (27)$$

Therefore, the centroid shift due to microlensing is calculated by taking the difference between the microlensing centroid and the lens-source centroid,

$$\vec{\delta}_{lum}(\vec{u}) = \vec{\theta}_m - \vec{\theta}_{LS} = \frac{\Theta_E}{1 + g} \left[\frac{1 + g(u^2 + 3 - u\sqrt{u^2 + 4})}{u^2 + 2 + gu\sqrt{u^2 + 4}} \right] \vec{u} \quad (28)$$

Assuming a dark lens ($F_L = 0$), the standard centroid shift result (Hog et al., 1995; Miyamoto & Yoshii, 1995; Walker, 1995) is obtained,

$$\vec{\delta}_{dark}(\vec{u}) = \frac{\Theta_E}{u^2 + 2} \vec{u}. \quad (29)$$

In contrast to the photometric effect, δ_{dark} is maximal at $u = \sqrt{2}$, with $\delta_{dark}(\sqrt{2}) = \sqrt{2}\Theta_E/4$, and is scaled by Θ_E . Assuming a rectilinear trajectory (Eq. 20) $\vec{\delta}_{dark}$, traces an ellipse, with parallel (\parallel) and perpendicular (\perp) components to the relative lens-source motion given by,

$$\vec{\delta}_{\parallel}(\vec{u}) = \frac{\tau\Theta_E}{u_0^2 + \tau^2 + 2} \hat{\mu}_{rel}, \quad \vec{\delta}_{\perp}(\vec{u}) = \frac{u_0\Theta_E}{u_0^2 + \tau^2 + 2} \hat{u}_0. \quad (30)$$

Here, $\tau = (t - t_0)/t_E$. Fig. 12 shows examples of the astrometric signal for a range of u_0 values. Notice that the parallel component of the signal contains an asymmetrical deflection about u_0 whereas the perpendicular component and overall signal magnitude are both symmetric about u_0 . In a similar fashion to the photometric signal, for

¹ This turns out to be reasonable for the astrometric events considered in this thesis. This is because the lenses considered turn out to be nearby and therefore bright, meaning the lens light dominates the blend flux.

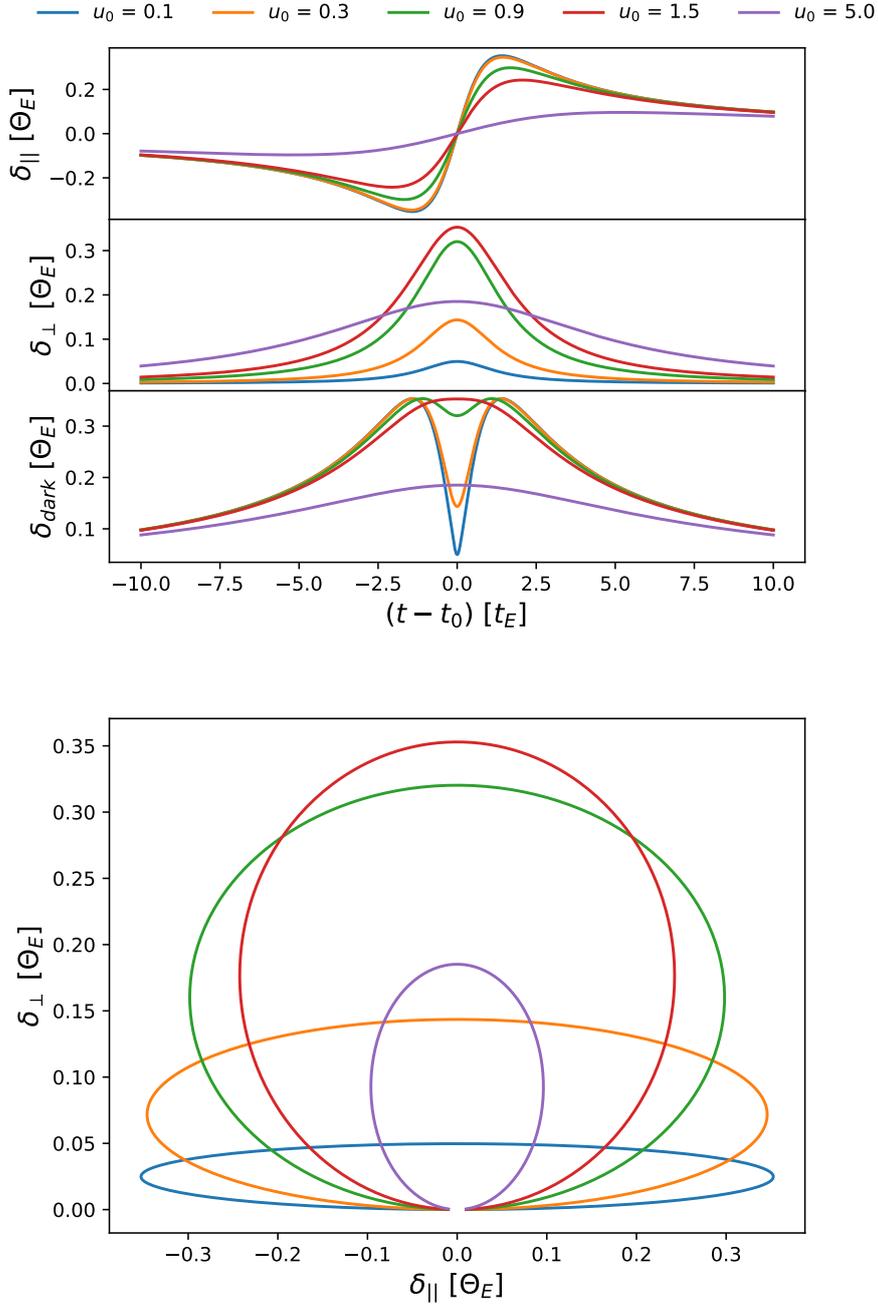


Figure 12: Example of the astrometric lensing signal for a range of u_0 values and a rectilinear lens-source trajectory. Plotted is the unresolved astrometric shift with a dark lens. **Top:** Size of the deflection in the direction parallel (δ_{\parallel}), and perpendicular (δ_{\perp}) to the relative lens-source proper motion. Overall size of the signal is shown in the bottom panel. Note that for events with $u_0 < \sqrt{2}$, the deflection is not maximal at u_0 . **Bottom:** Both components of the astrometric deflections showing that the signal traces an ellipse.

$u \gg 1$, we can Taylor expand Eq. (28) in powers of u to find how quickly the astrometric effect dies off with lens-source separation,

$$\delta_{\text{lum}}(u) \stackrel{u \gg 1}{\approx} \frac{1}{(1+g)u} \Theta_E. \quad (31)$$

In contrast to the photometric effect, the astrometric signal dies off much more slowly with lens-source separation ($\propto 1/u$). Similarly to the photometric signal, blended flux from a luminous lens acts to suppress the signal by a factor of $(1+g)$, for $u \gg 1$. The astrometric signal can also in principle be affected by finite source effects (Walker, 1995; Nucita et al., 2017). However, these effects are only important for small impact parameters of the same size as the source radius, $\beta \approx R_s$ (Walker, 1995). For the astrometric events considered in this thesis we are always in a regime where $\beta \gg R_s$, so finite source effects are not important and we do not consider them further.

For the unresolved astrometric signals, it is important to note that we are measuring an astrometric shift due to microlensing with respect to the lens-source centroid, C_{LS} , and not the unlensed source position. The lens-source centroid only coincides with the source position in the case of a dark lens ($g = F_L = 0$). Notably, the expression for δ_{lum} is incorrect in Boden et al. (1998) due to the authors calculating the shift due to microlensing from the unlensed source position and not the lens-source centroid.

2.3.2 The partially-resolved regime

Due to the astrometric capabilities of modern instrumentation (e.g. Gaia and HST), it is possible to observe microlensing events in the partially-resolved regime² (Sahu et al., 2017; Zurlo et al., 2018). Specifically, for some lensing configurations, it is possible to measure a detectable signal at wide lens-source separations larger than the instrument angular resolution, or where $\theta_+ > \theta_{\text{res}}$. In this case, the observer sees the major source image at position $\vec{\theta}_+$, and a blend of the lens-minor image centroid, C_{LI-} (see Fig. 10).

In this case, an astrometric shift of the source from the unlensed source position (not the lens-source centroid), can be calculated as (e.g. Sahu et al., 2017),

$$\vec{\delta}_+(\vec{u}) = \vec{\theta}_+ - \vec{u} = \frac{\Theta_E}{2} \left[\sqrt{u^2 + 4} - u \right] \hat{u}. \quad (32)$$

² This term was originally coined by Bramich (2018).

This signal behaves similarly to the dark lens shift, $\vec{\delta}_{\text{dark}}$ in Eq. (29) and pictured in Fig. 12, in that the signal traces an ellipse for a rectilinear trajectory, and possess both symmetrical and asymmetrical components around u_0 . Moreover, its magnitude also falls off $\approx \Theta_E/u$ for $u \gg 1$. However, δ_+ is maximal at the smallest lens-source separation at which the major image can be resolved or at $\delta_+ > \theta_{\text{res}}$, and not at $u = \sqrt{2}$ as for δ_{dark} .

Crucially, the magnitude of δ_+ is not suppressed by, and does not depend on, the flux from the lens (concretely, δ_+ does not depend on g). This is in contrast to the unresolved centroid shift in Eq. (22) which is suppressed by flux from the lens. The implication of this is that provided the lens-source separation is large enough such that θ_+ can be resolved from the lens, the astrometric lensing signal of a faint background source can in principle be detected in the presence of comparatively bright foreground lens.

The regime of a bright lens and resolvable major source image turns out to be important for the predictable astrometric microlensing events studied in this thesis. This is because predictable astrometric microlensing events are typically caused by nearby, and therefore bright lenses ($g \approx 10^3$, see e.g. Proft et al., 2011). However, HST and Gaia are typically able to resolve source separations of order of 100 mas. As an example, consider a $0.5M_\odot$ lens at a distance $D_L = 10$ pc, lensing a source in the Galactic bulge ($D_S = 8000$ pc). This system has $\Theta_E \approx 20.0$ mas. At $\beta_0 = 200$ mas corresponding to $u = 10$, $\theta_+ = 201.9$ mas (likely resolvable with HST and Gaia), and $\delta_+ = 1.9$ mas which is detectable with typical HST astrometric precision (e.g. Kains et al., 2017, 2018).

In principle, there is also a resolved photometric amplification of the major source image, A_+ which is given in Eq. (23). Similarly to the unresolved photometric amplification, A_+ falls off rapidly with increasing lens-source separation ($A_+ \approx 1 + 1/u^4$ for $u \gg 1$). As a result of this sharp fall off, for typical predictable microlensing events at $\theta_{\text{res}} \approx 100$ mas, A_+ is undetectable.

In the partially resolved regime, there is also in principle, lensing effects on the lens-minor image centroid. Taking the amplification-weighted image position of the lens and minor image, the shift from the unlensed lens position is given by Bramich (2018) as,

$$\vec{\theta}_{\text{L.L.}}(\vec{u}) = \frac{A_- \vec{\theta}_-}{A_- + g} = -\frac{(u^2 + 1)(\sqrt{u^2 + 4} - u) - 2u}{u^2 + 2 + (2g - 1)u\sqrt{u^2 + 4}} \Theta_E \hat{u}. \quad (33)$$

The accompanying photometric amplification of the lens-minor image blend is also given by [Bramich \(2018\)](#) as,

$$A_{\text{LI}_-}(\mathbf{u}) = \frac{A_- + g}{g} = \frac{u^2 + 2 + (2g - 1)u\sqrt{u^2 + 4}}{2gu\sqrt{u^2 + 4}}. \quad (34)$$

Notice that both A_{LI_-} and θ_{LI_-} are dependent on, and are suppressed by, the lens flux. Revisiting our above lensing configuration example and assuming $g = 10^3$, we find $A_{\text{LI}_-} - 1$ and $\theta_{\text{LI}_-} \approx 10^{-6}$ and $\approx 10^{-6}$ mas, respectively. In general, and for the astrometric events considered in this thesis, A_{LI_-} and θ_{LI_-} are too small to be detectable and therefore can be safely ignored.

Now that we have reviewed the microlensing signals in the partially resolved regime, it is worth pointing out the apparent confusion in the literature about the partially resolved lensing signal. In many previous microlensing event prediction studies (e.g. [Proft et al., 2011](#); [Harding et al., 2018](#); [Klüter et al., 2018a](#)), authors find events with bright lenses (large g), but with wide enough impact parameters that they can be observed in the partially resolved regime. The authors of these studies quite rightly conclude that, because the event is partially resolved, the large flux of the lens does not suppress the microlensing signal. However, this conclusion is accompanied by the incorrect assertion that deflection of the background source is given by $\vec{\delta}_{\text{dark}}$ instead of the correct expression of $\vec{\delta}_+$. Although, for $u \gg 1$, both δ_{dark} and $\delta_+ \approx \Theta_E/u$, for intermediate impact parameters, the functional forms of $\vec{\delta}_{\text{dark}}, \vec{\delta}_+$ differ, and are fundamentally different measurements. Overall, it is important that the correct deflection is used.

2.3.3 *Following an event through both regimes*

Let us return to our example lensing configuration introduced in the previous section ($M_L = 0.5M_\odot$, $D_L = 10$ pc, $D_S = 8000$ pc, $\beta_0 = 200$ mas). Instead of an impact parameter of $\beta_0 = 200$, let us instead assume a much tighter alignment of $\beta_0 = 10$ mas. Furthermore, let us also assume that the lens is 100 times brighter than the background source or that $g = 100$. If we were observing this event with Gaia ($\theta_{\text{res}} \approx 100$ mas), we may be able to detect a lensing signal as the source passes through both the resolved and unresolved regimes.

[Fig. 13](#) shows the magnitude of the lensing signals, for the example event, as the source traverses the different resolution regimes. Again

we assume a rectilinear lens-source trajectory. We can see that both A_+ , and δ_+ constitute the signals when the lens-minor image blend and major image are resolved (partially resolved regime) and then A_{lum} and δ_{lum} when the system is completely unresolved around its maximum. When the system is unresolved, the observable signals are suppressed by light from the luminous lens. For predictable microlensing events this scenario turns out to be common. Overall, it is important to consider both the resolved and unresolved signals in combination with instrumental resolution when deciding if a given event is likely to be detectable.

2.3.4 What can we infer from the lensing signals?

In the case of typical photometric events found via the monitoring channel, usually completely unresolved, and in the absence of any higher-order effects detectable in the light curve (e.g. parallax or finite source effect), t_E is the only parameter that can be measured from the photometric signal which contains physical information about the lens. However, this physical information is triple degenerate in the lens-mass, lens-source transverse velocity, and relative lens-source distance. This renders these events useless for the determination of specific lens masses.

In the case of predictable microlensing events, where the lens and source identity are known, degeneracies can be broken and the lens mass can be determined (Paczynski, 1995, 1996b). If an astrometric lensing signal is detected then Θ_E can be measured. Then, because we were able to predict the event, we readily have distance information (parallax) on the source and lens. Then the mass of the lens can be determined by inverting the expression for Θ_E and is given by Paczynski (1995) as,

$$M_L = 0.123 \frac{\Theta_E^2}{\varpi_{\text{rel}}} M_\odot, \quad \varpi_{\text{rel}} = \frac{1}{D_L} - \frac{1}{D_S}. \quad (35)$$

Here, Θ_E and ϖ_{rel} are in units of mas. For a predicted photometric event θ_E can be determined using $\theta_E = t_E \mu_{\text{rel}}$. For sufficiently distant sources ($D_S > 1000$ pc) and close lenses ($D_L < 100$ pc), Θ_E is insen-

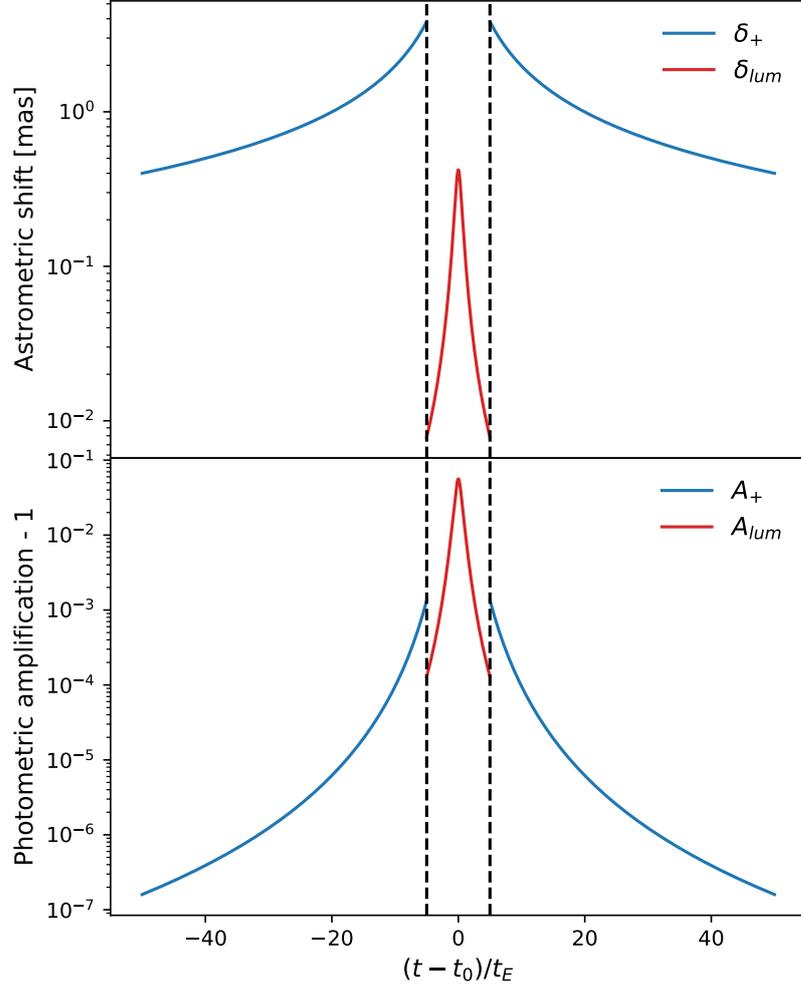


Figure 13: Following the lensing signals through the unresolved and partially resolved regimes. This is for the example event with parameters: $M_L = 0.5M_\odot$, $D_L = 10$ pc, $D_S = 8000$ pc, and $\beta_0 = 10$ mas, and assuming a rectilinear trajectory. Vertical dashed lines indicate a resolution limit of 100 mas.

sitive to D_S and can be approximated by (Proft et al., 2011; Lépine & DiStefano, 2012; Sahu et al., 2017) as,

$$\Theta_E \approx \sqrt{\frac{4GM_L}{c^2 D_L}} \quad \text{if } D_L \ll D_S. \quad (36)$$

The implication of this is that, provided you have accurate distance information on the lens, and you are sure, or willing to assume that

the source is distant, Θ_E and therefore M_L can be determined without a precise estimate for D_S (Paczynski, 1995).

Finally, it's worth noting that in general, the astrometric signals of a microlensing event are detectable over much longer timescales compared to their photometric counterparts. Photometric signals are only significant at $u \approx 1$ and fall off rapidly with increasing lens-source separation. For little or no blended light ($g \approx 0$), t_E characterize the event timescale well and is typically ≈ 20 days for stellar lenses in the Galaxy (e.g Mróz et al., 2019). Conversely, the astrometric signals fall off much more slowly with lens-source separation and so are detectable for $u \gg 1$. Therefore, the astrometric timescale depends on the measuring instrument's astrometric precision, σ_{ast} , (Belokurov & Evans, 2002; Dominik & Sahu, 2000; Honma, 2001; Proft et al., 2011). For a rectilinear lens-source trajectory and $g = 0$, this timescale is approximated by (Dominik & Sahu, 2000; Belokurov & Evans, 2002) as,

$$t_{\text{ast}} = \frac{1}{5\sqrt{2}} \left(\frac{t_E \Theta_E}{\sigma_{\text{ast}}} \right). \quad (37)$$

Assuming an example $\sigma_{\text{ast}} \approx 0.1$ mas, the duration of typical astrometric signals is of the order of months to years for stellar lenses in the Galaxy (Dominik & Sahu, 2000; Belokurov & Evans, 2002).

2.4 PREDICTING MICROLENSING EVENTS

In this section I outline the general search algorithm used to predict microlensing events. This method was applied largely unchanged between data sets and was used to predict the microlensing events in Chapters 3, 4, and 5. However, as I learnt more about the subtleties of predicting microlensing events, small improvements to the algorithm were made. The specifics of the search algorithm used are explicitly detailed in each relevant Chapter.

2.4.1 Future projections of celestial positions

Given a reference position $\vec{\phi}_{\text{ref}}$, and time t_{ref} , a proper motion $\vec{\mu}$, and parallax $\bar{\omega}$ of a celestial body, its projected future position at time t

on the celestial sphere to good approximation, is (e.g. [Urban & Seidelmann, 2014](#), Section 7.2.2.3),

$$\vec{\Phi}(t) \approx \vec{\Phi}_{\text{ref}} + (t - t_{\text{ref}})\vec{\mu} + \bar{\omega}\mathbf{J}^{-1}\vec{\mathbf{R}}_{\oplus}(t), \quad (38)$$

$$= \begin{bmatrix} \alpha_{\text{ref}} \\ \delta_{\text{ref}} \end{bmatrix} + (t - t_{\text{ref}}) \begin{bmatrix} \mu_{\alpha^*} / \cos \delta_{\text{ref}} \\ \mu_{\delta} \end{bmatrix} + \bar{\omega}\mathbf{J}^{-1}\vec{\mathbf{R}}_{\oplus}(t). \quad (39)$$

Here, $\vec{\mathbf{R}}_{\oplus}(t)$ are Cartesian Barycentric solar system coordinates in au of the Earth at time t . For the work presented in this thesis $\vec{\mathbf{R}}_{\oplus}(t)$ was retrieved via the Astropy Python package ([Astropy Collaboration et al., 2013, 2018](#)) which uses values computed from NASA JPL's Horizons Ephemeris². \mathbf{J}^{-1} is the inverse Jacobian matrix of the transformation from Cartesian to spherical coordinates projected at the reference position. This is defined as,

$$\mathbf{J}^{-1} = \begin{bmatrix} \sin \alpha_{\text{ref}} / \cos \delta_{\text{ref}} & -\cos \alpha_{\text{ref}} / \cos \delta_{\text{ref}} & 0 \\ \cos \alpha_{\text{ref}} \sin \delta_{\text{ref}} & \sin \alpha_{\text{ref}} \sin \delta_{\text{ref}} & -\cos \delta_{\text{ref}} \end{bmatrix}. \quad (40)$$

Of course, Eq. (39) is an approximation and several higher-order effects are being neglected. Among the more potentially significant higher order effects is that Eq. (39) does not account for radial motion. Radial motion can cause secular changes in an object's proper motion (perspective acceleration) over time (e.g. [Dravins et al., 1999](#)). However, these effects are only significant and detectable over multiple decades ([Lindegren & Dravins, 2021](#)), so as long as Eq. 39 isn't used to project positions far into future, it remains reasonably accurate.

Eq. (39) also assumes that there is no motion due to a companion. Binary systems can pull the centre of light away from the standard astrometric solution if the system is resolved (visual binary e.g. [Bond et al., 2017a](#)) or even if unresolved (astrometric binary e.g. [Torres, 2006](#); [Belokurov et al., 2020](#)). These binary astrometric deviations will cause Eq. (39) to be inaccurate. This problem can be mitigated by only using Eq. (39) to project the position of objects with a good fit to standard five-parameter astrometric solution, indicating the presence of no significant high-order astrometric deviations.

² <https://ssd.jpl.nasa.gov/?horizons>

2.4.2 Finding the closest approach

For a given candidate lens-source pair, the first step in predicting microlensing signals is to compute their angular separation $\beta(t) = \vec{\phi}_S(t) - \vec{\phi}_L(t)$. To find the peak microlensing signals, we need to compute the lens-source time of closest approach. Formally, we need to find t_{\min} such that,

$$t_{\min} = \min_t \beta(t) = \min_t \|\vec{\phi}_L(t) - \vec{\phi}_S(t)\|. \quad (41)$$

Due to the relative non-linear parallax motion of the source and lens it is possible that Eq. 41 has local minima. In practice this means standard gradient-based numerical optimizers can fail to find the global solution. Therefore, the global minimum to Eq. (41) is found by a two-step procedure. First an approximate time of closest approach, \tilde{t}_{\min} , is found by ignoring the parallax motion of the source and lens. This problem is linear in time and can be minimized analytically (by determining where the gradient $d\beta(t)/dt = 0$ without parallax terms). Assuming both the lens and source have the same t_{ref} ,

$$\tilde{t}_{\min} = t_{\text{ref}} - \frac{\Delta\mu_\delta(\delta_{\text{ref},L} - \delta_{\text{ref},S}) + \Delta\mu_{\alpha*}(\alpha_{\text{ref},L} - \alpha_{\text{ref},S}) \cos \delta_{\text{ref},L}}{\Delta\mu_\delta^2 + \Delta\mu_{\alpha*}^2}. \quad (42)$$

Here, the subscripts identify the lens (L) and source (S) astrometric parameters, and

$$\Delta\mu_{\alpha*} = (\mu_{\alpha*,L} - \mu_{\alpha*,S}), \quad \Delta\mu_\delta = (\mu_{\delta,L} - \mu_{\delta,S}). \quad (43)$$

\tilde{t}_{\min} is then used as an initial starting point for a global optimization algorithm (e.g. differential evolution or basin hopping; [Vernet et al., 2011](#); [Wales & Doye, 1997](#)) which is used to find t_{\min} including lens and source parallax motion. It is worth noting that the use of a global optimization algorithm without the first analytical initialization step (Eq. 42) reliably finds the global solution. However, I found that including the analytic initialization did speed the optimization up by a factor of ≈ 2 . This proved helpful for scaling the search algorithm to large data sets.

2.4.3 Estimating microlensing signals

For a given lens-source candidate pair and having obtained the lens-source closest approach, the predicted microlensing signals can be

computed with an estimate for Θ_E . An estimate for Θ_E can be obtained with a photometric estimate for the lens mass, M_L , using appropriate models fitted to the broad-band photometry or spectrum of the lens, or using empirical mass-luminosity relationships. The uncertainty or distribution in predicted trajectories and lensing signals are calculated using a Monte Carlo simulation by drawing samples from the appropriate distributions of the reported lens and source astrometric parameters and their respective errors.

2.4.4 *Finding candidate lens-source pairs*

For large modern astrometric data sets containing $\approx 10^9$ sources (e.g. GDR1 and GDR2), it is unfeasible to check if all possible lens-source pairs will align sufficiently closely with each other to produce detectable lensing signals. The first step to making this problem tractable is to first select lenses that have a high probability of lensing a background source (Salim & Gould, 2000; Proft et al., 2011). This is simply achieved by selecting a high proper motion ($\approx > 100$ mas/year) sample of stars as lenses. The reason why these stars are likely to be lenses is threefold. Firstly, high proper motion stars traverse large parts of the sky in a given time, making them more likely to align with a background source (Di Stefano, 2008a,b). Secondly, high proper motion stars are close by, meaning they have large Θ_E and so can lens sources at wider separations, overall probing greater areas of sky. Thirdly, high proper motion stars cause shorter duration lensing signals. This is useful for following up astrometric signals where a detectable variation in the astrometric signal will happen over a reasonably short period of time ($\approx < 1$ year), permitting a dedicated follow-up campaign (Dominik & Sahu, 2000; Sahu et al., 2017; Zurlo et al., 2018).

For each selected high-probability lens it's sensible to only compute microlensing signals for sources located nearby or close to the projected path of the lens. This can be achieved by selecting sources in a rectangular box around the projected proper motion of the lens (Proft et al., 2011). The advantage of this matching strategy is that the box width can be chosen to select all sources with a detectable signal (at a given threshold angular separation from the lens). The major drawback, however, is that the match is rectangular and is much slower than highly optimized circular matches (Koposov & Bartunov, 2006). For this reason I adopted a circular matching strategy

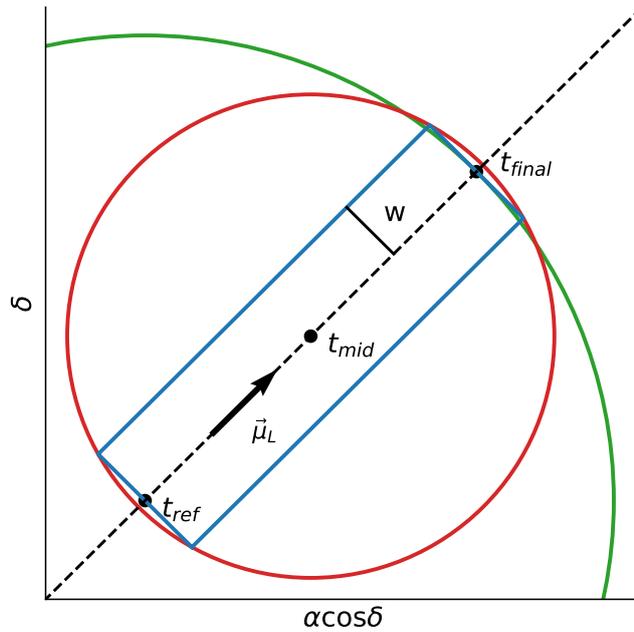


Figure 14: Different lens and source matching strategies. Dashed line indicates the projected path of the lens with proper motion $\vec{\mu}_L$. t_{ref} , t_{mid} and t_{final} indicate the position of the lens at the reference time and the middle and end times at which the predicted events occur. Sources lying within the blue rectangle which are located at a threshold separation (w) away from the lens, may have a detectable signal and should be matched. The blue match is inefficient. Alternatively, the fast green circular match can be performed (centre at t_{ref}), with the drawback of matching more sources than necessary. Optimally, the red circular match (centre at t_{mid} , radius $\sqrt{(t_{mid} - t_{ref})^2 \mu_L^2 + w^2}$) can be performed, followed by the blue match.

for the searches in Chapters 3 and 4 which was more than sufficient for matching sources and finding events with maximums in the near future. This was at the cost of matching more sources than necessary. Finally, I converged on the optimal matching strategy which was used in Chapter 5. The strategy was to first perform a large circular match followed by the rectangular match. All matching strategies are shown in Fig. 14.

2.5 BAYESIAN STATISTICS

This section provides a brief primer on aspects of Bayesian statistical methodology underpinning some of the work in Chapters 3, 4, and 7, and loosely follows some of the introductory material in Gelman et al. (2013).

Consider a model \mathcal{M} , encompassing all its assumptions, with unobserved parameters $\vec{\theta}$. We wish to make inferences on $\vec{\theta}$, conditioned on some observed data \mathcal{D} . Inferences on $\vec{\theta}$ are computed using Bayes rule,

$$p(\vec{\theta}|\mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\vec{\theta}, \mathcal{M})p(\vec{\theta}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})} \propto p(\mathcal{D}|\vec{\theta}, \mathcal{M})p(\vec{\theta}|\mathcal{M}). \quad (44)$$

Here, $p(\bullet|\bullet)$ are conditional probability density distributions, respectively. Let us detail each of the terms in Eq. (44).

- $p(\vec{\theta}|\mathcal{M})$ or the prior, encapsulates our knowledge (or ignorance) of the model parameters before any data is seen. $p(\vec{\theta}|\mathcal{M})$ can be informative and constrain model parameters to specific regions of the parameter space, or uninformative and contain little or no information on $\vec{\theta}$.
- $p(\mathcal{D}|\vec{\theta}, \mathcal{M})$ or the likelihood, encapsulates our hypothesis and assumptions on how the data, \mathcal{D} , was generated and corrupted by noise.
- $p(\vec{\theta}|\mathcal{D}, \mathcal{M})$ or the posterior, summarizes our knowledge of the model parameters under our modeling assumptions and given the observed data.
- $p(\mathcal{D}|\mathcal{M}) = \int p(\vec{\theta}|\mathcal{M})p(\mathcal{D}|\vec{\theta}, \mathcal{M})d\vec{\theta}$ or the Bayesian evidence, is the likelihood integrated over the prior volume. $p(\mathcal{D}|\mathcal{M})$ is a constant factor and does not depend on $\vec{\theta}$ leading to the proportionality in Eq. (44).

The general procedure to fit a model to data in the Bayesian framework is as follows. Firstly, construct a model that explains the data you would like to make inferences about, including how the data was corrupted by noise. Secondly, construct prior distributions for all model parameters that capture your knowledge or ignorance about each parameter. Thirdly, calculate the posterior distribution for the model parameters. Finally, check that the posterior inferences agree with the data and explore strengths and weaknesses of the model. In all but the simplest linear models, the posterior, $p(\vec{\theta}|\mathcal{M})$, is analytically intractable so we resort to approximate numerical methods to obtain samples from it (e.g. Markov Chain Monte Carlo; MCMC, or Nested Sampling; [Foreman-Mackey et al., 2013](#); [Speagle, 2020](#)).

Given that parts of this thesis (Chapters 3 and 4) are concerned with modeling predicted microlensing events, where we have prior

information on both the source and lens, which we want to combine with follow-up data of the event, a Bayesian analysis framework turns out to be the natural choice.

A PREDICTED ASTROMETRIC LENSING EVENT BY A NEARBY WHITE DWARF

This Chapter is split into three parts. In Section 3.1 we present a predictive microlensing search of GDR1, where we find one promising event where the lens is a nearby white dwarf (LAWD 37). We then obtained astrometric HST follow-up data for the predicted event and my analysis of that data is presented in Section 3.2. In Section 3.3 we discuss the implications of a gravitational mass determination for LAWD 37 from the astrometric microlensing event.

3.1 A PREDICTIVE MICROLENSING SEARCH OF GDR1

Based on work originally published in McGill et al. (2018).¹

We used the TGAS, part of GDR1, to search for candidate astrometric microlensing events expected to occur within the remaining lifetime of the Gaia satellite. Our search yielded one promising candidate. We predict that the nearby DQ type white dwarf LAWD 37 (WD 1142-645) will lens a background star and will reach closest approach on November 11th 2019 (± 4 days) with impact parameter 380 ± 10 mas. We predicted that this will produce an apparent maximum deviation of the source position of 2.8 ± 0.1 mas. We explore astrometric follow-up prospects for this event with both Gaia and HST.

3.1.1 Candidate event search

To search for events, a high proper motion (> 150 mas/year) sample of 13,206 lens stars from the TGAS catalog was selected. To narrow our search, the lens sample was cross-matched with the GDR1 source catalogue. Each lens was paired with all sources within a search radius of 10 times its proper motion. This produced a catalogue of

¹ This work was completed in collaboration with Leigh Smith (LS), Vasily Belukorov (VB), Wyn Evans (WE), and Richard Smart (RS). The search and event analysis in this section was performed entirely by myself. RS provided some of the text in Section 3.1.3. LS, VB, and WE provided useful comments on early stages of the draft of the published manuscript on which this Section is based.

≈ 4000 lens-source pairs, which we investigated further. We calculated time of closest approach (Section 2.4.2, with the differential evolution optimization algorithm: Vernet et al., 2011). The parallax motion of the lens and the proper motion of the source, where available from the "Hot Stuff for One Year" proper motion catalogue (HSOY Altmann et al., 2017), was included.

We define a candidate lensing event as a lens-source pair, which has a closest approach within the remaining Gaia mission time, assumed to be between J2018 and J2022. This left 30 candidate events. Visual inspection of the stellar field around each event removed six suspected erroneous events, which could not be confirmed to be genuine in the images available to us. Of the 24 remaining events, only one had an estimated significant maximum centroid shift > 0.4 mas. It is this event that we report on in this Chapter.

3.1.2 The candidate

We predict that the white dwarf LAWD 37 will align with a background source (G magnitudes 18 and 11 respectively) with a closest approach of $\beta_0 = 380 \pm 10$ mas ($u_0 = 11.6 \pm 0.5$) on November $11^{\text{th}} 2019 \pm 4$ days. Figure 15 shows the stellar field around the event and the trajectory of LAWD 37 as it aligns with the source. The astrometric data for both LAWD 37 and background source can be found in Table 1. Errors in the event parameters were calculated with a Monte Carlo simulation drawing 10^4 samples from Gaussian distributions, consistent with the source and lens position, proper motion and parallax, and their respective errors, and provided by the TGAS and HSOY catalogues.

At a distance of 4.6 pc, LAWD 37 (also known as WD 1142-645) is the fourth nearest known white dwarf to the Sun (Sion et al., 2009). It is classified as spectral type DQ indicating the presence of carbon in its atmosphere (Koester & Weidemann, 1982). By fitting atmospheric models (Dufour et al., 2005) to the photometry of LAWD 37, estimates for its effective temperature ($T_{\text{eff}} = 7966 \pm 219\text{K}$) and surface gravity ($\log g = 8.09 \pm 0.02$) have been obtained (Giammichele et al., 2012). Assuming LAWD 37 follows the standard evolutionary model for CO core white dwarfs, Giammichele et al. (2012) estimates its mass to be $0.61 \pm 0.01 M_{\odot}$.

Assuming that $D_S \gg D_L$ and using Eq. (36), the mass estimate of Giammichele et al. (2012) and the TGAS parallax, we find the esti-

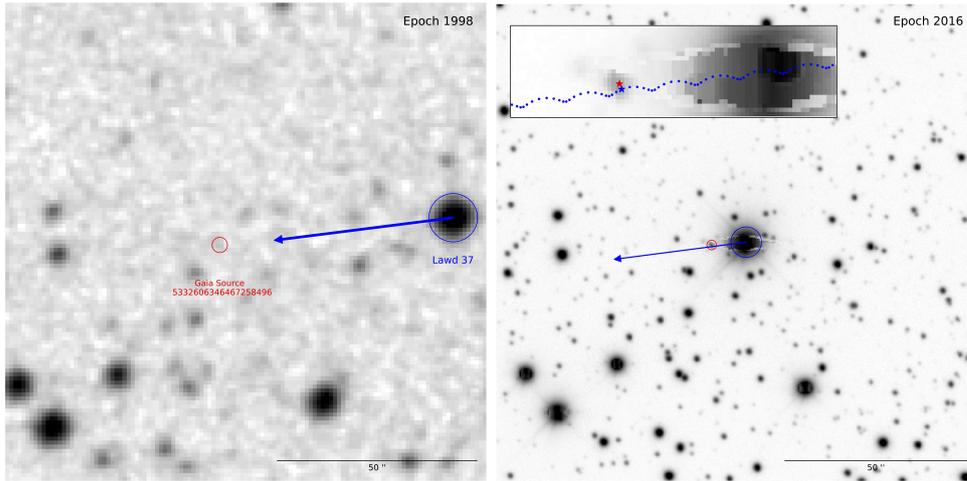


Figure 15: Images of the stellar field around the event. **Left:** Digitized Sky Survey image at epoch J1998. **Right:** Dark Energy Camera Plane Survey (DECAPS [Schlafly et al., 2018](#)) image at epoch J2016. On both images the blue circle indicates the position of the lens LAWD 37, the red circle indicates the position of the source and the blue arrow indicates the proper motion vector of LAWD 37. **Right inset:** Zoom of the DECAPS image. This shows the source and lens position at closest approach marked with red and blue stars respectively. The blue dashed line indicates the lens trajectory, which includes parallax motion.

mated Einstein Radius for LAWD 37 to be $\Theta_E = 32.8 \pm 0.3$ mas. Fig. 16 shows the estimated astrometric signal and separation of the lens and source during the event. At closest approach, the maximum astrometric shift of the source from its unlensed position is $\delta_+ = 2.8 \pm 0.1$ mas. Gaia’s resolution limit is a function of the orientation of the objects with respect to the focal plane and the magnitude difference of the objects. However, it is potentially as small as ≈ 100 mas², and is shown in Fig. 16. This suggests that Gaia is likely to observe all of this event in the partially resolved regime. Due to the event’s large impact parameter ($u_0 \gg 1$), the photometric signal is estimated to correspond to an apparent maximum brightening of the source of $< 10^{-4}$ mag. Therefore, the photometric signal is unlikely to be detected, so we consider constraining the mass of LAWD 37 from the astrometric signal only. With a closest approach of $\beta_0 = 380 \pm 10$ mas, a predicted astrometric deflection of $\delta_{+,max} = 2.8 \pm 0.1$ mas, and a lens-source magnitude difference of ≈ 7 , the viability of successfully observing this event has to be demonstrated.

² <https://www.cosmos.esa.int/web/gaia/science-performance>

	lens (LAWD 37)	background source
GDR1 Id	5332606518269523072	5332606346467258496
α_{ref} [deg \pm mas]	176.4549073 ± 0.2	176.46360456 ± 2
δ_{ref} [deg \pm mas]	-64.84295714 ± 0.2	-64.84329779 ± 2
μ_{α^*} [mas/yr]	2662.0 ± 0.2	-14 ± 3
μ_{δ} [mas/yr]	-345.2 ± 0.2	-2 ± 3
D_{\perp} [pc]	4.63 ± 0.03	-
G [mag]	11.410 ± 0.002	18.465 ± 0.005

Table 1: Lens (LAWD 37) and source astrometric data. Proper motions of the lens and source are from the TGAS and HSOY catalogues, respectively. The coordinates ($\alpha_{\text{ref}}, \delta_{\text{ref}}$) are from the Gaia DR1 source catalogue on the ICRF and at reference epoch $t_{\text{ref}} = \text{J2015.0}$. Distance to the lens D_{\perp} is obtained by inverting the lens parallax of 215.8 ± 0.2 mas from TGAS. G is the Gaia G band magnitude.

3.1.3 Observational outlook with Gaia

In order to assess the feasibility of observing the event with Gaia, we use the Gaia Observation Schedule Tool (GOST)³ to predict the dates and scan direction of the expected Gaia observations. Due to the scanning law, observations are unevenly spaced and the scan direction, which is an important predictor of the possible centroiding precision, constantly changes. In Fig. 17, we show the projected positions of the source and lens when Gaia is predicted to observe LAWD 37. We have also plotted the scan direction of Gaia and the estimated direction of deflection.

The final centroiding precision of Gaia will be determined by a combination of the scan direction and the relative positions of the two objects. Particularly for objects fainter than $G=13$, Gaia provides only binned line spread functions with very precise positions in the AL directions, but relatively low precision in the AC direction. For objects as bright as LAWD 37, Gaia will provide a window 2×1 arc-second in the AC and AL scan directions, respectively (Fabricius et al., 2016). This opens up the possibility of obtaining more precise astrometric measurements of the source with special processing when it is in the same window as LAWD 37, compared to the standard one dimensional processing pipeline for a $G = 18$ object. With both objects in the two dimensional window, it may be possible to measure the astrometric signal within a differential framework (Abbas et al., 2019). In this framework, our astrometric measurement would be the

³ <https://gaia.esac.esa.int/gost/>

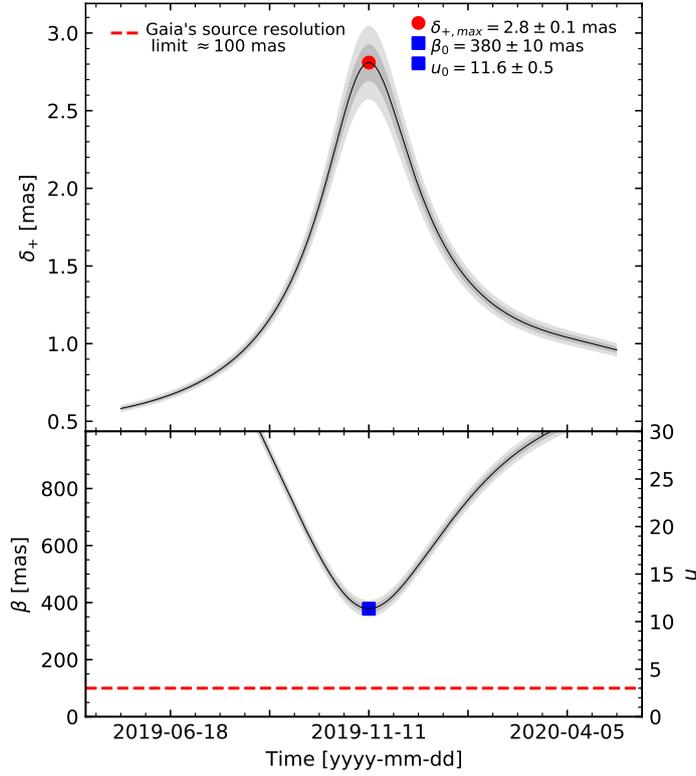


Figure 16: **Top Panel:** Magnitude of the partially resolved astrometric lensing signal around the time of closest approach (2019.86 ± 0.01 Julian Years or $2019 - 11 - 11 \pm 4$ days) for the event. Maximum astrometric shift is marked by the red circle. **Bottom panel:** Lens-source separation around the time of closest approach for the event. The minimum separation is marked with a blue square. Dark and lighter grey bands indicate 1σ and 2σ errors on both δ_+ and β , respectively. The red dashed line indicates Gaia's resolution limit of ~ 100 mas. The signal is asymmetrical about u_0 due to the relative lens-source parallax motion.

separation between the lens-minor image blend and major source image θ_+ . δ_+ could be extracted by comparisons of this separation with the projected true lens-source separation, β .

From one CCD transit, it is possible to obtain precision of 0.06 mas for a $G = 12$ object (Fabricius et al., 2016). However, because our primary measurement is of the distance between the apparent source position and lens, the floor will be set by the fainter source. When the objects are observed in the same window, the use of gates to stop LAWD 37 saturating will lead to a significantly reduced signal-to-noise ratio of the fainter source and a corresponding loss in preci-

sion. Even when in the same window, the higher AL precision will remain because the pixels are rectangular and approximately three times larger in the AC compared to the AL direction.

Let us assume an absolute best case scenario that the precision gains from a two dimensional window and the differential framework allow the apparent lens-source separation to be measured with 0.2 mas AL and 1 mas AC. This precision will be further improved by a factor of 3 as we have 9 independent estimates, one for each column in the focal plane. We simulated a uniform distribution of scan angles and assumed the AL and AC errors above and that the 9 observations provide independent measurements. From this, we find the per-epoch median error for the apparent lens source separation is $\sigma_{LS} = 0.24$ mas. Current GOST results from around the event maximum indicate that there will be approximately 30 scans in which the astrometric deflection will be $> 2\sigma_{LS}$.

Assuming that $\Theta_E = 32.8$ for LAWD 37, we may estimate the rough precision at which Gaia could determine its mass. At each Gaia transit with an expected astrometric deflection $> 2\sigma_{ls}$, we draw 10^6 samples from a Gaussian distribution centred at the expected deflection and with variance σ_{ls}^2 . We have assumed that the error on the true lens source separation is small compared with the error on the apparent lens source separation, so that $\sigma_{ls} \approx \sigma_{\text{deflection}}$, since the apparent lens-source separation is the sum of the true lens-source separation and the deflection. Using these samples and inverting the expression δ_+ (Eq. 32) for the mass of the lens, we calculate 10^6 simulated measurements for the mass of LAWD 37 at each transit. By taking the mean and variance of the mass measurement distributions for each transit and then calculating the inverse variance weighted average across all transits, we produce a final mass measurement and error. We estimate, in the best case scenario, and with special astrometric processing, that Gaia may be able to determine the mass of LAWD 37 to $\approx 3\%$ precision.

We note that, for this event and with AL and AC precision in line with the standard astrometric processing of a $G = 18$ source (Rybicki et al., 2018; Bramich, 2018) Klüter et al. (2020) finds that inference on LAWD 37's mass degrades to $\approx 19\%$. Overall, a high-precision mass will only be extracted by Gaia if precision gains can be made with special astrometric processing of the background source.

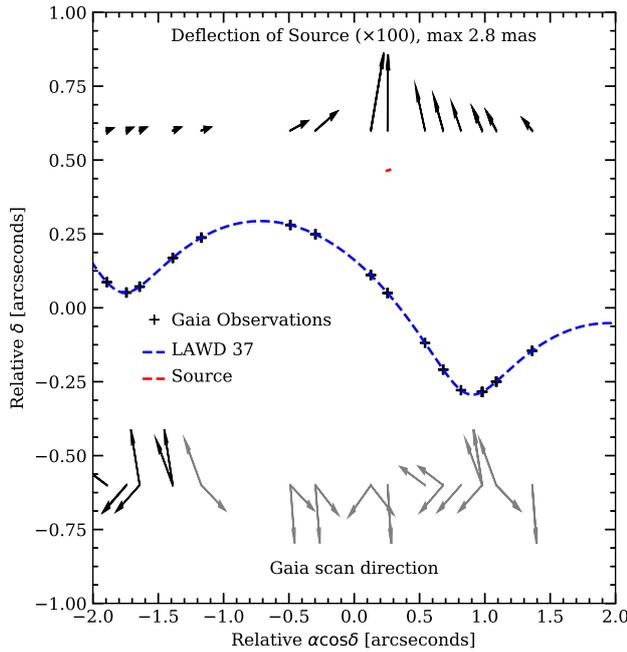


Figure 17: Blue and red dashed lines indicate LAWD 37's and the source's trajectory around the time of closest approach. Crosses mark the time of Gaia's predicted observations. The vectors at the top indicate the predicted source deflection direction and the largest deflection is 2.8 mas. The arrows at the bottom indicate Gaia's scan directions (gray arrows indicate provisional scan directions after June J2019). When the deflection arrow and Gaia's scan direction are aligned, the measurement is AL and they will be the most precise.

3.1.4 Observational outlook with HST

The outlook for astrometric follow-up with HST for this event is much more promising. This is mainly due to the fact that, unlike with Gaia, HST exposures can be optimized for maximizing astrometric centroiding precision of the source. Single measurement accuracies of ≈ 0.2 mas through pointed imaging by the Advanced Camera for Surveys and the Wide Field Camera 3 (WFC3) have been achieved with HST (Bellini et al., 2011). While spatial scanning modes with WFC3 have enabled astrometric measurements with a precision $20 - 40 \mu\text{as}$ Casertano et al. (2016), the magnitude difference of the source and lens in our event precludes using this technique (Casertano, private communication). However, for our event the maximum shift of the source position is estimated to be ≈ 2.8 mas, which is well within HST's capa-

bilities. The mass of white dwarf Stein 2051 B was determined with an accuracy $\approx 8\%$ via astrometric microlensing (Sahu et al., 2017). This event had a lens–source closest approach ≈ 100 mas. At the point that the lens and major image were still resolvable by HST (separation ≈ 500 mas), this produced a deflection of the background source position of the order of ≈ 2 mas. The Stein 2051 B event had a similar lens brightness and lens-source contrast ratio (g) to the LAWD 37 event and still the deflection was successfully measured by HST. Overall this provided an optimistic outlook for follow-up with HST.

3.1.5 Conclusion

In conclusion, we have predicted that the white dwarf LAWD 37 will lens the light from a background source, causing an apparent deflection of the source position. Maximally, this deflection will be 2.8 ± 0.1 mas on the 11th of November 2019 ± 4 days. Recent observations with HST of a comparable astrometric microlensing event have allowed the successful determination of the mass of white dwarf, Stein 2051 B, with $\approx 8\%$ accuracy. This provides an optimistic outlook for a precision mass determination of LAWD 37 from our event with HST.

3.2 ANALYSIS OF THE LAWD 37 ASTROMETRIC EVENT

Based on work done in collaboration⁴

Leveraging astrometry from Gaia and astrometric follow-up data from HST, we measure the astrometric deflection of the background source of the predicted astrometric microlensing event presented in Section 3.1. Modeling the deflection signal allows a gravitational mass for LAWD 37 to be obtained. The main challenge of this analysis is extracting the lensing signal of the faint background source whilst it is buried in the wings of LAWD 37’s point spread function (PSF). In order to model the background source position, LAWD 37’s PSF has to be removed, which introduces a significant amount of correlated noise. We find that this correlated noise can mimic the astrometric

⁴ This work was done in collaboration with Kailash Sahu, Jay Anderson, and Stefano Casertano. They carried out the HST data reduction described in Section 3.2.2.1. This included the lens point spread function subtraction and construction of the astrometric reference frame used for the rest of the analysis. Stefano Casertano provided some of the text for Section 3.2.2.1. The rest of the work in this Section was done entirely by myself.

lensing signal, and therefore has to be carefully modeled. We find a deflection model including correlated noise caused by the removal of LAWD 37's PSF best explains the data. This model yields a mass for LAWD 37 of $0.56_{-0.08}^{+0.08} M_{\odot}$. We conclude further astrometric follow-up data on the source is likely to improve the inference on LAWD 37's mass at the $\approx 3\%$ level and definitively rule out purely correlated noise explanations of the data.

3.2.1 *Outline*

The layout of this analysis is as follows. First, the two data sources, HST astrometric follow-up and GEDR₃, used in this analysis, are described in Section 3.2.2. Then, in Section 3.2.3, we describe four different models that are considered and fit to the data, and our Bayesian fitting method. In Section 3.2.4 we introduce the Bayesian Leave-One-Out cross-validation score and then use it to compare and criticize the different models. In Section 3.2.5, we present the inferences on Θ_E and LAWD 37's mass, and we investigate the sensitivity of the inferences to prior modeling assumptions.

3.2.2 *Data*

There are two sources of data used in this work. Firstly, we have single epoch HST astrometric measurements of the lensed source position over the predicted maximum of the event. Secondly, we have the astrometric solution of the source and lens from GEDR₃. These astrometric solutions provide information on the unlensed source and lens trajectories. The combination of these two sources of data will allow the astrometric shift, due to microlensing, to be measured and consequently, LAWD 37's mass to be determined. Below the characteristics of each data source are outlined.

3.2.2.1 *HST astrometric measurements*

Follow-up astrometric data for the source was obtained with HST between March J2019 and August J2020. In total, 86 images, which were divided into 9 epochs, using WFC₃ with the broad-band F814W filter, were taken. Within each epoch, all images were obtained within 47 minutes of one another. This permitted the positions of 22 stars (21

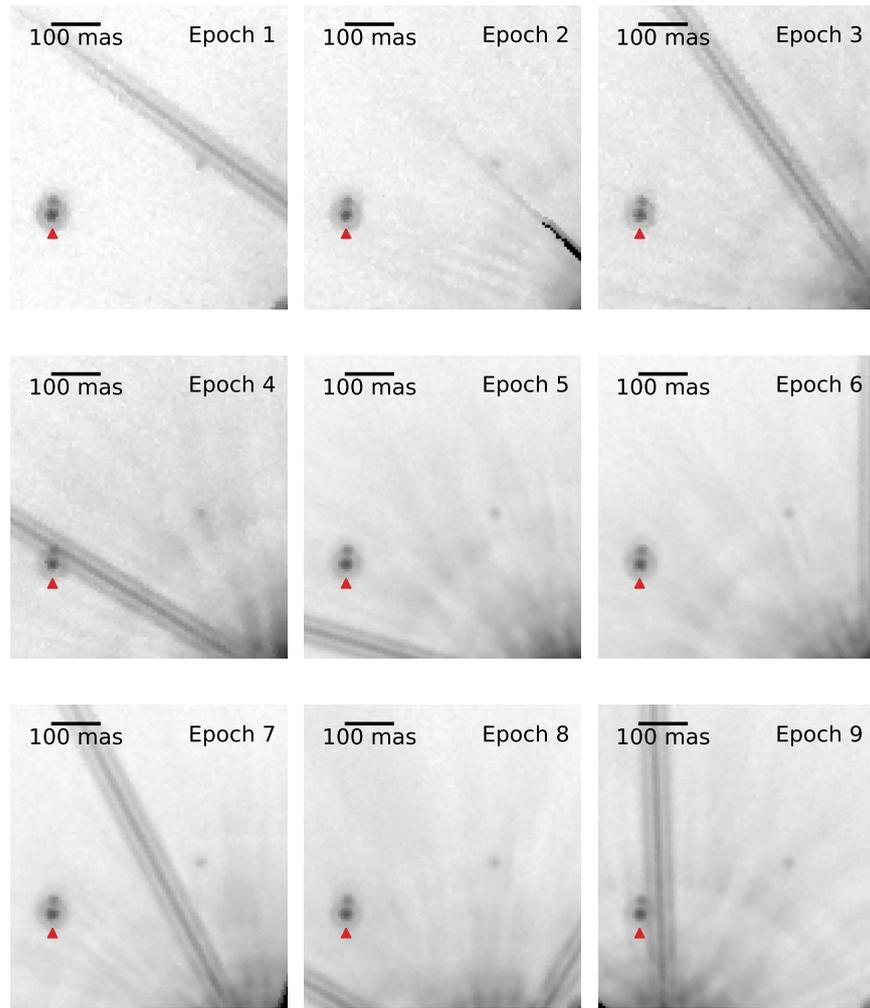


Figure 18: HST F814W-band image cutouts for each epoch of data during the LAWD 37 event. The source star is marked with a red triangle. In each image, LAWD 37 is located out of shot in the bottom or bottom right of each image. The images were made by stacking all images in a given epoch. The images shown have had LAWD 37's PSF subtracted. Streaks in all images are diffraction spikes from LAWD 37. Epoch 4 shows one of the diffraction spikes over the source.

reference stars plus the target source star) to be determined. Fig. 18 shows the stacked HST images over each of the 9 epochs.

For the reference stars, the raw positions (corrected for local geometric distortion) were measured via point spread function fitting in standard calibrated images, using the detector pixels. For all reference stars, this was a clear-cut measurement, subject primarily to white noise due to the pixel shot and readout noise. We adopt a nom-

inal position uncertainty of 0.8 mas (0.02 pixel) for each measurement. The PSF was determined from the available WFC₃ PSF library⁵. It is noted that the choice of PSF does not influence the measured positions.

Determining the raw position of the target source is more challenging. This is because of the bright image of LAWD 37 near the target source. The light from LAWD 37 needed to be estimated and subtracted before the position of the target source could be measured. For this purpose, library PSFs are not accurate enough. Instead, the light distribution for each epoch is obtained by a composite PSF computed from the images of LAWD 37 in the *other* epochs. Same-epoch images must not be used for this purpose to avoid the risk of partially subtracting the target source light. Each epoch has the target in a different location relative to LAWD 37, thus the target source in one epoch does not affect the predicted light of LAWD 37 near the target source in other epochs.

Since the PSF is known to vary with time, primarily as a consequence of HST focus drift, there is an uncertainty in the quality of the PSF for each epoch. Critically, this results in a highly-correlated, random scatter of the target source position within an epoch. Fig. 19 shows a realization of noise believed to corrupt the astrometric measurements of the target source within an epoch. The lens PSF subtraction introduces a highly-correlated, within-epoch scatter in addition to the white instrument noise scatter. In order to estimate the size of this correlated noise for each epoch, we repeat the PSF subtraction for each epoch with the PSF obtained by each of the other epochs in succession. We use the distribution of residuals as a proxy for the size of the within-epoch correlated scatter in the target position. The values recovered for the estimated size of the within-epoch correlated scatter are shown in Table 2.

Next, the reference frame was constructed using the measured positions and the GEDR₃ astrometry for all 21 reference stars and was aligned with the GEDR₃ reference epoch of J2016.0. For each image, a four-parameter transformation to define its position in the reference frame (two offsets, rotation, and scale) was adopted. For each reference star, a five-parameter standard astrometric solution, including reference position, proper motion, and parallax, was used. The astrometric parameters of the reference stars were left free, but a weak prior based on GEDR₃ (each of the astrometric parameters were al-

⁵ <https://www.stsci.edu/hst/instrumentation/wfc3>

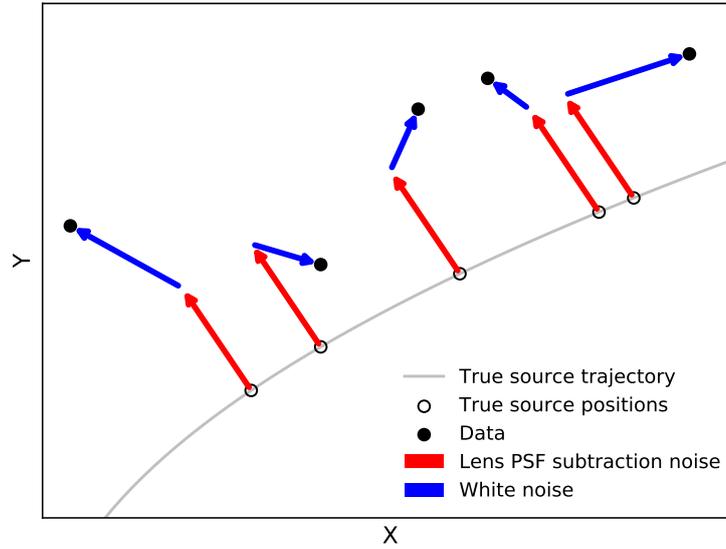


Figure 19: Diagram of how the source position data is believed to be generated within an epoch. The data receive a perfectly correlated scatter from lens PSF subtraction, which is constant for all data within an epoch (red). The data are also scattered with white noise (blue).

low to vary within five times the GEDR3 reported uncertainty), was adopted. Next a global optimization of all the parameters: four parameters for each image, plus five parameters for each star, was performed. Residuals from the solution indicate that one of the reference stars had large uncertainties and therefore was removed from the solution. As a result of this optimization, we can now transform all measurements of the target source into the desired common reference frame.

In most cases, it was possible to orient HST observations such that the diffraction spikes, or charge-bleeding columns from LAWD 37 did not lie over the source. However, during epoch 4, the possible observation orientations were too narrow to avoid the diffraction spike covering the source position (see Fig. 18). As a result, data from epoch 4 had to be discarded due to a diffraction spike causing large and hard to model uncertainties.

In the analysis that follows, all astrometric data are on the tangent plane projected at reference position right ascension $\alpha_{\text{ref}} = 176.46045$ 340° , and declination $\delta_{\text{ref}} = -64.84488414^\circ$, on the ICRS. $(\alpha_{\text{ref}}, \delta_{\text{ref}})$ corresponds to (1000, 1000) in the tangent plane with the first coordinate, X , having the direction of the local west unit vector and, the

	Epoch								
	1	2	3	4	5	6	7	8	9
$m_{\sigma_{e,\text{corr}}}$ mas	0.04	0.12	0.6	-	0.6	0.4	0.6	0.04	0.04

Table 2: Estimated sizes of the correlated noise standard deviation due to the PSF subtraction in each epoch of data. These values were obtained via simulated PSF subtraction and provided by Stefano Casertano. Note that the size of the correlated noise tends to increase when the lens and source are closest. Epoch 4 is omitted due to a diffraction spike falling over the source position.

second coordinate, Y , having the direction of the local north unit vector. Units in both of the coordinates are scaled to be 1 mas. In what follows we denote a position in this tangent plane as, $\vec{\zeta} = [X, Y]$.

3.2.2.2 Gaia astrometric solution

GEDR3 provides positions, proper motions, and parallax values for both the source and lens. Additionally, each parameter's standard errors and covariances, which are derived from a linear least squares fit of single-epoch astrometric measurements, are provided (Lindgren et al., 2021). For GEDR3, the astrometric solutions are based on measurements taken between July 2014 and May 2017⁶. The mean and covariance matrix for the (GEDR3 source Id 5332606350796955904) astrometric parameters in the tangent plane projection defined in Section 3.2.2.1 are,

$$\vec{m}_S^G = \begin{bmatrix} 35186.9625 \\ 45709.7257 \\ 8.89044 \\ 0.03676 \\ 0.19940 \end{bmatrix}, \quad \Sigma_S^G = 10^{-1} \begin{bmatrix} 0.2 & -0.01 & 0.04 & -0.06 & -0.04 \\ -0.01 & 0.2 & -0.05 & 0.003 & 0.05 \\ 0.04 & -0.05 & 0.3 & -0.02 & -0.02 \\ -0.06 & 0.003 & -0.02 & 0.3 & 0.07 \\ -0.04 & 0.05 & -0.02 & 0.07 & 0.3 \end{bmatrix}. \quad (45)$$

Here, the mean and covariance matrix have the parameter order, $X_{0,S}, Y_{0,S}$ source reference positions (mas), $\mu_{X,S}, \mu_{Y,S}$ proper motions in the X and Y directions (mas/year), and ϖ_S parallax amplitude

⁶ <https://www.cosmos.esa.int/web/gaia/earlydr3>

(mas). Similarly, for the lens (GEDR₃ source Id 5332606522595645952),

$$\bar{\mathbf{m}}_{\text{L}}^{\text{G}} = \begin{bmatrix} 45825.73799 \\ 46592.48286 \\ -2661.63959 \\ -344.93250 \\ 215.67527 \end{bmatrix}, \quad \Sigma_{\text{L}}^{\text{G}} = 10^{-3} \begin{bmatrix} 0.2 & -0.03 & -0.02 & -0.04 & -0.03 \\ -0.03 & 0.2 & -0.03 & 0.01 & 0.07 \\ -0.02 & -0.03 & 0.3 & -0.05 & 0.02 \\ -0.04 & 0.01 & -0.05 & 0.4 & 0.06 \\ -0.03 & 0.07 & 0.02 & 0.06 & 0.3 \end{bmatrix}. \quad (46)$$

This astrometric information from GEDR₃ for both the source and lens will be used in Gaussian priors for the models described in Section 3.2.3. Fig. 20 shows the GEDR₃ predicted unlensed lens and source trajectories, and the HST astrometric measurements of the source. We can see that there are clear offsets from the GEDR₃ unlensed source trajectory in the expected direction of predicted lensing signal, however, the data is clearly noisy.

3.2.3 Models

Fig. 20 suggests that we are in a low signal-to-noise regime. This is because the offset in the data is comparable to the size of the scatter at each epoch and the projected GEDR₃ unlensed position of the source. It is therefore important that we investigate a range of models. In this section we describe the four different models that were fitted to the data, and compared to one another. We fit models with and without the astrometric lensing signal and with and without a correlated noise component. We first address the choice of parameterisation of the astrometric lensing signal. Next, we outline the four different likelihoods used in each of the models. Finally, we detail the prior distributions used in each model and the MCMC methods used to sample from the posteriors in all models.

3.2.3.1 Parameterisation of the microlensing signal

Using the expression for $\vec{\delta}_+$ (Eq. 32), and calculating $\vec{\beta}$, using the lens and source trajectories, the lensed source position is dependent on 11 parameters,

$$[X_{0,S}, Y_{0,S}, \mu_{X,S}, \mu_{Y,S}, \varpi_S, X_{0,L}, Y_{0,L}, \mu_{X,L}, \mu_{Y,L}, \varpi_L, M_L]. \quad (47)$$

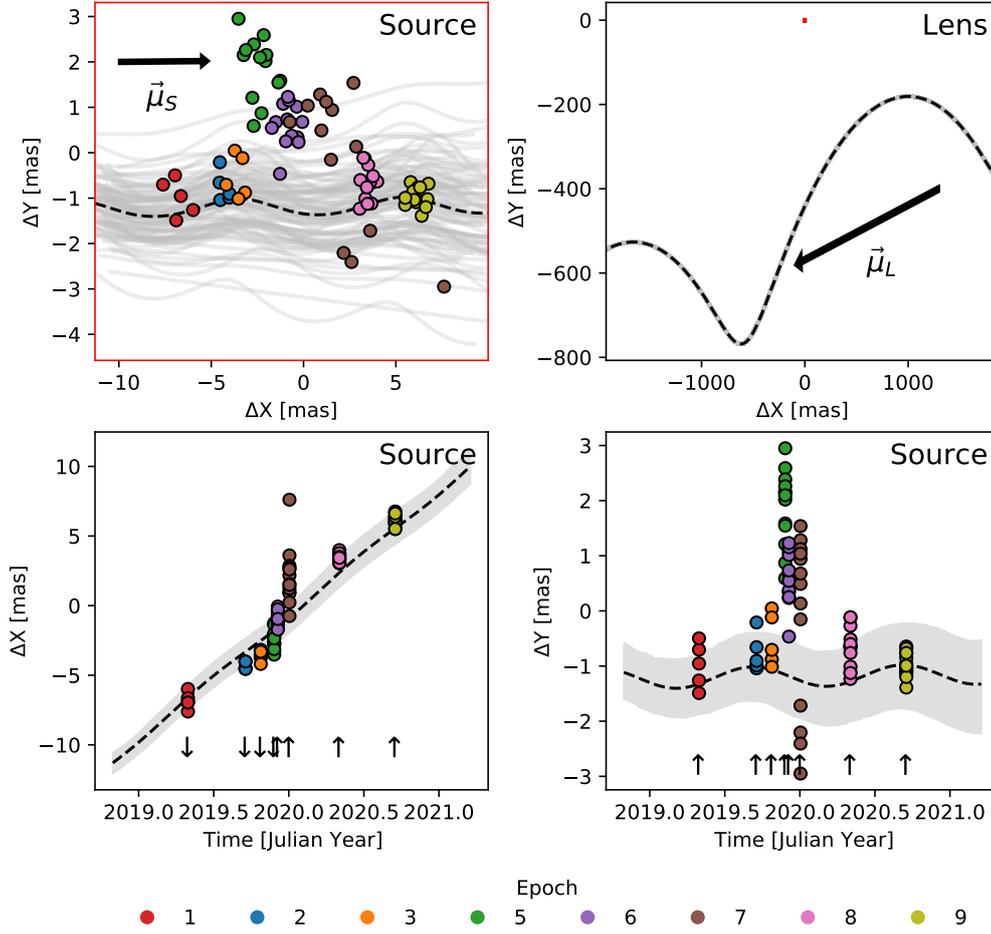


Figure 20: HST astrometric follow-up data during the predicted microlensing event by LAWD 37. **Top Left:** Single HST astrometric measurements coloured by time are shown as circles. Measurements are clustered together in time within eight distinct epochs of data. Also shown are 100 random samples of the source unlensed projected trajectory from the GEDR3 astrometric solution. Specifically, projected trajectories corresponding to samples $\vec{\theta}_S^{\text{ast}} \sim \mathcal{N}(\vec{m}_S^{\text{G}}, \Sigma_S^{\text{G}})$, are shown in grey and the black dashed line corresponds to $\vec{\theta}_S^{\text{ast}} = \vec{m}_S^{\text{G}}$. **Top Right:** Zoomed out region showing the unlensed trajectory of LAWD 37 - the lens. 100 random samples of the GEDR3 projected unlensed lens trajectory are shown where $\vec{\theta}_L \sim \mathcal{N}(\vec{m}_L^{\text{G}}, \Sigma_L^{\text{G}})$, and mean trajectory $\vec{\theta}_L^{\text{ast}} = \vec{m}_L^{\text{G}}$. The region corresponding to the top left panel is shown as a red box. **Bottom:** Projections of the source data in the X and Y directions. Small arrows indicate the predicted direction of the astrometric deflection signal.

Specifically, the source ($\vec{\theta}_S^{\text{ast}} = [X_{0,S}, Y_{0,S}, \mu_{X,S}, \mu_{Y,S}, \varpi_S]$) and lens ($\vec{\theta}_L^{\text{ast}} = [X_{0,L}, Y_{0,L}, \mu_{X,L}, \mu_{Y,L}, \varpi_L]$) astrometric parameters and the lens mass, M_L . However, modeling the signal with these parameters is not straightforward. This is due to the fact that parallax enters the model in two different ways. Firstly, the source and lens parallax controls the lens-source trajectories and hence the lens-source angular separation, and the source parallax also controls the unlensed source trajectory. Secondly, the lens and source parallax enter as distance terms and control the size of the Θ_E . The problem arises due to the interpretation of negative source parallax values when trying to include the GEDR3 astrometric solution as priors in the model.

The GEDR3 reported value of the source parallax with standard error is 0.20 ± 0.16 mas, where some of the distribution $\varpi_S < 0$ (assuming a Gaussian distribution). Using negative parallax values when ϖ_S enters the model in the source trajectory is fine as this reflects uncertainty in the parallax component of source trajectory and is completely physical. However, a negative ϖ_S value entering the model as a distance term and re-scaling Θ_E , is not physical. In this case, a negative ϖ_S value would act to artificially increase Θ_E and therefore potentially bias the inference towards lower M_L .

There are a number of ways to mitigate this problem. Most simply, the dependence of Θ_E on the source and lens parallax could be ignored, and Θ_E could be fitted for instead of M_L . In this case the astrometric parameters of the model would be $[\vec{\theta}_S^{\text{ast}}, \vec{\theta}_L^{\text{ast}}, \Theta_E]$. Here the ϖ_S only enters the model as a trajectory term and hence negative values are permitted. This simplifying assumption allows us to include negative values of ϖ_S easily at the cost of ignoring a dependence in our model. In order to justify this simplifying assumption, we have to show that ignoring the dependence does not affect the value of the astrometric signal by a detectable amount.

Rewriting Eq. (2) in terms of the relative lens source parallax $\varpi_{\text{rel}} = \varpi_L - \varpi_S$ and using the large angular separation approximation for $\delta_+ \approx \Theta_E/u$, (which is valid as $u_0 \approx 11 \gg 1$), we can find the approximate change in δ_+ given a change in ϖ_{rel} ,

$$|\Delta\delta_+| \approx \left| \frac{\partial\delta_+}{\partial\varpi_{\text{rel}}} \right| \sigma_{\varpi_{\text{rel}}} = \frac{4GM_L}{c^2\beta} \sigma_{\varpi_{\text{rel}}} \approx 0.013\sigma_{\varpi_{\text{rel}}}. \quad (48)$$

In the last equation we have assumed $M_L = 0.61M_\odot$ and $\beta_0 = 380$ mas in line with the predicted values in Section 3.1. Taking $\sigma_{\varpi_{\text{rel}}} = \sqrt{\sigma_{\varpi,L}^2 + \sigma_{\varpi,S}^2}$, and $\sigma_{\varpi,L}^2 = 0.018$ mas, $\sigma_{\varpi,S}^2 = 0.16$ mas in line with

their GEDR3-reported uncertainties, we have $|\Delta\delta_+| \approx 0.002$ mas. This value is much smaller than the single epoch astrometric precision of HST (see e.g. [Kains et al., 2017](#)), and the estimated size of the noise introduced by the lens PSF subtraction (Table 2). We therefore conclude that it is safe to ignore the dependency of Θ_E on the lens and source parallax in the modeling and fit for Θ_E as a parameter rather than M_L . Finally, M_L can then be extracted after inferring a value for Θ_E .

3.2.3.2 Likelihoods

In order to model the data, we need to setup a likelihood which encodes our assumptions on how the data was generated. Specifically, for all models, we assume the process that generated the data at time t_i , for the source is of the form,

$$\vec{\zeta}^{\text{obs}}(t_i; \vec{\theta}, \mathcal{M}_{\text{TN}}) = \vec{\zeta}^{\text{T}}(t_i; \vec{\theta}^{\text{ast}}) + \vec{\epsilon}^{\text{N}}(t_i; \vec{\theta}^{\text{noise}}). \quad (49)$$

Here, $\vec{\zeta}^{\text{obs}}$ is the observed source position in the tangent plane given by model \mathcal{M}_{TN} with trajectory component T and noise component N. $\vec{\zeta}^{\text{T}}$ is the uncorrupted source position predicted by trajectory component T with astrometric parameters $\vec{\theta}^{\text{ast}}$. $\vec{\epsilon}^{\text{N}}$ is an additive Gaussian noise component N with parameters $\vec{\theta}^{\text{noise}}$. We consider two different trajectory models, with and without the astrometric lensing deflection term. The model without the deflection is, $\vec{\zeta}^{\text{T}}(t_i; \vec{\theta}^{\text{ast}}) = \vec{\zeta}^{\text{N}}(t_i; \vec{\theta}^{\text{ast}}) \equiv \vec{\zeta}(t_i; \vec{\theta}_S^{\text{ast}})$. Here, similarly to Eq. (39),

$$\vec{\zeta}(t_i; \vec{\theta}_S^{\text{ast}}) = \begin{bmatrix} X_{0,S} \\ Y_{0,S} \end{bmatrix} + (t_i - t_{\text{ref}}) \begin{bmatrix} \mu_{X,S} \\ \mu_{Y,S} \end{bmatrix} + \omega_S \mathbf{J}^{-1} \vec{\mathbf{R}}_{\oplus}(t_i) \quad (50)$$

\mathbf{J}^{-1} is evaluated at $(\alpha_{\text{ref}}, \delta_{\text{ref}})$, and $t_{\text{ref}} = \text{J2016.0}$ is the GEDR3 reference epoch. Eq. (50) is just the standard motion of the source with proper motion and parallax. The model with the deflection is $\vec{\zeta}^{\text{T}}(t_i; \vec{\theta}^{\text{ast}}) = \vec{\zeta}^{\text{D}}(t_i; \vec{\theta}^{\text{ast}}) \equiv \vec{\zeta}(t_i; \vec{\theta}_S^{\text{ast}}) + \vec{\delta}_+(t_i; [\vec{\theta}_S^{\text{ast}}, \vec{\theta}_L^{\text{ast}}, \Theta_E])$ where $\vec{\theta}^{\text{ast}} \equiv [\vec{\theta}_S^{\text{ast}}, \vec{\theta}_L^{\text{ast}}, \Theta_E]$.

For all models, we assume that the noise is uncorrelated in the X and Y directions. With this in mind, we can write the likelihood for a set of data K_e points within an epoch, e . Let the data in epoch e be denoted as $D_e = \{\vec{t}_e, \vec{X}_e, \vec{Y}_e\}$, where the elements of D_e are vectors of length K_e , and represent the times, X positions and Y positions of

all data points within epoch e , respectively. The likelihood of data within in epoch e , is then,

$$\begin{aligned} p(D_e|\vec{\theta}_e, \mathcal{M}_{\text{TN}}) = & \mathcal{N}\left(\vec{X}_e|\vec{X}_e^T(\vec{t}_e; \vec{\theta}^{\text{ast}}), \Sigma_e^{\text{N}}(\vec{\theta}_e^{\text{noise}})\right) \\ & \times \mathcal{N}\left(\vec{Y}_e|\vec{Y}_e^T(\vec{t}_e; \vec{\theta}^{\text{ast}}), \Sigma_e^{\text{N}}(\vec{\theta}_e^{\text{noise}})\right) \end{aligned} \quad (51)$$

Here \mathcal{N} is the K_e dimensional multivariate Gaussian density, $\vec{\theta}_e \equiv [\vec{\theta}^{\text{ast}}, \vec{\theta}_e^{\text{noise}}]$, $\vec{X}_e^T(\vec{t}_e; \vec{\theta}^{\text{ast}})$ and $\vec{Y}_e^T(\vec{t}_e; \vec{\theta}^{\text{ast}})$ are vectors of X and Y source positions of length K_e obtained by evaluating the vector of times \vec{t}_e for the trajectory model component T . $\Sigma_e^{\text{N}}(\vec{\theta}_e^{\text{noise}})$ is a $K_e \times K_e$ covariance matrix. We consider two noise models for the covariance matrix. An uncorrelated white noise model with $\Sigma_e^{\text{N}}(\vec{\theta}_e^{\text{noise}}) = \Sigma_e^{\text{W}}(\vec{\theta}_e^{\text{noise}}) \equiv \sigma_{\text{white}}^2 \mathbf{I}$, where \mathbf{I} is the $K_e \times K_e$ unit diagonal identity matrix, and $\vec{\theta}_e^{\text{noise}} \equiv [\sigma_{\text{white}}]$. We also consider a correlated and white noise model with $\Sigma_e^{\text{N}}(\vec{\theta}_e^{\text{noise}}) = \Sigma_e^{\text{C}}(\vec{\theta}_e^{\text{noise}}) \equiv \Sigma_e^{\text{W}}([\sigma_{\text{white}}]) + \sigma_{e,\text{corr}}^2 \mathbf{1}$, where $\mathbf{1}$ is the $K_e \times K_e$ full ones matrix, and $\vec{\theta}_e^{\text{noise}} \equiv [\sigma_{\text{white}}, \sigma_{e,\text{corr}}]$. The correlated noise model corresponds to the data generation process shown in Fig. 19.

For all models, under the assumptions of independence between N_e epochs of data, the likelihood over the full data set is,

$$p(\mathcal{D}|\vec{\theta}, \mathcal{M}_{\text{TN}}) = \prod_{e=1}^{N_e} p(D_e|\vec{\theta}_e, \mathcal{M}_{\text{TN}}). \quad (52)$$

Here, $\vec{\theta}$ are all the model parameters and $\mathcal{D} = \{D_e\}_{e=1}^{N_e}$ is the full data set over all epochs. The consideration of two deflection model components and two noise model components leads to four distinct models to be investigated: a model with the deflection and correlation noise - \mathcal{M}_{DC} ; a model with the deflection and just white noise - \mathcal{M}_{DW} ; a model without the deflection and correlated noise - \mathcal{M}_{NC} , and a model without the deflection and just white noise - \mathcal{M}_{NW} . Table 3 contains a summary of the different models and their components.

3.2.3.3 Priors

For the source and lens astrometric parameters (10 parameters total), there is prior information from GEDR3. Specifically we assume multivariate normal distributions, $p(\vec{\theta}_S^{\text{ast}}) = \mathcal{N}(\vec{m}_S^{\text{G}}, \Sigma_S^{\text{G}})$, and $p(\vec{\theta}_L^{\text{ast}}) = \mathcal{N}(\vec{m}_L^{\text{G}}, \Sigma_L^{\text{G}})$ using the values in Eqs. (45) and (46).

There are three potential issues that need to be considered when using the GEDR3 astrometric solution as priors on the source and

	Model			
	\mathcal{M}_{DC}	\mathcal{M}_{DW}	\mathcal{M}_{NC}	\mathcal{M}_{NW}
Deflection	✓	✓	✗	✗
White noise	✓	✓	✓	✓
Correlated noise	✓	✗	✓	✗
# of parameters	20	12	14	6
parameters $\vec{\theta}$	$\vec{\theta}_S^{\text{ast}}, \vec{\theta}_L^{\text{ast}}, \Theta_E, \sigma_{\text{white}}, \vec{\sigma}_{\text{corr}}$	$\vec{\theta}_S^{\text{ast}}, \vec{\theta}_L^{\text{ast}}, \Theta_E, \sigma_{\text{white}}$	$\vec{\theta}_S^{\text{ast}}, \sigma_{\text{white}}, \vec{\sigma}_{\text{corr}}$	$\vec{\theta}_S^{\text{ast}}, \sigma_{\text{white}}$

Table 3: Summary of the components of the four considered models. Deflection indicates if the model contains the astrometric microlensing deflection term. $\vec{\sigma}_e = [\sigma_{1,\text{corr}}, \sigma_{2,\text{corr}}, \dots, \sigma_{N_e,\text{corr}}]$ is the vector of correlated noise parameters.

lens unlensed trajectory. The first two issues arise due to the implicit assumption that the GEDR3 astrometric solution for both the source and lens does not already contain some of the astrometric microlensing signal. This is a possibility as astrometric microlensing events typically have long tails (Dominik & Sahu, 2000; Belokurov & Evans, 2002), and could overlap with the data used to build the lens and source GEDR3 astrometric solutions. If the lens or source astrometric solution does contain a detectable part of the astrometric microlensing signal, this could potentially bias our inference as the lens and source astrometric parameters would not be representative of the true unlensed lens and source trajectories.

Firstly, for the source, we have to check that when the GEDR3 data was taken, there was not a significant lensing signal present. GEDR3 is based on data collected between July 2014 and May 2017. In May 2017, the predicted deflection of the source is < 0.2 mas. This is below the AL (AL) precision ($\sigma_{\text{AL}} \approx 0.8$ mas with the standard Gaia pipeline; Rybicki et al., 2018; Bramich, 2018; Everall et al., 2021) a $G \approx 18$ mag source. We therefore conclude that the astrometric lensing signal was not detectable during the time GEDR3 data was collected and therefore did not significantly influence the GEDR3 astrometric solution of the source. Secondly, for the lens, we have to check if the shift due to the blending with the minor image was detectable during GEDR3 (Eq. 33). In May 2017 $\theta_{\text{LL}} \approx 10^{-13}$ mas, we therefore safely conclude that the lens GEDR3 astrometric solution does not contain any astrometric lensing signal.

Finally, we have to consider if the GEDR3 astrometric solution of the $G \approx 18$ mag has likely been influenced by the presence of the comparatively bright $G \approx 11$ mag lens. The current Gaia processing

Parameter	Prior	Description
$\vec{\theta}_S^{\text{ast}}$	$\mathcal{N}(\vec{m}_S^G, \Sigma_S^G)$	GEDR3 prior for the source trajectory
$\vec{\theta}_L^{\text{ast}}$	$\mathcal{N}(\vec{m}_L^G, \Sigma_L^G)$	GEDR3 prior for the lens trajectory
σ_{white}	$\mathcal{N}(m_{\sigma_{\text{white}}}, \sigma_n^2)$	White component of the noise $m_{\sigma_{\text{white}}} = 0.8, \sigma_n = 0.1$
$\vec{\sigma}_{\text{corr}}$	$\mathcal{N}(\vec{m}_{\text{corr}}, \sigma_n^2 \mathbf{I})$	Correlated components of the noise for each epoch
Θ_E	$\mathcal{U}(20, 60)$	Uninformative prior of the angular Einstein radius

Table 4: Summary of the parameter priors used in the models.

pipeline is able to resolve sources for separations in the most optimal cases down to ≈ 200 mas (Arenou et al., 2018). While the lens-source contrast ratio is far from optimal in our case, in May 2017, the lens and source had a predicted separation of ≈ 6400 mas. At this separation, the lens and source were unlikely to be close to each other on the Gaia focal during the GEDR3 time baseline. Therefore, it is safe to conclude that the source GEDR3 astrometric solution is unlikely to have been significantly affected by the presence of the lens. Overall we conclude, for both the lens and source, the GEDR3 solutions are safe to use as priors on the unlensed source and lens trajectories in the models.

For the white noise parameter present in all models, we set $p(\sigma_{\text{white}}) = \mathcal{N}(\sigma_{\text{white}} | m_{\sigma_{\text{white}}}, \sigma_n^2)$ where $m_{\sigma_{\text{white}}} = 0.8$ mas and $\sigma_n = 0.1$ mas, reflecting the estimated instrument precision of WFC3. For the correlated noise parameters $\vec{\sigma}_{\text{corr}} = [\sigma_{1,\text{corr}}, \sigma_{2,\text{corr}}, \dots, \sigma_{N_e,\text{corr}}]$ we assume a Gaussian prior $p(\vec{\sigma}_{\text{corr}}) = \mathcal{N}(\vec{\sigma}_{\text{corr}} | \vec{m}_{\text{corr}}, \sigma_n^2 \mathbf{I})$, truncated at zero to avoid negative values. Here, $\vec{m}_{\text{corr}} = [m_{\sigma_{1,\text{corr}}}, m_{\sigma_{2,\text{corr}}}, \dots, m_{\sigma_{N_e,\text{corr}}}]$ are the estimated size of the correlated noise components in Table 2 and we have assumed no correlation between epochs. Finally, we then assume an uninformative uniform prior on Θ_E , as $p(\Theta_E) = \mathcal{U}(\Theta_E | \text{lower} = 20 \text{ mas}, \text{upper} = 60 \text{ mas})$. In all models, we build the full prior by taking the product over all the prior of the required parameters. Table 4 contains a summary of all parameter priors, and Fig. 21 shows the probabilistic graph illustrating all parameter dependencies and structure of the models.

3.2.3.4 Sampling the posterior

Now that we have constructed the prior and likelihoods for each model, we may compute the posterior distribution on the model parameters. Under Bayes rule, the posterior distribution or the probability distribution of the model parameters given the data, is given

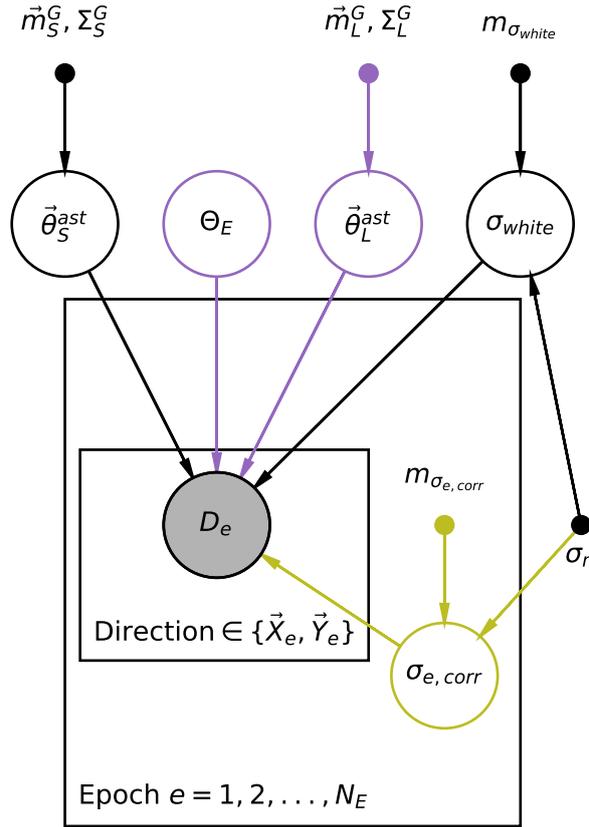


Figure 21: Probabilistic graphical model showing the dependence structure of the models considered in this work. An arrow from one node to another indicates a conditional dependence. No arrow between two nodes means there are conditionally independent. Unfilled circles are latent random variables in the models or parameters that are fitted for. Filled small circles are fixed values in the model (parameters for the informative prior distributions). The shaded circles are the observed data. Parameters inside a plate are repeated for each epoch and then direction. Parameters outside the plates are global parameters. Black parts of the graph are common to all models considered. Purple parts are common to models with an astrometric deflection. Green parts of the graph are common to models with correlated noise.

by Eq. (44). We obtain samples from the posterior distribution using an MCMC algorithm. Specifically, we use the No-U-Turn Sampler (NUTS) Hamiltonian Monte Carlo algorithm (Homan & Gelman, 2014) implemented by the PyMC3 Python package (Salvatier et al., 2016). NUTS allows samples from the posterior distribution to be obtained faster than other traditional MCMC samplers (e.g. Foreman-Mackey et al., 2013) because it uses first-order gradient information to efficiently step through the parameter space. We also take advantage of using the dense full mass matrix step implemented in the

Exoplanet Python package for further performance gains (Foreman-Mackey et al., 2021).

For each model investigated in this work, we run NUTS for 2000 tuning steps and then for a further 10,000 steps. This is done for two independent chains to permit between-chain convergence checks. Specifically, we compute the rank-normalized \hat{R} convergence diagnostic for each inference in this analysis. \hat{R} measures convergence by comparing between-chain and within-chain variance for each parameter. A value $\hat{R} > 1.01$ indicates poor convergence (Vehtari et al., 2021). We find for all parameters in all inferences considered, $\hat{R} = 1.0$, suggesting good convergence. Running both chains for a model typically took 10 Central Processing Unit (CPU) minutes.

3.2.4 Model comparison and criticism

3.2.4.1 Leave-one-out cross-validation

We are fitting four models to the data and we need to assess which model best explains the data and critically examine the strengths and weakness of each of the models. To do this, we use the Bayesian Leave-One-Out cross-validation score (LOO). LOO is one method to estimate point-wise out-of-sample prediction accuracy of a given model (Vehtari et al., 2017). LOO is calculated for a given model by fitting the model to the data set where one of the data points has been left out. The posterior samples of that fit are then projected through the model likelihood to assess how well the left-out data point is predicted by the model. The procedure is then repeated so each data point is left out in turn. For a given model, this provides a per data point score which can be totalled over the data to give an indication of overall model performance, or compared data point-wise with a different model allowing an interpretable comparison between models.

Specifically, for a model \mathcal{M} , which being fit to the full data set \mathcal{D} , the LOO score for the i th data point, $D_i = \{t_i, X_i, Y_i\}$, is

$$\text{LOO}_{i,\mathcal{M}} = \log p(D_i | \mathcal{D}_{-i}, \mathcal{M}). \quad (53)$$

Here \mathcal{D}_{-i} is the full data set \mathcal{D} with the i th data point or D_i removed. This is the log of the LOO predictive density conditioned on the data set without the i th data point. It can be written in terms of expected

value of the likelihood of the left-out data point over the posterior distribution obtained while fitting the model to \mathcal{D}_{-i} ,

$$p(D_i|\mathcal{D}_{-i}, \mathcal{M}) = \int p(D_i|\vec{\theta}, \mathcal{D}_{-i}, \mathcal{M})p(\vec{\theta}|\mathcal{D}_{-i}, \mathcal{M})d\vec{\theta}. \quad (54)$$

Practically, if we have S samples from the posterior distribution of the model parameters $\{\vec{\theta}_{-i}^s\}_{s=1}^S$, obtained by fitting model \mathcal{M} to the data set \mathcal{D}_{-i} , we can use these samples to compute Eq. (54) as,

$$p(D_i|\mathcal{D}_{-i}, \mathcal{M}) = \frac{1}{S} \sum_{s=1}^S p(D_i|\mathcal{D}_{-i}, \vec{\theta}_{-i}^s, \mathcal{M}). \quad (55)$$

Here, we have assumed that we have a sufficient number of samples to fully capture the posterior distribution (Gelman et al., 2014).

This quantity can be totaled over the data to give an overall score for model \mathcal{M} ,

$$\text{LOO}_{\mathcal{M}} = \sum_{i=1}^N \text{LOO}_{i,\mathcal{M}}. \quad (56)$$

Alternatively the difference can be used to compare the performance of two model \mathcal{M}_1 and \mathcal{M}_2 ,

$$\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2} = \text{LOO}_{\mathcal{M}_1} - \text{LOO}_{\mathcal{M}_2}. \quad (57)$$

Here, a positive value indicates that \mathcal{M}_1 has a higher computed out-of-sample predictive accuracy than \mathcal{M}_2 . The standard error on this difference is given by Vehtari et al. (2017) as,

$$\text{se}(\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}) = \sqrt{NV_{i=1}^N (\text{LOO}_{i,\mathcal{M}_1} - \text{LOO}_{i,\mathcal{M}_2})}. \quad (58)$$

$V_{i=1}^N$ is the variance of the point-wise difference over the full data set of N data points. This can be used to assess how significant the difference between two models is. For brevity, we define the significance of the difference as $\text{sig}(\bullet) = |\bullet|/\text{se}(\bullet)$ where $\bullet = \text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}$.

In the case of this analysis, it is also informative to sum over the point-wise LOO score over the data in a single epoch, e . This is because the data is tightly temporally clustered within an epoch, and in the correlated noise models ($\mathcal{M}_{DC}, \mathcal{M}_{NC}$), data within an epoch share correlated noise properties. We define the difference in LOO predictive accuracy over an epoch e , as $\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}^e$. This quantity is analogous to $\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}$ (defined in Eqs. 56 and 57), but instead of

summing the point-wise score over all data points, the sum is taken only over the data in epoch, e . The standard error, $\text{se}(\Delta\text{LOO}_{\mathcal{M}_1, \mathcal{M}_2}^e)$, is also calculated analogously by instead calculating the variance of the point-wise difference of data within the epoch. In this case, N in Eq. 58 is the total number of data points within epoch e . Overall, computation of $\Delta\text{LOO}_{\mathcal{M}_1, \mathcal{M}_2}^e$ will permit the comparison of models at epoch resolution.

Calculating all $\text{LOO}_{i, \mathcal{M}}$ terms for all models is computationally expensive. This is because it would require N full refits (obtaining samples from the posterior distribution) of the model with each data point left out in turn. In our case, $N = 81$ and a single refit of a model takes ≈ 10 CPU minutes, therefore computing all $\text{LOO}_{i, \mathcal{M}}$ terms for the four considered models would take ≈ 55 CPU hours of computation. While not completely unfeasible, the required computation is still significant. We therefore turn to an importance sampling approximation to compute the $\text{LOO}_{i, \mathcal{M}}$ terms for each model.

We use the Pareto Smoothed Importance Sampling (PSIS; Vehtari et al., 2015) approximation to compute the $\text{LOO}_{i, \mathcal{M}}$ terms for each model (Vehtari et al., 2017; Bürkner et al., 2020), implemented in the Arviz Python package (Kumar et al., 2019). Instead of refitting the model with each data point left out in turn, in the PSIS approximation the model is initially fit to the full data set. Then posterior samples from the full data set fit are re-weighted (via importance sampling) to approximate the effect of removing each data point in turn. Overall PSIS allows the fast and approximate computation of the LOO terms for a model with very few refits. Appendix A.1 contains the application details of the PSIS approximation along with checks of the approximation accuracy for the models considered in this work.

3.2.4.2 *The case for LOO over other comparison metrics*

LOO is just one of many metrics that can be computed to assess the performance of a model. In this section, I briefly justify the decision to use LOO as a model criticism and comparison tool compared to the two commonly used approaches in astronomy: a reduced $\Delta\chi^2$ approach, and use of the Bayesian model evidence.

Typically in microlensing event analyses, albeit in analyses of photometric microlensing events, a reduced $\Delta\chi^2$ approach is used to select between competing models (e.g. Bond et al., 2004; Smith et al.,

2005; Alcock et al., 2000b; Bennett et al., 2018).⁷ There is a multitude of reasons why we do not use it here and choose LOO instead. Firstly, because some of the models considered in this work contain Gaussian-correlated noise components, reduced $\Delta\chi^2$ is no longer fully descriptive of the likelihood of the model (reduced $\Delta\chi^2$ is only related to the likelihood of an uncorrelated Gaussian likelihood with a diagonal covariance matrix). Secondly, reduced $\Delta\chi^2$ is only valid for a model that is linear in its parameters; none of the models considered in this Chapter are linear. Thirdly, reduced $\Delta\chi^2$ fails to account for any posterior uncertainty on the parameters. Andrae et al. (2010) gives an extensive account of the pitfalls of using reduced $\Delta\chi^2$ for comparison of non-linear models.

Comparison of models using the Bayesian evidence (Eq. 44) is becoming popular in astrophysics due to nested sampling algorithms that readily allow its computation (e.g. Skilling et al., 2006; Higson et al., 2019; Speagle, 2020). The Bayesian evidence has the appealing properties of fully capturing parameter uncertainty and naturally penalizes more complex models that do not significantly explain the data better. The critical downside, however, is that the evidence is sensitive to the choice of prior distribution (see e.g. Fong & Holmes, 2020). This becomes a problem when comparing models possessing parameters with uninformative and somewhat arbitrarily set prior distributions (see e.g. Section 7.2 of Gelman et al., 2013). For example, for Θ_E in this analysis, the arbitrary choice of a large width for its uninformative prior can arbitrarily change the model evidence without any resulting change of the posterior distribution.

Comparatively, LOO is not sensitive to the model priors. This is due to each $\text{LOO}_{i,\mathcal{M}}$ term being computed when the model is conditioned on the rest of data (Eq. 54), so the prior is always overwhelmed by the likelihood and has little effect. Moreover, the Bayesian evidence only provides a single summary statistic for the whole model and data, shedding little light on precisely where a model fails, whereas LOO provides an interpretable per data point score.

⁷ See Eq. (99) in Appendix A.2 for an example of a χ^2 optimization for the models considered in this work.

Parameter	Model			
	\mathcal{M}_{DC}	\mathcal{M}_{DW}	\mathcal{M}_{NC}	\mathcal{M}_{NW}
$X_{0,S}$ mas	35186.997 $^{0.127}_{-0.128}$	35187.027 $^{0.124}_{-0.123}$	35186.943 $^{0.126}_{-0.125}$	35186.946 $^{0.122}_{-0.125}$
$Y_{0,S}$ mas	45709.705 $^{0.125}_{-0.124}$	45709.675 $^{0.121}_{-0.122}$	45709.726 $^{0.125}_{-0.124}$	45709.741 $^{0.123}_{-0.122}$
$\mu_{X,S}$ mas/year	8.964 $^{0.044}_{-0.043}$	8.992 $^{0.040}_{-0.040}$	9.011 $^{0.045}_{-0.045}$	9.042 $^{0.045}_{-0.044}$
$\mu_{Y,S}$ mas/year	0.028 $^{0.047}_{-0.047}$	-0.012 $^{0.046}_{-0.047}$	0.159 $^{0.046}_{-0.046}$	0.289 $^{0.043}_{-0.044}$
ω_S mas	0.123 $^{0.110}_{-0.111}$	0.043 $^{0.101}_{-0.098}$	0.127 $^{0.121}_{-0.121}$	0.070 $^{0.115}_{-0.114}$
σ_{white} mas	0.838 $^{0.044}_{-0.041}$	0.993 $^{0.046}_{-0.044}$	0.882 $^{0.046}_{-0.045}$	1.225 $^{0.051}_{-0.048}$
$X_{0,L}$ mas	45825.738 $^{0.014}_{-0.014}$	45825.738 $^{0.013}_{-0.014}$	-	-
$Y_{0,L}$ mas	46592.483 $^{0.015}_{-0.015}$	46592.483 $^{0.015}_{-0.015}$	-	-
$\mu_{X,L}$ mas/year	-2661.640 $^{0.018}_{-0.018}$	-2661.640 $^{0.018}_{-0.018}$	-	-
$\mu_{Y,L}$ mas/year	-344.932 $^{0.019}_{-0.019}$	-344.932 $^{0.020}_{-0.019}$	-	-
ω_L mas	215.675 $^{0.018}_{-0.018}$	215.676 $^{0.018}_{-0.018}$	-	-
Θ_E mas	31.353 $^{2.077}_{-2.184}$	34.164 $^{1.386}_{-1.440}$	-	-
$\sigma_{1,corr}$ mas	0.081 $^{0.077}_{-0.055}$	-	0.084 $^{0.083}_{-0.057}$	-
$\sigma_{2,corr}$ mas	0.127 $^{0.090}_{-0.075}$	-	0.206 $^{0.097}_{-0.103}$	-
$\sigma_{3,corr}$ mas	0.597 $^{0.097}_{-0.099}$	-	0.627 $^{0.096}_{-0.094}$	-
$\sigma_{5,corr}$ mas	0.620 $^{0.096}_{-0.093}$	-	0.741 $^{0.082}_{-0.080}$	-
$\sigma_{6,corr}$ mas	0.380 $^{0.098}_{-0.099}$	-	0.477 $^{0.087}_{-0.083}$	-
$\sigma_{7,corr}$ mas	0.681 $^{0.087}_{-0.085}$	-	0.719 $^{0.083}_{-0.081}$	-
$\sigma_{8,corr}$ mas	0.078 $^{0.077}_{-0.053}$	-	0.120 $^{0.094}_{-0.078}$	-
$\sigma_{9,corr}$ mas	0.085 $^{0.078}_{-0.057}$	-	0.105 $^{0.090}_{-0.069}$	-

Table 5: Parameter posterior summaries for each model. Values are the posterior medians, uncertainties are the 84th-50th and 50th-16th posterior percentiles. ‘-’ indicates that the considered model does not contain the parameter.

3.2.5 Results

3.2.5.1 Performance of the models

All models (\mathcal{M}_{DC} , \mathcal{M}_{DW} , \mathcal{M}_{NC} , \mathcal{M}_{NW}) were fitted to the data. Posterior parameter summaries for all parameters can be found in Table 5, and posterior projections of the model over the data are shown in Fig. 22. Additionally, LOO scores were computed for all models.

Fig. 23 shows the pairwise comparison between all model combinations. \mathcal{M}_{DC} (deflection and correlated noise model) has the best overall LOO score when compared to all other models. \mathcal{M}_{NW} is the least preferred model with all other models having a higher score. The model most competitive with \mathcal{M}_{DC} , is \mathcal{M}_{NC} with $\Delta\text{LOO}_{\mathcal{M}_{DC},\mathcal{M}_{NC}} = 8.58$ and $\text{sig}(\Delta\text{LOO}_{\mathcal{M}_{DC},\mathcal{M}_{NC}}) = 1.48$. This suggests that the correlated

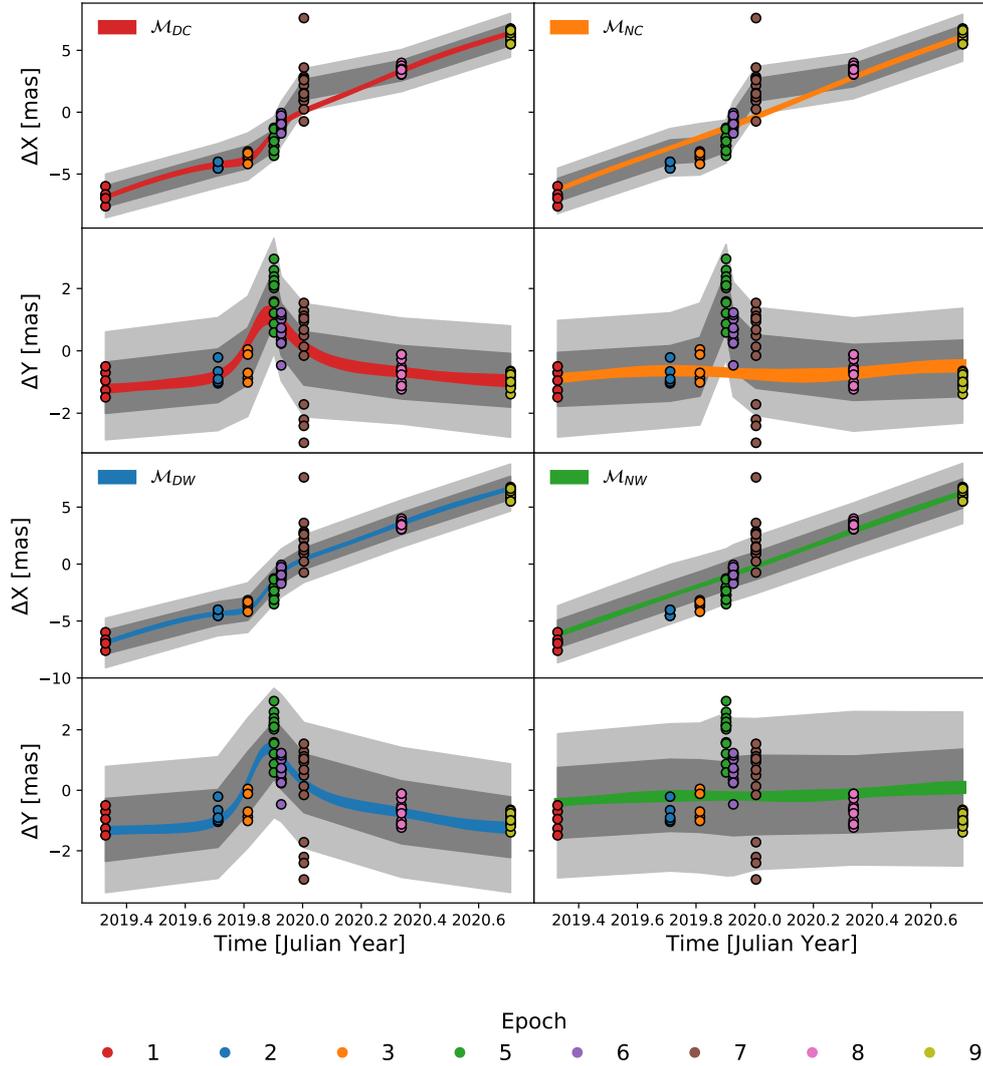


Figure 22: Posterior realizations plotted over the data in the X and Y directions and for each of the considered models. Coloured bands show the 84th-16th posterior percentiles on the inferred source trajectory for each model and direction. Dark (light) grey bands are the posterior data realizations 16th-84th (2nd-98th) percentile bands. Specifically this includes the trajectory and the noise model component realizations. The posterior data realizations are discontinuous between epoch because the noise model is only defined within an epoch and on the data grid.

component of the noise is an important component of the model because the other deflection model with just white noise (\mathcal{M}_{DW}), is

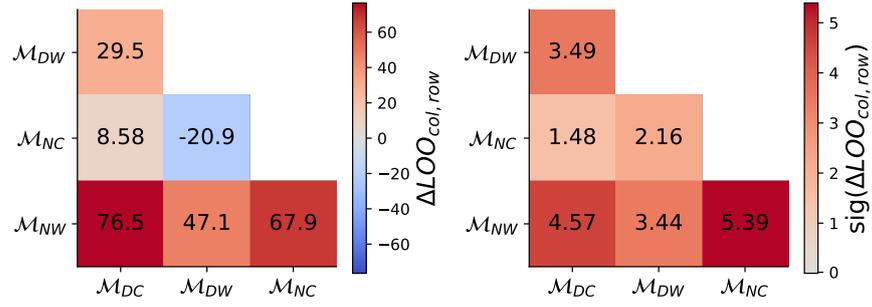


Figure 23: **Left:** Pairwise difference in LOO scores for all models considered. Red or a positive number means the model in the column is preferred. Blue or a negative number means the model in the row is preferred. **Right:** Significance of the scores in the left panel. Specifically, each score in the left panel divided by its standard error estimate. A higher number means the difference is more significant.

comparably not competitive with \mathcal{M}_{NC} . In fact, Fig. 23 shows that \mathcal{M}_{NC} is preferred over \mathcal{M}_{DW} with $\Delta LOO_{\mathcal{M}_{NC}, \mathcal{M}_{DW}} = -20.9$ and $sig(\Delta LOO_{\mathcal{M}_{NC}, \mathcal{M}_{DW}}) = 2.16$. That is, the non-deflection correlated noise model is preferred over the model with the deflection and just white noise. While \mathcal{M}_{DC} is the overall preferred model, it is informative to understand how exactly \mathcal{M}_{NC} is able to explain the data with no deflection term and even beat the deflection model with just white noise, \mathcal{M}_{DW} .

The starting point for understanding the good performance of \mathcal{M}_{NC} is to examine the per-epoch LOO scores. Fig. 24 shows the per-epoch LOO scores for \mathcal{M}_{DW} and \mathcal{M}_{DC} compared to \mathcal{M}_{NC} . For the comparison of the two correlated noise models (\mathcal{M}_{DC} versus \mathcal{M}_{NC}), it is shown that \mathcal{M}_{DC} marginally beats \mathcal{M}_{NC} in every epoch with the exception of epoch 2 where \mathcal{M}_{DC} is more clearly preferred and epoch 7 where \mathcal{M}_{NC} beats \mathcal{M}_{DC} . The reason why \mathcal{M}_{NC} can explain the epoch with the largest deflection terms (epochs 5, 6, and 7), is it inflates the correlated noise and alters the unlensed source trajectory.

Fig. 25 shows the priors and posterior distributions of the noise parameter in all of the models. For the correlated noise parameters, \mathcal{M}_{NC} inflates the size of the noise parameters relative to \mathcal{M}_{DC} and the prior for epochs 5, 6 and 7. \mathcal{M}_{NC} does this to try and explain the deflection signal. For epoch 2, the correlated noise term in \mathcal{M}_{NC} is slightly inflated compared to \mathcal{M}_{DC} and the prior. At first glance, this seems counterintuitive because the deflection at epoch 2 is small. This begs the question as to what signal \mathcal{M}_{NC} is trying to explain away with increased correlated noise. The reason for this is that, al-

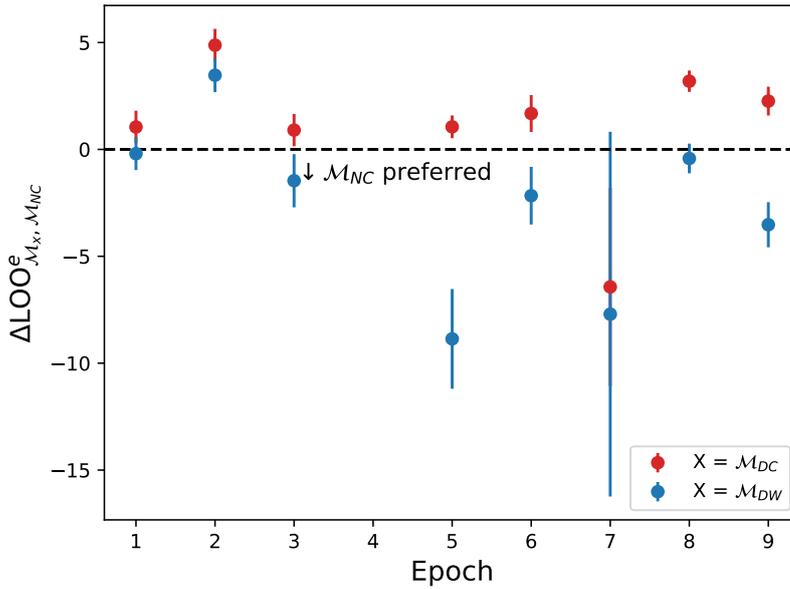


Figure 24: LOO score of models \mathcal{M}_{DC} and \mathcal{M}_{DW} compared to \mathcal{M}_{NC} within each epoch. Error bars are one standard error.

though the signal at epoch 2 is small, so is the estimated magnitude of the correlated noise, and so epoch 2 has one of the highest signal-to-noise ratios of all the epochs. Overall, this suggests correlated noise can mimic the deflection signal. It is noted that in epoch 7 and for the white noise both \mathcal{M}_{NC} and \mathcal{M}_{DC} inflate the noise terms relative to the prior. This suggests that the priors underestimate both of these quantities. It is also noted that for models missing either the correlated noise or the deflection signal, the white noise is increased to compensate relative to \mathcal{M}_{DC} and the prior.

Fig. 26 shows the GEDR3 priors and posterior distributions of the source astrometric parameters. While there is broad agreement between all models for the source proper motion in the X direction and the source parallax, \mathcal{M}_{NC} and \mathcal{M}_{DW} disagree on the Y direction proper motion posterior. \mathcal{M}_{DC} infers $\mu_{Y,S} = 0.028^{+0.047}_{-0.047}$ mas/year, whereas \mathcal{M}_{NC} infers $\mu_{Y,S} = 0.159^{+0.046}_{-0.046}$ mas/year. The relatively high value inferred by \mathcal{M}_{NC} is caused by \mathcal{M}_{NC} trying to explain away the positive Y direction deflection (see e.g. Fig. 22) by altering the source trajectory. This suggests that further data taken after the event to pin down $\mu_{Y,S}$ could completely rule out \mathcal{M}_{NC} as a viable model. Encouragingly, we note that for all models the lens and source astrometric parameters are consistent with the GEDR3 priors (see Figs. 61 and 60 in Appendix A.3).

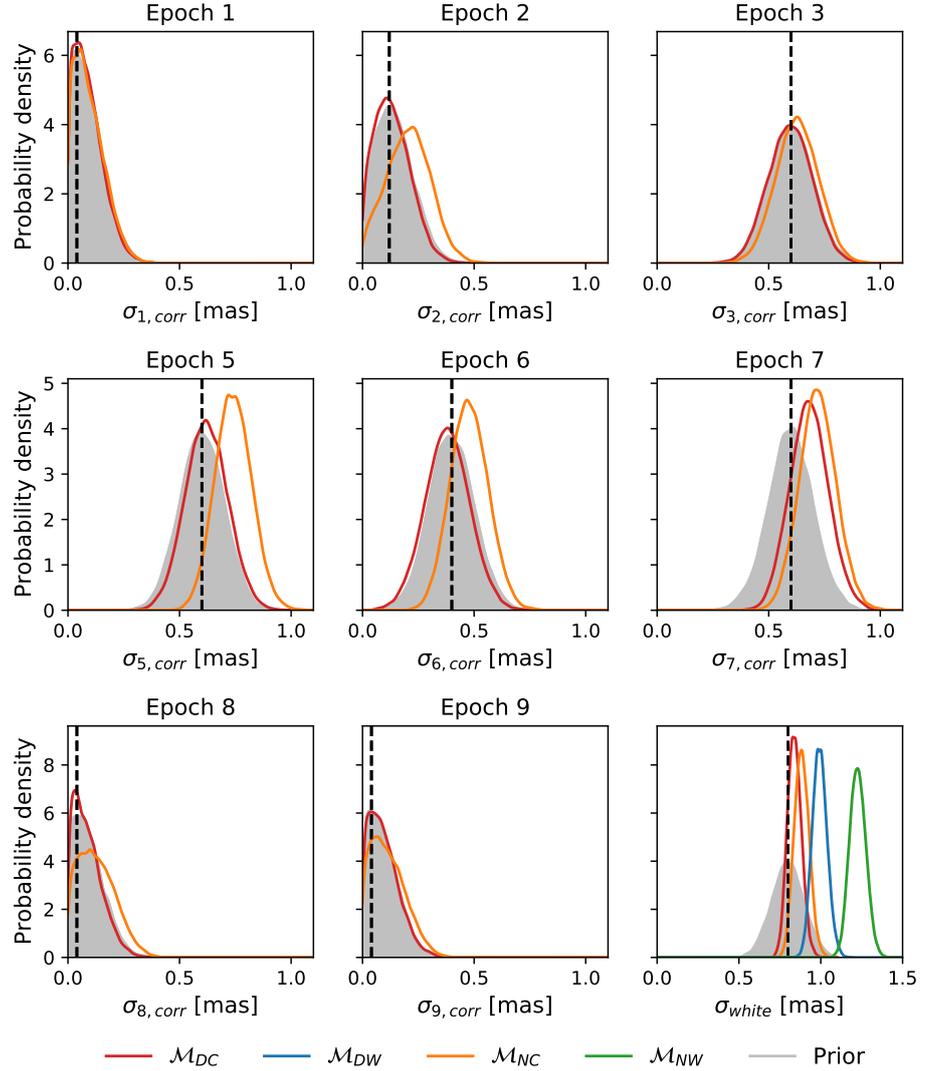


Figure 25: Prior and marginal posterior probability density functions for the noise parameters in the different models. Vertical dashed lines show the values of these parameters estimated from the PSF subtraction simulations (Table 2). The prior probability density function is shaded to aid differentiation with the posteriors. Note that the models with no correlated noise, \mathcal{M}_{DW} and \mathcal{M}_{NW} , do not have correlated noise parameters and therefore, only the posteriors on σ_{white} are shown.

The reason for the better performance of \mathcal{M}_{DW} compared with \mathcal{M}_{DC} in epoch 2 is a high signal-to-noise deflection (as mentioned earlier), combined with the inflexibility of the source trajectory to be altered to explain away an offset in the negative X direction. This is due to the asymmetrical deflection in the X direction being in the

negative (before the closest approach) then positive (after the closest approach) X direction. The source trajectory cannot be altered in \mathcal{M}_{NC} to account for both of these offsets, so \mathcal{M}_{NC} performs worse than \mathcal{M}_{DC} in epoch 2. The better performance of \mathcal{M}_{NC} compared to \mathcal{M}_{DW} in epoch 7 is due to the outlying data within that epoch (see Fig. 22). These outlying data are located at lower Y values which are further away from the deflection trajectory of \mathcal{M}_{DC} than the unlensed trajectory of \mathcal{M}_{NC} .

For the per-epoch comparison of \mathcal{M}_{NC} and \mathcal{M}_{DW} , Fig. 24 shows that \mathcal{M}_{NC} beats \mathcal{M}_{DW} in every epoch apart from epoch 2. The relative performance of \mathcal{M}_{NC} and \mathcal{M}_{DW} can be explained by the same reasoning used above, for the \mathcal{M}_{DC} and \mathcal{M}_{NC} epoch comparison. In epoch 5, Fig. 24 shows \mathcal{M}_{DW} clearly performs worse than \mathcal{M}_{NC} , despite there being a large deflection signal in epoch 5. This is due to epoch 5 also having a large correlated noise estimate ($m_{\sigma_{e,\text{corr}}} = 0.6$ mas) which white noise \mathcal{M}_{DW} cannot explain away, even by inflating the size of the white noise (see Fig. 25).

3.2.5.2 Inference on the angular Einstein radius

Both models including the astrometric lensing deflection signal (\mathcal{M}_{DW} and \mathcal{M}_{DC}), provide a posterior inference on Θ_{E} . Fig. 27 shows the Θ_{E} marginal posterior distribution for both the \mathcal{M}_{DW} and \mathcal{M}_{DC} models, along with the prior used in both models. For \mathcal{M}_{DW} and \mathcal{M}_{DC} , the inferred values are $\Theta_{\text{E}} = 34.2_{-1.4}^{+1.4}$ mas and $\Theta_{\text{E}} = 31.4_{-2.2}^{+2.1}$ mas, respectively. Here the values and upper and lower bars represent the 50th, 84th-50th, and 16th-50th posterior percentiles, respectively. Fig. 27 shows that \mathcal{M}_{DW} provides a tighter constraint and slighter higher value of Θ_{E} compared with \mathcal{M}_{DC} . It is also shown that the \mathcal{M}_{NW} Θ_{E} posterior distribution is slightly asymmetrical with more probability mass in the left hand tail (towards lower Θ_{E} values).

The difference in the Θ_{E} posteriors between \mathcal{M}_{DW} and \mathcal{M}_{DC} is consistent with the findings in Section 3.2.4. In Section 3.2.4, it was shown that the correlated noise part of the model (with help from a positive $\mu_{Y,S}$) can mimic or explain away some of the microlensing signal. Therefore, jointly fitting both the correlated noise and the microlensing signal means the observed offset on position explanation is shared amongst the microlensing signal and correlated noise, causing a slightly lower posterior median Θ_{E} (as $\delta_+ \propto \Theta_{\text{E}}$) for \mathcal{M}_{DC} compared with \mathcal{M}_{DW} . Conversely, because the white noise model in \mathcal{M}_{DW} cannot explain correlated noise at each epoch, so it inflates the

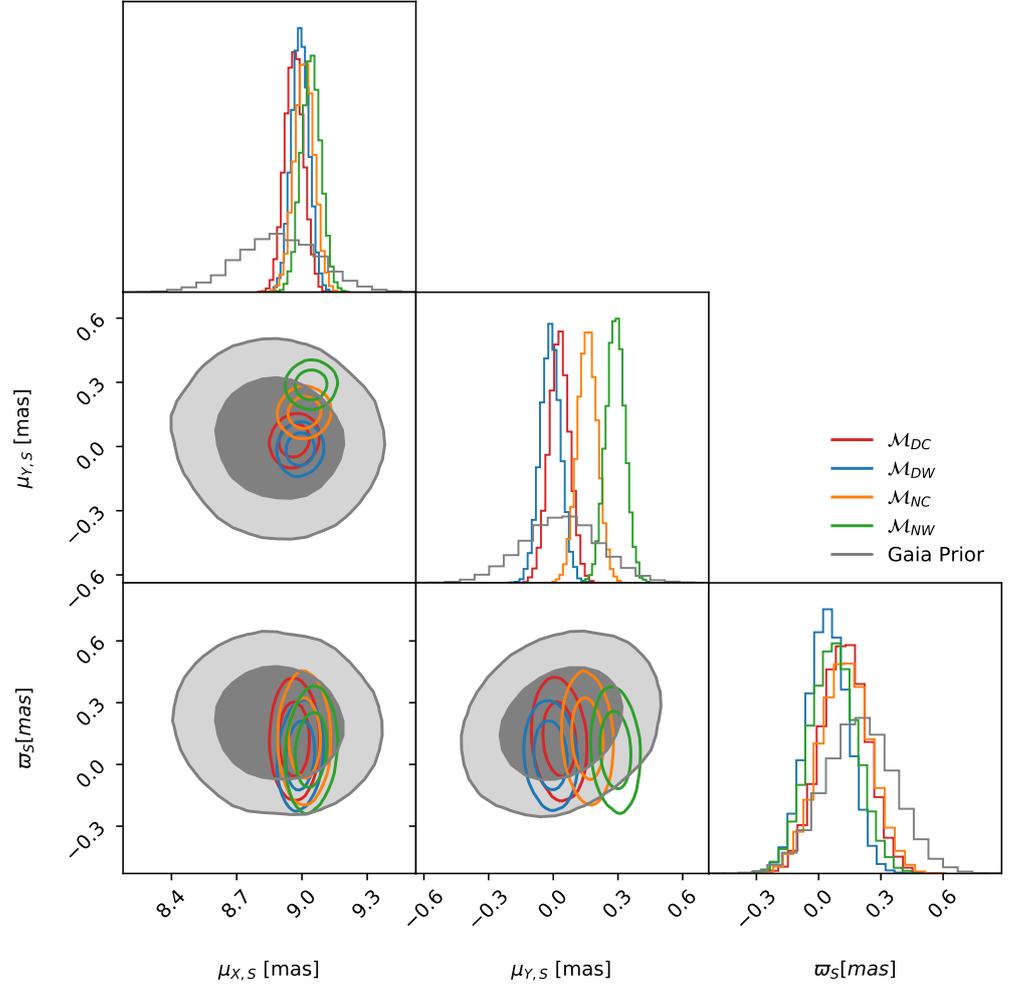


Figure 26: Comparison of the source GEDR3 astrometry used as a prior in all of the models compared with the posterior on the astrometry from each of the models. All panels show a probability density. For the 2D plots, the inner and outer contours contain 68% and 95% of the probability mass, respectively. The histograms show the marginal probability densities for each parameter. We have omitted the GEDR3 prior and model posteriors for the source reference positions $(X_{0,s}, Y_{0,s})$ because we found good agreement between GEDR3 priors and all the model posteriors (see Fig. 60 for the full corner plot).

values of Θ_E to compensate. The larger spread in the Θ_E posterior for \mathcal{M}_{DC} is likely similarly due to the fact the correlated noise can mimic the microlensing signal, suggesting some degeneracy between them, which ultimately leads to a less certain inference on Θ_E in \mathcal{M}_{DC} .

We now use the posterior samples to calculate an inferred value for M_L or LAWD 37's mass. Inverting Eq. (2), we can write the M_L in terms of Θ_E and $\varpi_L - \varpi_S$. Of course, we run into the same diffi-

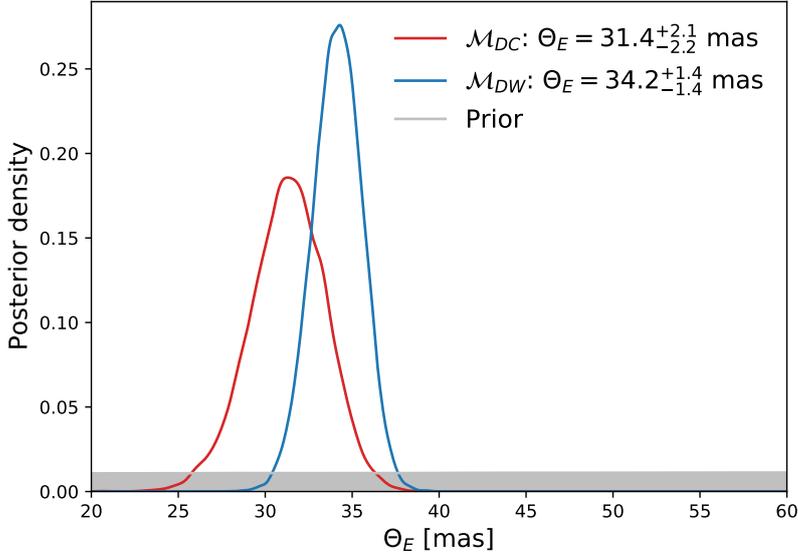


Figure 27: Posterior distribution on Θ_E for the deflection model with correlated noise \mathcal{M}_{DC} and the deflection model with white noise \mathcal{M}_{DW} . Reported Θ_E values are the 50th posterior percentile and the 84th-16th posterior percentile uncertainties. The prior used for Θ_E in both models is shown in grey.

culties described in Section 3.2.3.1 of negative ϖ_S values entering as distance terms in Θ_E . This is not a problem because the source is so distant, $\varpi_S \approx 0.1$ mas, compared with the lens, $\varpi_L \approx 215$ mas, we can approximate $\varpi_L - \varpi_S \approx \varpi_L$. Under this approximation we get M_L for the correlated noise model, $M_L = 0.56^{+0.08}_{-0.08} M_\odot$, and M_L for the white noise model, $M_L = 0.66^{+0.06}_{-0.05} M_\odot$. Here, we report the 50th, 84th-16th, and 50th-16th posterior percentiles. We note inclusion of ϖ_S both with negative posterior samples and with the negative posterior values truncated, does not change the reported mass values or uncertainties for either model.

3.2.5.3 Prior sensitivity

We can check how sensitive the posterior inference on Θ_E is to the prior assumptions in \mathcal{M}_{DC} . Specifically, we can test how tightening or relaxing the prior parameter distributions affects the Θ_E posterior distribution with the view of determining which of our assumptions strongly affect our inferences. For both the lens and source unlensed trajectory, we used informative Gaussian priors from GEDR3. The right panel of Fig. 28 shows how changing (inflating or shrinking by a multiplicative factor) the EDR3 prior covariance matrix on the source

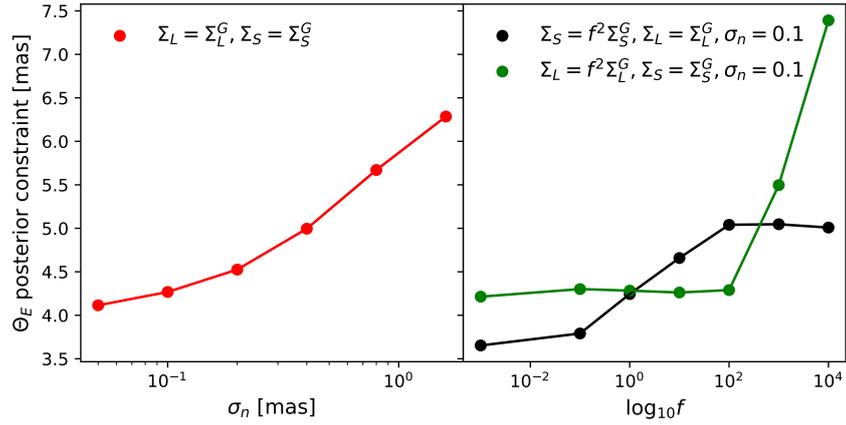


Figure 28: Sensitivity of the posterior constraint on Θ_E (16th-84th posterior percentile) to changes in the priors in the deflection and correlated noise model \mathcal{M}_{DC} . **Left:** Posterior constraint versus σ_n , the standard deviation on the Gaussian prior on each of the correlated noise and white noise parameters. In this case, the priors on the source and lens astrometry are fixed to the GEDR3 Σ_G^S and Σ_G^L , respectively. **Right:** Posterior constraint versus altering the source or lens astrometric Gaussian prior covariance matrix by a multiplicative factor f , whilst keeping all the other priors fixed.

(Σ_S^G) or lens (Σ_L^G), changes the Θ_E posterior constraint. Fig. 28 shows that the inference on Θ_E is insensitive to changing Σ_L^G . Specifically, the posterior constraint on Θ_E does not degrade until Σ_L^G is inflated by a factor of 1000^2 . Moreover, it is also shown that shrinking Σ_L^G (even shrinking by a factor of 1000^2) does not improve the posterior constraint on Θ_E . This suggests that further data on the lens position, either by future Gaia data releases or further HST monitoring, is unlikely to improve the constraint on Θ_E .

The right panel of Fig. 28 also shows the effect of changing the source unlensed trajectory prior covariance matrix Σ_S^G . It shows that the analysis is somewhat more sensitive to the prior on the source unlensed trajectory compared with the lens trajectory. Fig. 28 shows that inflating the prior by a factor of 10^2 immediately starts to degrade the inference on Θ_E , although the degradation does level off at an inflation factor of (100^2) . This is likely because at an inflation level beyond 100^2 , the source unlensed trajectory is completely determined by the HST data and the source unlensed trajectory prior becomes uninformative.

More importantly however, shrinking the prior on the unlensed source trajectory does improve the posterior constraint on Θ_E . Specif-

ically, an improvement of a factor of 10^2 in the prior covariance of the source unlensed trajectory would lead to a maximum improvement in the 16th-84th posterior constraint on Θ_E of ≈ 1 mas (corresponding to a 7% improvement on 84th-16th percentile constraint on M_L). While an improvement of a factor of 10^2 of the source astrometry covariance matrix is unlikely to be achieved, this suggests that further astrometric data pinning down the unlensing source trajectory either by future Gaia data releases or HST will likely improve the posterior constraint on Θ_E . Assuming a more realistic improvement of 2^2 on the source astrometry covariance matrix, corresponds to a 3% improvement on 16th-84th percentile constraint on M_L .

The left panel of Fig. 28 shows the effect of inflating or shrinking the standard deviation on the Gaussian prior for the noise parameters (σ_n), both correlated and white. The means of these priors were informed by lens PSF subtraction simulations and WFC3 instrument precision, in the case of the correlated and white noise terms, respectively. The standard deviations of these priors, (σ_n), was however chosen to be 0.1 mas. Fig. 28 shows that the analysis is sensitive to changing σ_n . Specifically, if σ_n is increased from 0.1 mas to 0.4 mas, the Θ_E 16th-84th posterior percentile degrades by ≈ 1 mas. In contrast, shrinking σ_n to 0.01 mas only marginally improves the Θ_E 16th-84th posterior percentile by ≈ 0.2 mas. This sensitivity is reflected in the posterior distributions of the correlated noise parameters shown in Fig. 25. This is because for the majority of epochs the size of the correlated noise is not constrained by the data and defaults the prior distribution, for the adopted value of $\sigma_n = 0.1$, indicating the priors are informative.

However, for epoch 7, the data does inform the value of the size of the correlated noise for \mathcal{M}_{DC} and the posterior is slightly shifted to higher values than the prior. This highlights the important balance between picking a reasonably tight prior to inform the size of the correlated noise but one that reflects reasonable uncertainty so the data can alter the value if there is constraining information. The adopted value of $\sigma_n = 0.1$ mas in the analysis strikes a reasonable balance between our confidence in the size of the correlated noise at each epoch, whilst also giving sufficient flexibility for the data to further inform the size of the noise. Overall, the posterior constraint on Θ_E is sensitive to our choice of $\sigma_n = 0.1$ mas and our prior knowledge on the correlated component of the noise in general.

3.2.6 Discussion and conclusions

We have analyzed HST follow-up data of a predicted astrometric microlensing event caused by the nearby white dwarf, LAWD 37. Specifically, we used HST astrometric data of the source in combination with Gaia astrometry (GEDR3) of the source and lens to infer a gravitational mass for LAWD 37. We consider and fit four different models to the data. Models with and without the astrometric deflection term, and then each with and without a correlated noise component due to the lens PSF subtraction within an epoch (\mathcal{M}_{DC} , \mathcal{M}_{DW} , \mathcal{M}_{NC} , \mathcal{M}_{NW}). We find the model with the deflection term and correlated noise (\mathcal{M}_{DC}) best explains the data according to the overall LOO score.

The model that provides the next best explanation for the data according to the LOO score is the model without an astrometric deflection but with correlated noise (\mathcal{M}_{NC}). This model is able to provide an explanation for the deflection signal in the Y direction by increasing the size of the correlated noise above the prior expectation and altering the source proper motion in the Y direction. Therefore, additional follow-up data on the source after the lensing event, which would further constrain the source proper motion, will likely definitively rule out this model. \mathcal{M}_{NC} , however, is unable to explain the asymmetric deflection signal in the X direction. This is most prominently seen in epoch 2 which has a large signal-to-noise deflection in the direction opposite to the source X proper motion direction.

The model with the deflection and just white noise, \mathcal{M}_{DW} , provides a comparatively poor explanation of the data according to the LOO score. \mathcal{M}_{DW} fails to explain the data in epoch 5 where there is a predicted high amount of correlated from the lens PSF subtractions. The failure in epoch 5 is clear, when compared to \mathcal{M}_{NC} , despite \mathcal{M}_{DW} increasing the size of the white noise and epoch 5 having a large deflection signal. Consequently, because the correlated noise can mimic the deflection signal, \mathcal{M}_{DW} interprets some of the unmodeled correlated noise as additional deflection signal and infers a high mass for LAWD 37 of $M_L = 0.66^{+0.06}_{-0.05} M_\odot$. Overall, we rule out this model, and its mass inference, due to its poor LOO score compared with \mathcal{M}_{DC} and \mathcal{M}_{NC} .

We performed checks on the sensitivity of our inferences in our chosen model, \mathcal{M}_{DC} , to our prior modeling assumptions. We find that, for \mathcal{M}_{DC} , our inference on M_L is not sensitive to the lens GEDR3 prior. This means that collecting additional follow-up astrometric data on

the lens (LAWD 37) to improve its astrometric solution is unlikely to improve the inference on M_L . In contrast, we find that, for \mathcal{M}_{DC} , the inference on M_L is sensitive to the GEDR3 prior on the source astrometry. Therefore, we conclude that collecting additional astrometric data to further constrain the source astrometric solution is likely to improve the inference on M_L by $\approx 3\%$. For the noise parameters in \mathcal{M}_{DC} , both correlated and white, we found our inference on M_L is sensitive to their chosen prior distributions. We found that simultaneously tightening the prior distributions on all the noise parameters does not significantly improve the constraint on M_L .

Overall, we find that the model with the microlensing deflection signal and correlated noise due to the lens PSF subtraction (\mathcal{M}_{DC}) provides the best explanation of the data. \mathcal{M}_{DC} provides an inference of the mass of LAWD 37 of $M_L = 0.56^{+0.08}_{-0.08} M_\odot$. A no deflection and correlated noise model, \mathcal{M}_{NC} , provides a worse but competitive explanation of the data with \mathcal{M}_{DC} . However, \mathcal{M}_{NC} , has to increase the size of the correlated noise above our PSF simulation expectations to explain the data in many epochs. Moreover, \mathcal{M}_{NC} cannot explain the high signal-to-noise negative X direction offset in epoch 2 which is well explained by \mathcal{M}_{DC} . We therefore conclude the astrometric deflection signal is present and we measure LAWD 37's mass to be $M_L = 0.56^{+0.08}_{-0.08} M_\odot$.

Appendix A.2 contains a comparison of my fitting method to obtain Θ_E with other methods used in the literature (e.g. Sahu et al., 2017; Zurlo et al., 2018).

3.3 THE ASTROPHYSICS OF LAWD 37

At a distance of 4.6 pc, LAWD 37 is the second nearest single white dwarf to the sun after van Maanen's Star. Direct imaging of the region around LAWD 37 with HST shows no evidence of visible companions down to detection limits (Schroeder et al., 2000). Comparing Hipparcos and GDR2 astrometry, Kervella et al. (2019) identified a proper motion anomaly for LAWD 37 which could be explained by a massive companion. However, upon comparison between Hipparcos and the more precise GEDR3 astrometry for LAWD 37, this has largely been ruled out (Lindgren & Dravins, 2021). Overall, at the present time it appears that LAWD 37 has no detectable companions.

Being a close by and bright target (V-band ≈ 11.50), LAWD 37 has been the subject of numerous studies (e.g. Koester & Weidemann,

1982; Weidemann & Koester, 1995; Dufour et al., 2005; Bergeron et al., 2001; Holberg et al., 2008; Subasavage et al., 2009; Giammichele et al., 2012; Sion et al., 2009; Coutu et al., 2019). As a result, a wealth of photometric and spectroscopic information has been gathered on LAWD 37. LAWD 37 has a spectral type DQ indicating a helium dominated atmosphere and the presence of carbon lines in its spectrum. The presence of carbon in DQ white dwarf atmospheres is well explained by models of carbon diffusing upwards from the core and being brought to the photosphere by convection (Pelletier et al., 1986; Dufour et al., 2005). In line with all helium rich white dwarfs LAWD 37 is expected to have a thin hydrogen layer, $q_{\text{H}} = 10^{-10}$ (Dufour et al., 2005).

Fitting⁸ DQ-type white dwarf atmospheric models (Blouin et al., 2018; Blouin & Dufour, 2019) to the broad-band photometry (V,R,I,J,H, and Ks), and spectroscopy of LAWD 37 (Kilic et al., 2020), its inferred atmospheric parameters are, $T_{\text{eff}} = 7836 \pm 185\text{K}$, $\log g = 7.98 \pm 0.03$, and $\log[\text{C}/\text{He}] = -5.61$. This fitting also permits the solid angle of LAWD 37 to be determined which in turn allows the radius of LAWD 37 to be calculated directly using distance (parallax) information from GEDR3. This implies the radius of LAWD 37 to be $R = 0.0127 \pm 0.0003R_{\odot}$. Assuming that LAWD 37 follows the standard evolutionary theory for CO core white dwarfs (Bédard et al., 2020) with a thin hydrogen layer ($q_{\text{H}} = 10^{-10}$) its radius corresponds to a mass of $0.57 \pm 0.02M_{\odot}$.

The direct gravitational mass determination from the astrometric microlensing caused by LAWD 37 is completely independent of all atmospheric and evolutionary modeling assumptions. This gravitational mass can therefore be used to test the theoretical models. Fig. 29 shows position of LAWD 37 on the theoretical MRRs, obtained from the Montreal theoretical cooling tracks⁹, using the gravitational mass from the astrometric microlensing event. Fig. 29 shows excellent agreement between the gravitational microlensing mass and predicted value from the evolutionary theory of CO core white dwarfs, $0.57 \pm 0.02M_{\odot}$, and assuming a thin hydrogen layer. Also plotted in Fig. 29 is the MRR for CO core white dwarfs with a thick hydrogen layer. While the microlensing mass is most consistent with the thin hydrogen layer expected to be possessed by LAWD 37,

⁸ Pierre Bergeron performed the atmospheric models fits and provided the resulting parameters and uncertainties.

⁹ Obtained from interpolation of the Montreal theoretical cooling tracks available at <https://www.astro.umontreal.ca/~bergeron/CoolingModels/>.

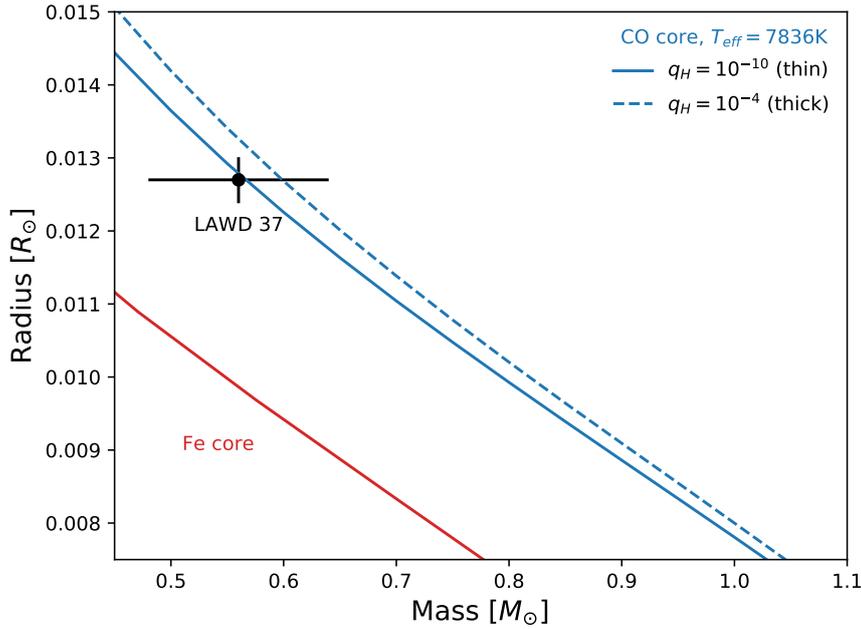


Figure 29: Comparison of the mass of LAWD 37 inferred from the astrometric lensing event with theoretical MRR relationships for white dwarfs. Blue lines show the MRR for CO core white dwarfs with a thin ($q_{\text{H}} = 10^{-10}$), and thick ($q_{\text{H}} = 10^{-4}$) hydrogen layer, and effective temperature equal to that of LAWD 37. For comparison, also shown is the theoretical MRR for a zero temperature white dwarf with an iron (Fe) core (Hamada & Salpeter, 1961).

there is not sufficient mass precision to rule out the thick hydrogen layer. For comparison, also shown in Fig. 29, is the theoretical MRR for zero temperature white dwarfs with an iron (Fe) core (Hamada & Salpeter, 1961), which is definitively ruled out by the microlensing mass.

Fig. 30 shows position of LAWD 37 on the theoretical Hertzsprung-Russell diagram (luminosity versus effective temperature) along with theoretical cooling tracks (assuming $q_{\text{H}} = 10^{-10}$) from the Montreal database for a range of masses. The position of LAWD 37 in the Hertzsprung-Russell diagram is in excellent agreement with the expected position of a white dwarf with the inferred microlensing mass of $0.56^{+0.08}_{-0.08} M_{\odot}$. Also shown in Fig. 30 are isochrones for 1.0, 1.2, and 1.5×10^9 years. By interpolation of the Montreal theoretical cooling tracks, the implied cooling age of LAWD 37 is $1.15^{+0.09}_{-0.06} \times 10^9$ years.

In conclusion, the gravitational mass for LAWD 37 obtained via astrometric microlensing in this Chapter is in excellent agreement with theoretical MMR and cooling tracks expected from the evolutionary theory of CO core white dwarfs. While the mass of LAWD 37 is most

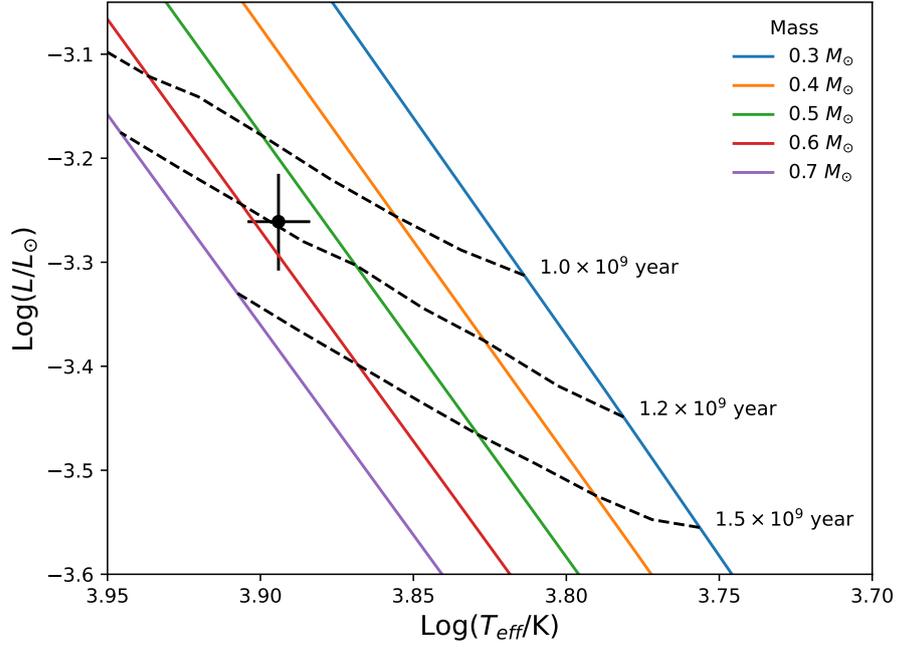


Figure 30: Hertzsprung-Russell diagram for LAWD 37. Coloured lines show the Montreal cooling tracks for a selection of white dwarf masses. Dashed lines indicate isochrones. The position of LAWD 37 agrees with the inferred microlensing mass of $0.56 \pm 0.08 M_{\odot}$. The implied age of LAWD 37 is $1.15^{+0.09}_{-0.06} \times 10^9$ years.

consistent with a thin hydrogen layer ($q_{\text{H}} = 10^{-10}$) expected for objects with helium-rich atmospheres like LAWD 37, the microlensing mass precision is currently insufficient to rule out the presence of a thick hydrogen layer ($q_{\text{H}} = 10^{-4}$). This work provides the first ever semi-empirical test of the white dwarf MRR for an isolated white dwarf and lends support to current white dwarf evolutionary theory. This work also marks only the third time that the astrometric microlensing effect has been detected via the prediction channel.

MICROLENS MASS DETERMINATION FOR GAIA'S PREDICTED EVENTS

Based on work originally published in McGill et al. (2019a).¹

We used GDR2 to search for upcoming photometric microlensing events, identifying two candidates with high amplification. In the case of candidate 1, a spectrum of the lens confirms it is a usdM3 subdwarf with mass $\approx 0.11M_{\odot}$ while the event reaches maximum amplification of 20_{-10}^{+20} mmag on 3rd November 2019 ($\pm 1d$). For candidate 2, the lens is a metal-poor M dwarf with mass $\approx 0.38M_{\odot}$ derived from spectral energy distribution fitting, and maximum amplification of 10_{-10}^{+40} mmag occurs on 3rd June 2019 ($\pm 4d$). This permits the use of a new algorithm for mass inference on the microlens. Given the predicted time, the photometric light curve of these events can be sampled by targeted photometric campaigns. The light curve is a function of the unknown lens mass together with eight other parameters for all of which Gaia provides measurements and uncertainties. Leveraging this prior information on the source and lens provided by Gaia's astrometric solution, combined with a photometric campaign, we show that the mass of the lenses in candidate event 1 and 2 are likely to be inferred with 20 – 27 and 30 – 42 % precision, respectively. We show that for both candidate events, the precision at which the lens masses can be inferred will likely be limited by the photometric measurements and not by the prior information provided by GDR2.

4.1 INTRODUCTION

The advent of data from the Gaia satellite has caused a flurry of interest in the prediction of future microlensing events whether astrometric or photometric. Given precise stellar positions, parallaxes, and

¹ This work was carried out in collaboration with Leigh Smith (LS), Wyn Evans (WE), Vasily Belokurov (VB) and Z. H. Zhang (ZZ). The original idea for the project was mine and I carried out the full search and Bayesian analysis presented in this Chapter. ZZ obtained and reduced the spectrum for lens of candidate 1 events and produced Figs. 32 and 33. ZZ also provided parts of the text relating to the reduction of the spectral data in section 4.4. LS, WE, and VB provided useful comments on early versions of this Chapter.

proper motions, it is a straightforward task to estimate whether a background object lies sufficiently close to the estimated Einstein radius of a foreground lens to produce a detectable lensing signal. The Gaia data releases have proved to be a treasure trove for finding such events. First, using data from GDR1, [McGill et al. \(2018\)](#) (Chapter 3), reported on a predicted astrometric microlensing event caused by the white dwarf LAWD 37 (WD 1142 – 645). Subsequent to GDR2 ([Gaia Collaboration et al., 2018](#)), two ongoing astrometric microlensing events were identified by [Klüter et al. \(2018a\)](#). Next a systematic search for microlensing events was carried out by [Bramich \(2018\)](#) (hereafter [B18](#)), who rediscovered the astrometric events caused by LAWD 37 and Stein 2051 B, reported earlier by [Sahu et al. \(2017\)](#) (hereafter [S17](#)). [B18](#) also identified nine further events which may exhibit detectable photometric and astrometric signatures. [Bramich & Nielsen \(2018\)](#) then extended [B18](#)'s work and presented an almanac of 2509 predicted microlensing events with closest approaches within the next century. Finally, [Klüter et al. \(2018b\)](#) (hereafter [K18b](#)) presented their systematic search with different astrometry and event detectability cuts to [B18](#). There have also been studies of predicted astrometric lensing events with pulsars ([Ofek, 2018](#)), and photometric events by nearby stars potentially hosting exoplanets ([Mustill et al., 2018](#)).

Here we discuss two photometric events found in our own search through GDR2 from a different viewpoint. It is well known that the photometric light curve of a microlensing event is degenerate and the mass of the lens cannot be extracted unless further information is available (e.g. [An et al., 2002](#)). By itself, the microlensing light curve for a point source lensed by a point mass provides a constraint only on the degenerate combination of mass and lens–source relative proper motion and parallax. However, if both lens and source are present in Gaia DR2 then there are measurements for all these quantities with errors. This suggests a new method for mass measurement of predicted photometric microlensing events with Gaia. If the light curve is densely sampled, as is possible for predicted events, then the extraction of the mass is a straightforward Bayesian inference problem using the astrometric quantities from GDR2.

This Chapter is organized as follows. We present our parametrization and model of the photometric microlensing signal, which permits a mass determination of lens by leveraging prior information from GDR2. We then identify suitable photometric candidates from

GDR2, and use Monte Carlo simulations to demonstrate the efficiency of our new method for measuring the masses of microlenses. Finally, we highlight further implications of our work.

4.2 A MODEL OF THE PHOTOMETRIC MICROLENSING SIGNAL

In order to model the photometric signal of a predictable microlensing event, we choose a parametrization of the light curve that naturally allows prior knowledge of astrometric quantities derived from GDR2 data on the source and lens to be used easily.

For the purpose of the parallax projection, we fix the reference position for both the lens and source at the reference position of the source, $(\alpha_{\text{refS}}, \delta_{\text{refS}})$. Using Eq. (39), allows us to write the angular separation of a source (S) and lens (L), at time t , in terms of their relative astrometric quantities,

$$\begin{aligned} \vec{\beta}(t) &\approx \begin{bmatrix} \alpha_{\text{refL}} - \alpha_{\text{refS}} \\ \delta_{\text{refL}} - \delta_{\text{refS}} \end{bmatrix} + [t - t_{\text{ref}}] \begin{bmatrix} [\mu_{\alpha^*L} - \mu_{\alpha^*S}] / \cos \delta_{\text{refS}} \\ \mu_{\delta L} - \mu_{\delta S} \end{bmatrix} \\ &+ (\varpi_L - \varpi_S) \mathbf{J}^{-1} \vec{\mathbf{R}}_{\oplus}(t) \\ &= \begin{bmatrix} \alpha_{\text{ref,rel}} \\ \delta_{\text{ref,rel}} \end{bmatrix} + [t - t_{\text{ref}}] \begin{bmatrix} \mu_{\alpha^*,\text{rel}} / \cos \delta_{\text{refS}} \\ \mu_{\delta^*,\text{rel}} \end{bmatrix} + \varpi_{\text{rel}} \mathbf{J}^{-1} \vec{\mathbf{R}}_{\oplus}(t). \end{aligned} \quad (59)$$

Here, \mathbf{J}^{-1} is projected at the source reference position $(\alpha_{\text{refS}}, \delta_{\text{refS}})$. It is now possible to write a model for the photometric microlensing signal parameterized by relative astrometric quantities and base lens and source fluxes that we have prior knowledge of from GDR2.

In the case of an unresolved luminous lens and source system, we may write the observed flux of the blend in a band and location denoted by the index k (assuming the only contaminating flux is from the lens) as,

$$F_{\text{obs}}^k(\mathbf{u}) = F_S^k \mathcal{A}(\mathbf{u}) + F_L^k, \quad (60)$$

where F_L^k is the flux of the lens, F_S^k is the unlensed flux of the source, and $\mathcal{A}(\mathbf{u}) = \mathcal{A}_+(\mathbf{u}) + \mathcal{A}_-(\mathbf{u})$ (Eq. 19) is the amplification from both the major and minor images. Explicitly rewriting Eq. (60) as a function of time with its explicit parameter dependencies, we obtain our model for the photometric signal at time t_i as,

$$F_{\text{obs}}^k(t_i; \vec{\theta}, F_S^k, F_L^k) = F_S^k \mathcal{A}(t_i; \vec{\theta}) + F_L^k + \epsilon_i, \quad (61)$$

where we assume ϵ_i is independent Gaussian noise. Here, $\vec{\theta}$ is the vector of parameters that control the amplification and hence the shape of the microlensing light curve and is,

$$\vec{\theta} = [\alpha_{\text{ref,rel}}, \delta_{\text{ref,rel}}, \mu_{\alpha^*,\text{rel}}, \mu_{\delta,\text{rel}}, \varpi_{\text{rel}}, M_L]. \quad (62)$$

We note that Eq. (61) does weakly depend on our choice of reference position $(\alpha_{\text{ref,S}}, \delta_{\text{ref,S}})$. However, as a simplifying assumption in the model, we assume that we know it exactly.

Our task is to fit the light-curve model outlined in Eq. (61) to a set of N observed photometric data points, $\mathcal{D}^k = \{t_i, f_i^k, \sigma_{f,i}\}_{i=1}^N$, where f_i^k is the observed flux of the lens-source blend at time t_i , with measurement variance $\sigma_{f,i}^2$. Assuming that noise terms ϵ_i are independent, we may write the likelihood function of the microlensing light curve as,

$$p(\mathcal{D}|\vec{\theta}, F_S^k, F_L^k) = \prod_{i=1}^N \mathcal{N} \left[f_i^k | F_{\text{obs}}^k \left(t_i; \vec{\theta}, F_S^k, F_L^k \right), \sigma_{f,i}^2 \right], \quad (63)$$

where $\mathcal{N}(x|\mu_x, \sigma_x^2)$ is the normal density in the random variable x with mean μ_x and variance σ_x^2 . Using Bayes theorem, we write the posterior distribution as,

$$p(\vec{\theta}, F_S^k, F_L^k | \mathcal{D}) \propto p(\mathcal{D}|\vec{\theta}, F_S^k, F_L^k) \times p(\vec{\theta}, F_S^k, F_L^k). \quad (64)$$

Here, $p(\vec{\theta}, F_S^k, F_L^k)$ denotes the prior on the model parameters. For all parameters apart from the mass of the lens M_L , if the lens and source are present in GDR2, Gaussian priors can be derived for these quantities. Finally, we place a uninformative flat uniform prior on M_L as our aim is to infer M_L from the data (see Table 6 for details on the priors). We note that, to avoid drawing unphysical samples during sampling from the posterior in Eq. (64) ($\varpi_{\text{rel}}, M_L < 0$), we frame the model in terms of $\ln \varpi_{\text{rel}}, \ln M_L$, and transform the priors in Table 6 accordingly. Overall, using the light curve for a predicted microlensing event combined with prior knowledge of the source and lens from their GDR2 astrometric solutions, will allow the determination of the lens mass.

Parameter	Unit	Prior
$\alpha_{\text{ref,rel}}$	mas/cos $\delta_{\text{ref}S}$	$\mathcal{N}(\alpha_{\text{ref}L} - \alpha_{\text{ref}S}, \sigma_{\alpha_{\text{ref}L}}^2 + \sigma_{\alpha_{\text{ref}S}}^2)$
$\delta_{\text{ref,rel}}$	mas	$\mathcal{N}(\delta_{\text{ref}L} - \delta_{\text{ref}S}, \sigma_{\delta_{\text{ref}L}}^2 + \sigma_{\delta_{\text{ref}S}}^2)$
$\mu_{\alpha^*,\text{rel}}$	mas/yr	$\mathcal{N}(\mu_{\alpha^*L} - \mu_{\alpha^*S}, \sigma_{\mu_{\alpha^*L}}^2 + \sigma_{\mu_{\alpha^*S}}^2)$
$\mu_{\delta,\text{rel}}$	mas/yr	$\mathcal{N}(\mu_{\delta L} - \mu_{\delta S}, \sigma_{\mu_{\delta L}}^2 + \sigma_{\mu_{\delta S}}^2)$
ϖ_{rel}	mas	$\mathcal{N}(\varpi_L - \varpi_S, \sigma_{\varpi_L}^2 + \sigma_{\varpi_S}^2)$
F_L^G	e-/s	$\mathcal{N}(F_L^G, \sigma_{F_L^G}^2)$
F_S^G	e-/s	$\mathcal{N}(F_S^G, \sigma_{F_S^G}^2)$
M_L	M_{\odot}	$\mathcal{U}(0.0001, 1)$

Table 6: Priors used for each parameter. $\mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean μ and variance σ^2 . $\mathcal{U}(a, b)$, $b > a$ is the uniform distribution and is constant in the interval $[a, b]$ and zero otherwise. As a simplifying assumption we have ignored the covariances between the astrometric parameters provided by GDR2. We found inclusion of the generally small covariance terms did not effect our results but significantly slowed our simulations due to the required inversion of the covariance matrix.

4.3 CANDIDATE SEARCH

In order to find predicted photometric microlensing events, we take a high proper motion ($\mu_{\text{tot}} > 150$ mas/year) sample of 168,734 lens stars from the GDR2 source catalogue. In order to try and remove spurious high proper motion objects from our sample, we make further photometric cuts. We take high proper motion objects with Gaia G-band magnitude $G < 19$, to allow for visual confirmation of the candidate events using current image data, and all objects with a measured Gaia $G_{\text{BP}} - G_{\text{RP}}$ colour, so photometric mass estimates may be obtained for the lens. These further photometric cuts leave a sample of 136,791 lens stars. To narrow our search, we cross-match each star in our lens sample with all GDR2 sources within a radius of 10 times the proper motion of the lens. This leaves $\approx 10,000$ lens-source pairs which we investigate further. We find the closest approach of each lens-source pair by the method in Section 2.4.2 with the basin-hopping algorithm (Wales & Doye, 1997) implemented by SciPy. Uncertainties in the event parameters were calculated using a Monte Carlo simulation with 10^4 samples drawn from the priors in Table 6. We search for events with a closest approach within the remaining nominal Gaia mission time which we assume to be J2018–J2022, and look for events with a closest approach separation < 10 mas which

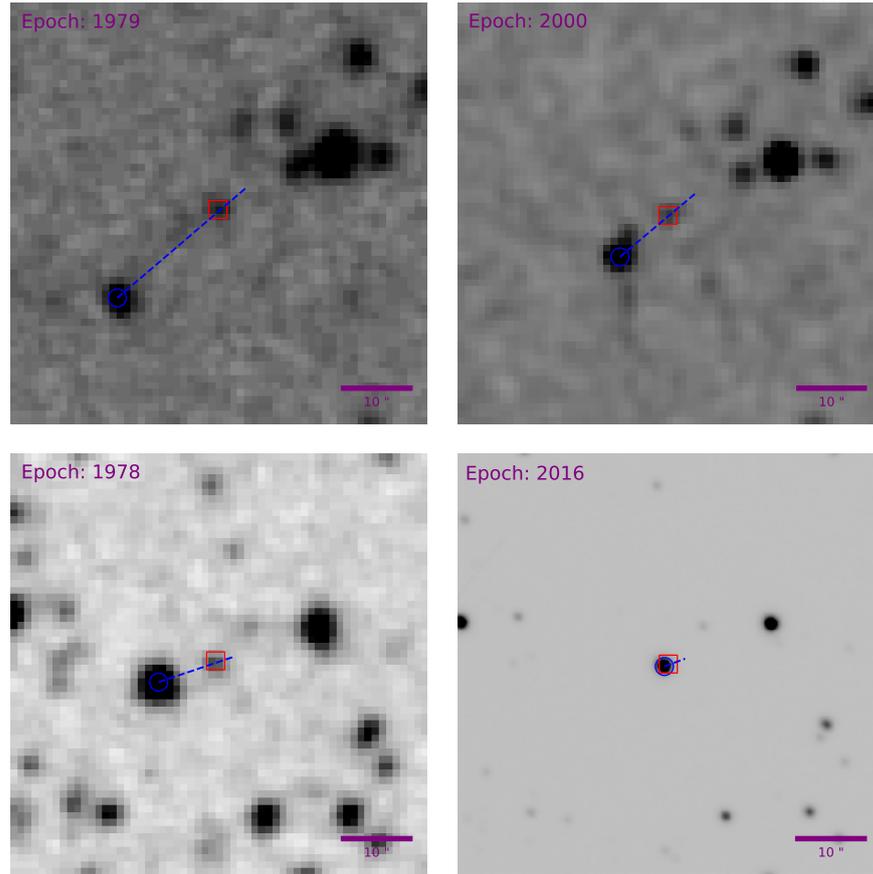


Figure 31: Stellar field around each of the candidate events. **Top:** shows the stellar field around candidate 1. Left shows a Digitized Sky Survey (DSS) image at epoch J1979. Right shows a 2MASS image at epoch J2000 (Skrutskie et al., 2006). **Bottom:** shows the stellar field around candidate 2 at two different epochs. Left shows a DSS image at epoch J1978. Right shows a Dark Energy Camera Plane Survey at epoch J2016 (Schlafly et al., 2018). In each image the positions of the lens and source at the image epoch are indicated with a blue circle and red square, respectively. The projected lens trajectory along its proper motion vector is shown with a blue dashed line in each case.

are likely to have a detectable photometric signal. We find two such events.

4.4 CANDIDATE 1

We predict that the lens with GDR2 5840411363658156032 (hereafter L1) will lens the light of a background star on J2019 November 3rd (± 1 day) or J2019.839 (± 0.003), with a closest approach of $\beta_0 = 5.75^{+1.22}_{-1.16}$ mas. The top row of Fig. 31 shows the stellar field around L1 at two different epochs. Table 7 contains information for both

L1 and the background source. $G_{BP} - G_{RP}$ colour and estimated absolute G-band magnitude G_{abs} , are consistent with L1 being a cool subdwarf.

We obtained the 300 – 2480 nm wavelength spectrum of L1 with the XSHOOTER (Vernet et al., 2011) on the VLT on the 8th of July 2018 under a seeing of 0.61 arcseconds and an average airmass of 1.61. The XSHOOTER spectrum was observed in an AB nodding mode with slits of 1.0 arcseconds in the ultraviolet-blue (UVB), and 0.9 arcseconds in the visible (VIS), and near-infrared (NIR) arms providing a resolving power of 5100, 8800, and 5100, respectively. The integration time was 2×223 seconds in the UVB, 2×235 seconds in the VIS, and 2×250 seconds in the NIR. A wavelength and flux-calibrated 2D spectrum of L1 was first reduced with European Southern Observatory (ESO) Reflex (Freudling et al., 2013). Then we extracted a 1D spectrum from the 2D spectrum with IRAF1 task APSUM. Telluric absorption correction in the VIS and NIR arms were achieved with a B9.5 V type telluric standard (HIP 69011) observed 2.3 hours before our target at an airmass of 1.12. Note the telluric correction of L1 is not quite clean, as the telluric standard was observed at a lower airmass than the target.

The original XSHOOTER spectrum of L1 has a Signal-to-Noise ratio (S/N) of 30 at 490 nm, 74 at 830 nm, 41 at 1200 nm, 39 at 1700 nm, and 52 at 2150 nm, respectively. Note that the original spectrum was smoothed (using a boxcar smooth with IRAF splot) by 101 pixels in the UVB and VIS, and 51 pixels in the NIR for display in Figs. 32 and 33 which increased the S/N by about 10 and 7 times. Fig. 32 shows the UVB and VIS spectrum of L1. We classified it as an M3 ultra subdwarf (usdM3), as it fits well to the usdM3 subdwarf, SDSS J141929.45+154819.1 observed by the SDSS (York et al., 2000). The classification of M subdwarfs is based on the CaH and TiO optical absorption bands (Lépine et al., 2007). The TiO absorption band gets weaker with the decrease in metallicity of M subdwarfs.

As the CaH band is sensitive to T_{eff} and the TiO band is sensitive to $[Fe/H]$, we therefore fitted our observed spectrum with BT-Settl model spectra (Allard, 2014), primarily by their CaH and TiO absorption bands, and by the overall profile in a non-standard approach. We also applied linear interpolation between some models where this was able to improve the fit. We used a fixed gravity because early-type M subdwarfs have similar gravity around $\log g = 5.0$ according to model predictions. We gradually changed the T_{eff} and $[Fe/H]$ of

Table 7: Lens, background source and event data for the two candidates. Data for all objects are from the GDR2 source catalogue. The coordinates (α, δ) are on the ICRF and at epoch J2015.5. F_L^G is the Gaia G-band flux, G_{abs} is the absolute G-band magnitude calculated via $G_{\text{abs}} = G + 5 + 5 \log_{10}(\varpi/1000)$. AEN Sig is the excess astrometric noise significance parameter provided in GDR2.

	Candidate 1 Lens	Candidate 2 Lens
GDR2 id	5840411363658156032	5862333044226605056
α_{refL} (deg \pm mas)	196.460398500 \pm 0.07	196.506734481 \pm 0.03
δ_{refL} (deg \pm mas)	-72.300995137 \pm 0.06	-63.532796061 \pm 0.03
μ_{α^*L} (mas/year)	-362.72 \pm 0.13	-209.59 \pm 0.05
$\mu_{\delta L}$ (mas/year)	306.51 \pm 0.12	79.60 \pm 0.05
ϖ_L (mas)	9.52 \pm 0.08	6.71 \pm 0.04
G (mag)	17.18	15.18
F_L^G (e-/s)	2500 \pm 1.95	16000 \pm 12.67
$G_{\text{BP}} - G_{\text{RP}}$ (mag)	2.14	2.16
G_{abs} (mag)	12.07	9.31
AEN Sig	1.75	5.42
	Candidate 1 Source	Candidate 2 Source
GDR2 id	5840411359350016128	5862333048529855360
α_{refS} (deg \pm mas)	196.459018289 \pm 0.13	196.506239773 \pm 0.36
δ_{refS} (deg \pm mas)	-72.300626103 \pm 0.11	-63.532706855 \pm 0.51
μ_{α^*S} (mas/year)	-12.95 \pm 0.29	-9.02 \pm 0.56
$\mu_{\delta S}$ (mas/year)	2.92 \pm 0.23	-4.42 \pm 0.91
ϖ_S (mas)	0.59 \pm 0.13	-0.06 \pm 0.37
G (mag)	18.17	18.09
F_S^G (e-/s)	1000 \pm 1.62	1100 \pm 8.89
$G_{\text{BP}} - G_{\text{RP}}$ (mag)	1.43	-
AEN sig	0	8.30
	Candidate Event 1	Candidate Event 2
t_0 (Julian year)	2019.839 \pm 0.003	2019.42 \pm 0.01
β_0 (mas)	5.83 $^{+1.26}_{-1.20}$	6.48 $^{+3.38}_{-3.36}$
M_L (M_{\odot})	0.11 \pm 0.01	0.38 \pm 0.06
Θ_E (mas)	2.82 \pm 0.12	4.56 $^{+0.35}_{-0.37}$
u_0 (Θ_E)	2.03 $^{+0.48}_{-0.47}$	1.41 $^{+0.78}_{-0.76}$
t_E (day)	2.2 \pm 0.1	7.7 \pm 0.6
$A_{\text{lum,max}}$ (mmag)	20 $^{+20}_{-10}$	10 $^{+40}_{-10}$
$\delta_{\text{lum,max}}$ (mas)	0.077 \pm 0.007	0.20 $^{+0.04}_{-0.03}$
$\delta_{+,max}$ (mas)	0.569 $^{+0.010}_{-0.089}$	0.04 \pm 0.01

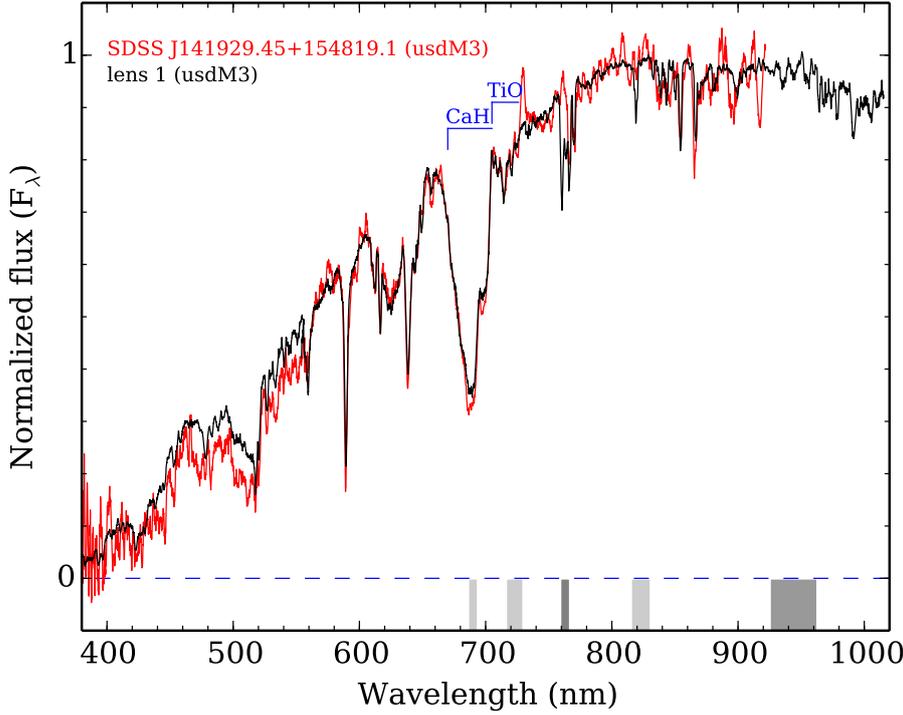


Figure 32: The UVB and VIS spectrum of L1 compared to a usdM3 subdwarf, SDSS J141929.45+154819.1. Telluric absorption regions are indicated with grey bands (corrected). Lighter and thicker shaded bands indicate regions with weaker and stronger telluric effects, respectively.

models by steps of 50 K and 0.1 dex to find the best-fitting model by visual inspection focused on the CaH, TiO bands, and overall profile. The best-fitting BT–Settl model for L1 has $T_{\text{eff}} = 3400$ K, $[\text{Fe}/\text{H}] = -1.8$, and $\log g = 5.0$. The atmospheric parameter uncertainties for L1 are around 100 K for T_{eff} , 0.2 dex for $[\text{Fe}/\text{H}]$, and 0.25 dex for $\log g$. Fig. 33 shows the full XSHOOTER spectrum of L1 compared to a best-fitting BT–Settl model spectrum. We placed L1 in a T_{eff} versus $[\text{Fe}/\text{H}]$ space (e.g. Fig. 9 in Zhang et al., 2017), and found that it has a mass of $0.11 \pm 0.01 M_{\odot}$, according to 10×10^9 year iso-mass contours, predicted by evolutionary models (Baraffe et al., 1997; Chabrier & Baraffe, 1997).

The background source in this event is likely a cool subdwarf at a distance of 1693^{+490}_{-310} pc (Gaia Collaboration et al., 2018) according to its location ($G_{\text{abs}} = 7.03^{+0.44}_{-0.55}$ mag and $G - G_{\text{RP}} = 0.81$ mag) on the Hertzsprung–Russell diagram (e.g. Fig. 21 in Zhang et al., 2018). BT–Settl Allard (2014) model spectral energy distribution (SED) fitting with the virtual observatory SED Analyzer (VOSA; Bayo et al., 2008) shows that background source likely has $T_{\text{eff}} \approx 4400$ K and $[\text{Fe}/\text{H}] \approx -3.5$ identifying it as a likely K subdwarf (e.g. Fig. 9 in

Zhang et al., 2017). Note that spectra of cool subdwarfs hotter than 4000 K have a smooth continuum and have no optical broad absorption bands like CaH and TiO. The background source is about 1 mag fainter than L1 in Gaia G-band. At the time we obtained the XSHOOTER spectrum of L1, (8th July 2018) L1 and the background source were separated by ≈ 0.63 arcseconds, therefore possible contaminating flux from the background source in the spectrum of L1 has to be considered. If the spectrum of L1 was contaminated by a small fraction of continuum flux from the background source the actual T_{eff} and $[\text{Fe}/\text{H}]$ of L1 would be slightly lower than the best-fitting model in Fig. 33, and its mass would be slightly lower than our initial estimation. To evaluate the flux contamination of the background source in the spectrum of L1, we compared the Gaia and 2MASS photometric spectral energy distributions of L1 and the background source to the XSHOOTER spectrum of L1 in Fig. 33. The SED of L1 fitted well to its spectrum across optical and NIR. A combined SED of L1 and the background source would be bluer than that of L1 alone. Combined SEDs of L1 and the background source are also plotted in Fig. 33, one with 20% and one with all flux from background source. Therefore, we conclude that flux contamination from the background source is likely not present in the spectrum of L1. Moreover, even if the spectrum of L1 is contaminated by a small fraction ($< 20\%$) of continuum flux from the background source, the impact on its mass estimation would be small and within the initial uncertainty of $0.01M_{\odot}$.

Using the mass of L1 derived from atmospheric and evolutionary models, we estimate its Einstein radius $\Theta_E = 2.82 \pm 0.12$ mas. This corresponds to a peak photometric amplification of the lens–source blend flux of $1.02_{-0.01}^{+0.02}$ or a change in magnitude of 20_{-10}^{+20} mmag. The Einstein time-scale for this event is 2.2 ± 0.1 days. Also, as pointed out by B18, we would see a boosted signal by a careful choice of the photometric filter (factor of ≈ 1.3 in G_{BP}). At the point L1 is still resolvable from the background source by Gaia, the maximum astrometric shift of the major source image is $\delta_{+, \text{max}} = \delta_+ (103 \text{mas}/\Theta_E) = 0.077 \pm 0.007$ mas. At closest approach, the maximum shift of the lens–source blend will be $\delta_{\text{lum}, \text{max}} = 0.569_{-0.089}^{+0.010}$ mas.

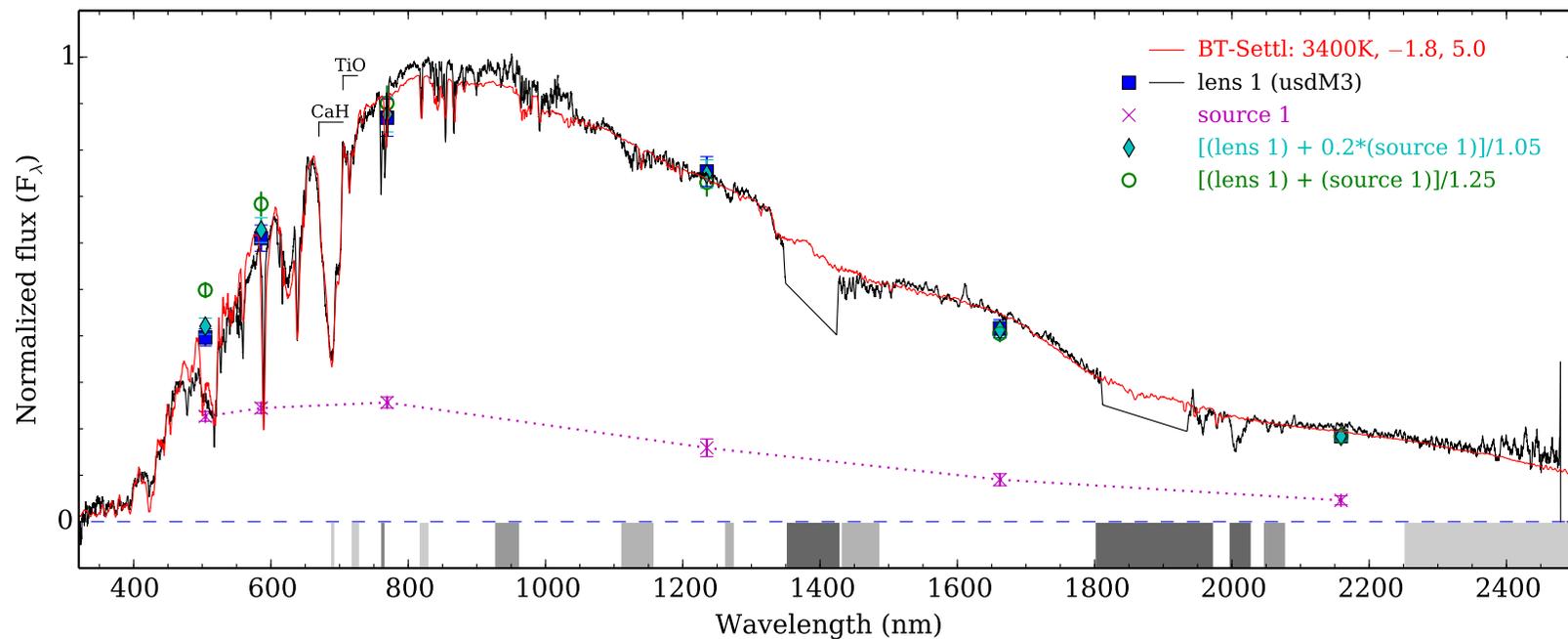


Figure 33: The XSHOOTER spectrum of candidate L1 (black) compared to its best-fitting BT-Settl model spectrum (red). Telluric absorption regions are indicated with grey bands (corrected). Lighter and thicker shaded bands indicate regions with weaker and stronger telluric effects. The Gaia and 2MASS photometric SEDs of L1 (blue squares) and the background source (source 1, magenta crosses) have the same scale. Combined SEDs (cyan diamonds and green open circles) are normalized to that of L1 at K-band. An uncertainty of 0.05 mag is used for Gaia photometry.

This event was also independently discovered as candidate ME19 in **B18**. While we are in agreement with **B18**'s closest approach and separation for this event, our predicted signal strengths differ considerably. The reason for this is that **B18** classifies L1 as a late-type M dwarf (**B18**: sections 6.3 and 7), and adopts a mass of $0.25M_{\odot}$, $\approx 225\%$ higher than our own. Consequently, **B18** predicts significantly higher peak photometric and astrometric signals for this event. **K18b** also independently predicts this event as candidate number 4. Similarly to **B18**'s predictions, we are in good agreement with **K18b**'s astrometric calculations for this event, however, **K18b** adopts a mass of $0.17M_{\odot}$ for L1 (**K18b**: Table 2), $\approx 50\%$ higher than ours. As a result, **K18b** also predicts higher peak photometric and astrometric signals for this event.

4.5 CANDIDATE 2

We also predict the lens L2, with GDR2 5862333044226605056, will lens the light of a background star with closest approach on 3rd June J2019 (± 4 days) or $J2019.42 \pm 0.01$, with closest approach of $\beta_0 = 6.48^{+3.38}_{-3.36}$ mas. Fig. 31 shows the stellar field around L2 at two different epochs, and Table 7 contains details on the lens, source, and event.

L2 is a known high proper motion object, visually confirmed by **Smith et al. (2018)** with VIRAC ID 323066023. L2's $G_{BP} - G_{RP}$ and G_{abs} are consistent with L2 being a mildly metal-poor M dwarf. Its tangential velocity hints at thick disc kinematics, further evidence for a somewhat low metallicity. In order to obtain a mass estimate for L2, we first fit Gaia G_{BP} , G , and G_{RP} , VISTA VVV survey J-band, and WISE W_1 and W_2 photometry to model SEDs using the VOSA. We omitted the VVV Z- and Y-band photometry due to relatively poor calibration in those bands, and H and K-band photometry due to saturation. We note that flux from the background source will contaminate the WISE photometric measurements, but given its relative faintness, contamination should be minor. We omitted the W_3 and W_4 upper limits from the SED fit but retained the W_1 and W_2 bands, as the additional wavelength coverage offsets the negative impact of the potential contamination. We fit the photometry to the BT-Settl and BT-Dusty models (**Allard et al., 2012**) with the following restrictions, $2000K < T_{eff} < 5000K$, $3 < \log g < 6$, and $-1.0 < [Fe/H] < -0.5$. The best fit from both models is $T_{eff} = 3500 \pm 50K$, $\log g = 4 \pm 0.25$, and

$[\text{Fe}/\text{H}] = -0.5 \pm 0.25$. We note that the minor contamination from the background source in W_1 and W_2 is just about apparent when comparing the SED to the best-fitting model. We then used the isochrones Python package (Morton, 2015) with the Dartmouth model grid (Dotter et al., 2008) to obtain a mass estimate for L2 of $0.38 \pm 0.06 M_{\odot}$.

Using the mass derived from L2’s photometry, we estimate its Einstein radius is $\Theta_E = 4.56_{-0.37}^{+0.35}$ mas. Consequently, we find a peak photometric amplification of the lens–source blend flux of $1.01_{-0.01}^{+0.04}$ or a change in magnitude of 10_{-10}^{+40} mmag. We note that both the source and lens are present in 2MASS (Skrutskie et al., 2006) with H-band magnitudes of 12.19 and 13.64, respectively. Because of this more favourable contrast ratio between the source and lens, we expect a boosted signal by a factor of ≈ 1.2 if this event were observed in the H-band. $t_E = 7.7 \pm 0.6$ days for this event. At the point L2 is still resolvable from the background source, there is a maximum astrometric shift due to the major source image of $\delta_{+, \text{max}} = \delta_+ (103 \text{mas} / \Theta_E) = 0.20_{-0.03}^{+0.04}$ mas. Over the event maximum, when the lens and source are blended, we predict a peak astrometric shift of the lens–source light centroid of $\delta_{\text{max}} = 0.04 \pm 0.01$ mas. This event was also independently predicted by K18b, as event number 3. We find we are in good agreement with K18b’s astrometric calculations, mass estimate for the lens, and consequently the predicted photometric and astrometric signals for this event. This event is not in B18’s sample as both L2 and the background source fail B18’s astrometric excess noise significance quality cuts.

4.6 OBSERVATIONAL OUTLOOK

4.6.1 Astrometric signal

Both candidates 1 and 2 have relatively low astrometric deflection magnitudes which will be challenging to detect. This is due to a combination of the specific event geometry and contaminating flux from the lens suppressing the astrometric signal when the lens and source are unresolved around the event maximum. In this section, we assess the feasibility of determining the mass of L1 and L2 from astrometric monitoring by Gaia and HST.

4.6.1.1 *Gaia*

For Gaia, an important predictor of the precision at which a deflection can be measured, is the scan direction relative to the deflection direction. For measurements in which the deflection is aligned AL, measurements will be precise, whereas measurements aligned in the AC direction will be considerably less precise (Fabricius et al., 2016). Recent simulations of the astrometric centroiding precision of Gaia have been carried out by Rybicki et al. (2018). For objects with G-band magnitude ≈ 15 , 17 and 18, Rybicki reports AL precisions $\sigma_{AL} \approx 0.1$, 0.5, and 0.7 mas, and AC precisions $\sigma_{AC} \approx 3$, 20, and 60 mas (see Rybicki et al., 2018, Tables 1 and 2), for astrometric precision as a function of Johnson V mag, and Fig. 4. in B18, for the conversion to Gaia G.

In the case of candidate 1, the peak astrometric deflection of the $G \approx 18$ source at the point the lens and source are still resolvable, is $\delta_{+,max} \approx 0.077$ mas which is $< \sigma_{AL}$. However, around event maximum, the $G \approx 17$ lens–source blend will shift by $\delta_{lum,max} \approx 0.57$ mas which is $> \sigma_{AL}$. This means candidate 1 should be borderline astrometrically detectable by Gaia around its maximum, but for only the most favourable scan directions. We find for candidate 2 both when the lens and source are resolved $\delta_{+,max} < \sigma_{AL}$ and around closest approach when they are not, $\delta_{lum,max} < \sigma_{AL}$. This means both the resolved tails and the unresolved maximum astrometric shift of candidate 2 are unlikely to be detected by Gaia. Overall, this presents a pessimistic outlook for mass measurements of L1 and L2 solely derived from the astrometric deflection measurements by Gaia.

4.6.1.2 *HST*

A better option may be to monitor the events with the WFC3 instrument on the HST. This is because, unlike Gaia, the exposure times with HST can be optimized to obtain the highest signal to noise per epoch for the event. Single-epoch pointed imaging measurements with accuracies of $\approx 0.1 - 0.2$ mas have been achieved with HST (see Bellini et al. (2011), S17). Moreover, S17 measured the mass of white dwarf, Stein 2051 B, to a precision of $\approx 8\%$ with WFC3, via astrometric microlensing. However, S17 was only able to measure astrometric deflection of the background source in the presence of the ≈ 6.5 mag brighter lens when they were separated by greater than ≈ 450 mas.

For candidate 1, the lens and background source only differ in brightness by ≈ 1 mag. Even in light of this more favourable contrast ratio between the lens and source, and assuming WFC3 could resolve the lens and source down to 103 mas, the resolved tails of this event are unlikely to be detected with HST. While the unresolved astrometric shift of the lens–source blend around maximum is likely to be detected by HST for candidate 1, the astrometric shift is now a function of the ratio of the lens–source flux (Eq. 28). This means that, in addition to precise astrometric measurements, precise photometric measurements of the lens and source will have to be obtained which will ultimately introduce more scatter into the final inference on the mass of L1.

For candidate 2, the background source is ≈ 3 orders of magnitude fainter than the lens. Again, assuming WFC3 can resolve the background source and lens at 100 mas, the resolved tails of this event will only be borderline detectable with HST. Over candidate 2’s event maximum, the unresolved astronomic shift is unlikely to be detected by HST. Overall the outlook for astrometric monitoring of both events with HST looks more optimistic than with Gaia, but still challenging. The precision at which the masses of both L1 and L2 can be determined from astrometric measurements with HST will depend on the closest separation at which the lens and background source can be resolved.

4.6.2 *Photometric signal*

While candidate 1 is visible from the Southern Hemisphere, it is only observable from the ground at high airmass over its event maximum. Therefore, ground-based photometric follow-up is unlikely to provide adequate measurements at a precision less than the signal amplification of ≈ 20 mmag. We instead consider a space-based photometric monitoring campaign.

Due to Gaia’s scanning law and the short duration of candidate 1’s amplification ($t_E = 2.2$ days), Gaia is unlikely to provide adequate photometric sampling of the event. Official Gaia scan forecasts are not yet available for J2019. If we assume that Gaia’s J2018 observing schedule will be similar to next year’s and consider ± 1 month around candidate 1’s maximum (November 3rd) the Gaia Observation Scheduling Tool indicates only two measurements will be taken. Instead, we consider a photometric monitoring campaign with HST,

as mmag precision is within HST's capabilities and we could reasonably expect to acquire between 10 and 20 measurements over the course of a month with a targeted HST campaign.

Candidate 2 is also visible from the Southern Hemisphere. In contrast to candidate 1, ground-based photometric monitoring of the event looks more optimistic. For example, candidate 2 is observable above the horizon around maximum amplification on J2019 June 3rd from the Cerro Tololo Inter-American Observatory. Photometric precision between ≈ 1 and 10 mmag is typically achieved from ground-based microlensing surveys (e.g. Udalski et al., 2015; Skowron et al., 2015), and \approx mmag precision is routinely achieved from the ground in the studies of transiting exoplanets (e.g. Gillon et al., 2016; Bayliss et al., 2018; Dittmann et al., 2017). These favourable observing conditions combined with candidate 2's longer event time-scale ($t_E = 7.7$ days), will permit a targeted multi-band observing campaign from the ground.

4.7 SIMULATED MASS DETERMINATION

In the previous section we have shown that monitoring the predicted astrometric shift of candidates 1 and 2 is unlikely to constrain the mass of L1 and L2. However, a space and ground-based photometric monitoring campaign will likely detect candidate 1 and 2's photometric signals, respectively. In this section, we turn our attention to assessing the precision at which the masses of the lenses can be inferred from photometric measurements combined with priors from the GDR2 astrometric solution.

To investigate the expected precision, we generate synthetic light curves according to Eq. (61). Specifically, we draw truth values for the mass of L1 and L2 from our mass estimates ($\sim \mathcal{N}(0.11, 0.01)M_\odot$, $\sim \mathcal{N}(0.38, 0.06)M_\odot$, respectively), and values from the priors outlined in Table 6 for the rest of the model parameters. Then, using these truth parameters, we generate data points according to a particular sampling and noise scenario for both candidates.

To infer the mass of L1 or L2 from the data, we sample from the posterior distribution outlined in Eq. (64) using the Affine Invariant MCMC Ensemble sampler, implemented by the EMCEE Python package (Foreman-Mackey et al., 2013), with the priors outlined in Table 6 derived from GDR2 quantities for the source and lens. We initialize 180 walkers in a small Gaussian ball around the maximum a pos-

teriori probability (MAP) estimate.³ The MAP estimate is the set of model parameters that maximize the right hand side of Eq. (64). We initially run a burn-in of 150 steps. We then run the sampler for a further 150 steps keeping the final 100 steps. Visual inspection of the MCMC chains for each parameter showed good mixing which can indicate good convergence. As a further convergence check we ran experiments with much longer chains of 10000 steps and found the posteriors on the lens mass of L1 and L2 did not change.

For example, we can generate 20 data points with ≈ 1 mmag measurement error, and corresponding independent Gaussian noise of $\epsilon_i \sim \mathcal{N}(0, 1)$ mmag on each point. We choose these points to be uniformly distributed within $20 t_E$ (40.4 days for candidate 1) of the maximum amplification, to allow for characterization of both the base lens–source blend and amplification shape. Fig. 34 shows the marginal distributions from the posterior for this set of generated event data for the candidate 1 event. The generated data and model samples from the inference are also shown. Fig. 34 shows that, for this particular set of generated data, we were able to recover a mass for L1 of, $M_L = 0.12 \pm 0.02 M_\odot$ which is consistent with the true lens mass used to generate the data $M_{L,true} = 0.115 M_\odot, \pm 17\%$.

Fig. 34 also shows a slight degeneracy in relative proper motion of the source and lens ($\mu_{\alpha*,rel}, \mu_{\delta,rel}$) and the lens mass M_L . There is also a degeneracy in the flux of the lens F_L^G and the flux of the source F_S^G . This is because the observed flux is always a blended combination of both the lens and source flux.

To estimate the average precision at which this method can determine the mass of L1 and L2 for a given set of data, we run the inference 100 times and stack the normalized marginal lens mass distributions, that is the samples of $(M_L - M_{L,true})/M_{L,true}$. This Monte Carlo simulation accounts for the spread in possible prediction values for the event and spread in reasonable lens masses. Overall this simulation quantifies the spread in possible constraints we are likely to obtain on the lens mass. In the following sections, we investigate a range of light-curve sampling and photometric precision scenarios.

4.7.1 Candidate 1 mass inference

As mentioned in section 4.6.2, we expect to follow-up candidate 1 with a space-based photometric observing campaign with HST. We assume we can obtain 10 to 20 measurements at ≈ 1 mmag preci-

³ The MAP is chosen as an initialization over the maximum likelihood estimate because we are using informative priors.

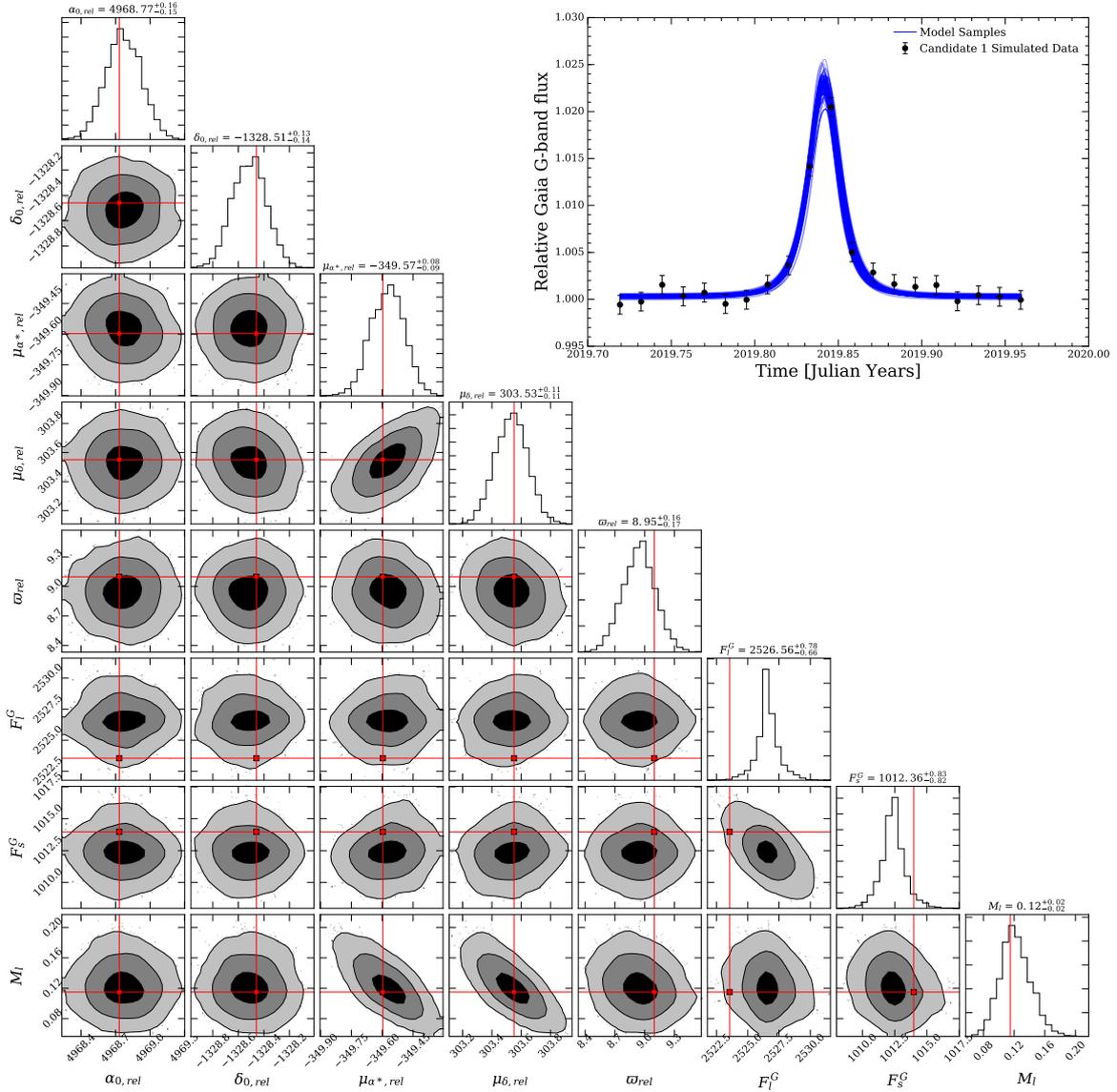


Figure 34: **Bottom left:** marginal posterior distributions for the eight parameters inferred from simulated microlensing light-curve data for candidate 1. $\delta_{0,rel}$, and $\alpha_{0,rel}$, are initial separations of the lens and source in units of mas and mas/cos δ_{0S} . $\mu_{\alpha^*,rel}$, and $\mu_{\delta,rel}$, are the relative lens–source proper motion in units of mas/year. ω_{rel} is the relative lens–source parallax in units of mas. F_L^G and F_S^G are the lens and source G-band fluxes in units of e-/s, and M_L is the mass of the lens in units of M_\odot . Contours show 1 σ , 2 σ , and 3 σ bands. Red lines indicate the truth values used to generate the data for each parameter. **Top right:** Black points indicate simulated data used for the inference with \approx mmag photometric precision. Blue lines show samples of the corresponding inferred light curves

sion over the course of the event. Fig. 35 shows a plot of the stacked marginal mass distributions for a range of different sampling and precision models. We investigate data points with 0.1, 1, and 10 mmag precision with corresponding noise $\epsilon_i \sim \mathcal{N}(0, 0.1, 1, 10)$ mmag. For our likely expected photometric precision with HST (1 mmag), we investigate how this affects the inferred mass of L1.

For a space-based monitoring campaign, we are most constrained by the number of measurements that can be obtained. For this reason, we investigate the effect of clustering the majority of the data points around the event maximum where the information about the lens mass is encoded, while still leaving some points to characterize the base lens–source blend flux. We compare a ‘clustered’ sampling regime in which we uniformly distribute 60% of the data points within $6t_E$, and the rest uniformly between 6 and $20t_E$ of maximum amplification, with the standard uniform sampling within $20t_E$.

Fig. 35 shows that for 1 mmag photometric precision and noise with 20 clustered measurements, we can expect a precision on the inference of L1’s mass of $\approx 20\%$. At 10 clustered data points at 1 mmag precision, we see the inference degrades to $\approx 27\%$. We can see the effect of clustering most clearly at 20 data points, where we obtain a $\approx 3\%$ precision increase compared with uniform sampling on the mass inference. Taking the worst and best case within the 1 mmag bracket, we conclude the mass of L1 is likely to be determined between 20 – 27% precision. Also included in Fig. 35, are two extreme cases of 20 clustered measurements with 0.1 and 10 mmag precision and noise, resulting in a mass precision of $\approx 9\%$ and $\approx 37\%$, respectively. The 0.1 mmag case is unlikely to be achieved with HST and demonstrates that this method will likely be limited by the photometric measurements of the light curve, and not by the priors provided by GDR2.

4.7.2 Candidate 2 mass inference

In contrast to candidate 1, candidate 2 is observable from the ground and has $t_E = 7.7$ days which is 3.5 times longer than candidate 1. Therefore, a dense sampling of the microlensing light curve is achievable. For candidate 2, we first assume 50 measurements can be obtained with the same clustered regime as described above. For this sampling, we investigate a range of different photometric precisions of 0.1, 1, 5, and 10 mmag with corresponding noise

$\sim \mathcal{N}(0, \{0.1, 1, 5, 10\})$ mmag. We then fix the photometric precision and noise at 1 mmag and investigate various numbers of measurements of the light curve (25, 50, 100, and 200 data points). We handle the negative parallax measurement for the source star (see Table 7) by setting the source parallax to zero. This means the prior on ϖ_{rel} and distribution used to generate the synthetic data is just $\varpi_{\text{rel}} \sim \mathcal{N}(\varpi_{\text{L}}, \sigma_{\varpi_{\text{L}}}^2)$.

Fig. 36 shows the stacked normalized inferred mass distributions for candidate 2. We can see for 1 – 5 mmag precision cases we expect to infer the mass of L2 at $\approx 30 - 42\%$ precision. For the 10 mmag case, the inference degrades to $\approx 47\%$. Fig. 36 also shows that, for 1 mmag photometric precision and noise, we only see marginal improvement for more data points. Specifically, there is only a $\approx 1\%$ improvement when the number of data points is increased by a factor of 4, from 50 to 200. The extreme case of 50 points at 0.1 mmag precision, which is unlikely to be achieved from the ground, is also shown in Fig. 36. Similarly to candidate 1, we are limited by the photometric precision and not the GDR2 priors for inference on the mass of L2. Overall, taking the 1 and 5 mmag cases, we expect to recover the mass of L2 with $\approx 30 - 42\%$ precision. We note that the clustered sampling scenarios typically offer gains of $\approx 5\%$ precision in the mass of L2 compared with the uniform sampling case for candidate 2.

4.7.3 Caveats

There are also some caveats. First, we have assumed a point mass and point lens, yet the lens could have a companion, which will make modeling significantly more challenging. Second, we have assumed the parallax of the background source for candidate 2 to be zero. The actual parallax of the background source will be greater than zero, although this should not effect the simulation too much as the lens is so close ($D_{\text{L}} \ll D_{\text{S}}$). However the parallax of the background source for candidate 2 will have an uncertainty which may degrade the inference on the mass of L2. And finally, we have assumed that there is no intrinsic variability in the flux both from the source and lens. This could be of the same order of the microlensing signal, although in principle this could also be modeled. However, variability may permit improved blending constraints for the event (Wyrzykowski et al., 2006).

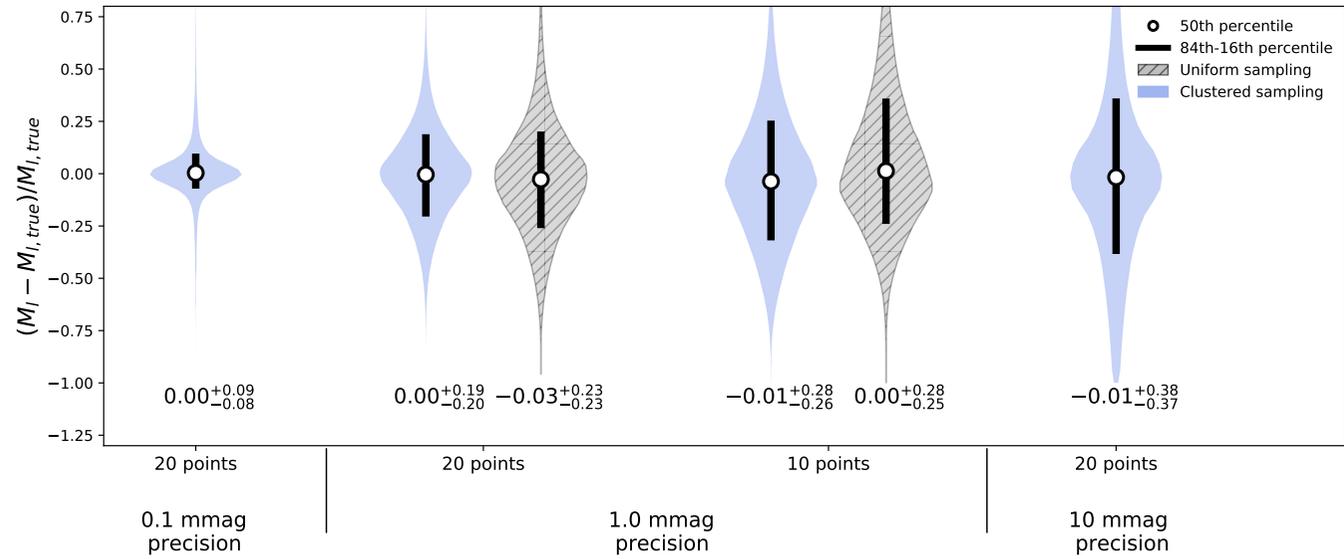


Figure 35: Violin plot of the expected precision at which the mass of L1 is expected to be recovered for a range of light-curve sampling scenarios. In each case, the circle represents the 50th percentile, the black line indicates the data between the 84th and 16th percentile, and the shaded region is a kernel density plot showing the shape of the distributions. Under each violin, the corresponding 50th percentile with uncertainty based on the 16th and 84th percentile, is shown. For example, sampling the light curve of candidate 1 with 20 measurements clustered around the maximum amplification at 1.0 mmag precision, we expect to recover the mass of L1 at $0.00^{+0.19}_{-0.21}$ i.e. with $\approx 20\%$ precision (see the text for further discussion).

4.8 DISCUSSION AND CONCLUSIONS

The Gaia satellite is proving a tremendous resource for microlensing (Proff et al., 2011; Belokurov & Evans, 2002). Here we have identified two high-amplitude microlensing events with a systematic search through the GDR2. Our focus here is to predict the time of closest approach between lens and source which is the time of maximum amplification. With this in hand, the lensing event can be densely sampled using ground- or space-based telescopes. Our candidate event 1 has a time of maximum amplification of November 3rd J2019 (± 1 day). It was also identified by B18, who assumed that the lens was an ordinary field M dwarf and as a result significantly overestimated its mass. Our SED fits and XSHOOTER spectrum are consistent with the lens being an M3 type star of mass $\approx 0.11M_{\odot}$. This is a double-edged sword.

On the one hand, mass estimates of cool subdwarfs from lensing are of greater scientific interest than M dwarfs. There are only a few metal-poor stars in close binary systems that have direct dynamical measurements (Jao et al., 2016). Furthermore, it is difficult to obtain a good quality spectrum of the secondary star of a close binary system for atmospheric parameter constraints. For example, in the binary system μ Cassiopeiae, μ Cassiopeiae B is the closest object to L1 in terms of mass and metallicity. It has a dynamical mass of $0.173 \pm 0.011M_{\odot}$ (Drummond et al., 1995) and metallicity between $[\text{Fe}/\text{H}] = -0.71$ (Karaali et al., 2003) and $[\text{Fe}/\text{H}] = -0.84$ (Soubiran et al., 2008). However, μ Cassiopeiae B is too close to its primary to obtain a good quality spectrum for further analysis. Dynamical mass measurements are often applied to close binaries with orbital periods between a few years to a few decades. The candidate 1 microlensing event presented in this paper is an ideal way to determine the mass of a subdwarf without the associated difficulty of obtaining the spectroscopic observations from binary systems. The uncertainty of the mass determined for L1 is expected to be around 20 – 27%. This is just good enough to compare to model predictions. There is no other star with a direct mass measurement in the parameter space of $[\text{Fe}/\text{H}] < -1.0$ and mass $< 0.15M_{\odot}$. Overall, measuring the mass of L1 through microlensing will provide an important benchmark in the characterization of low-mass stars of the halo.

On the other hand, the predicted Einstein radius is smaller than calculated by B18 because of the much lower mass. In contrast to

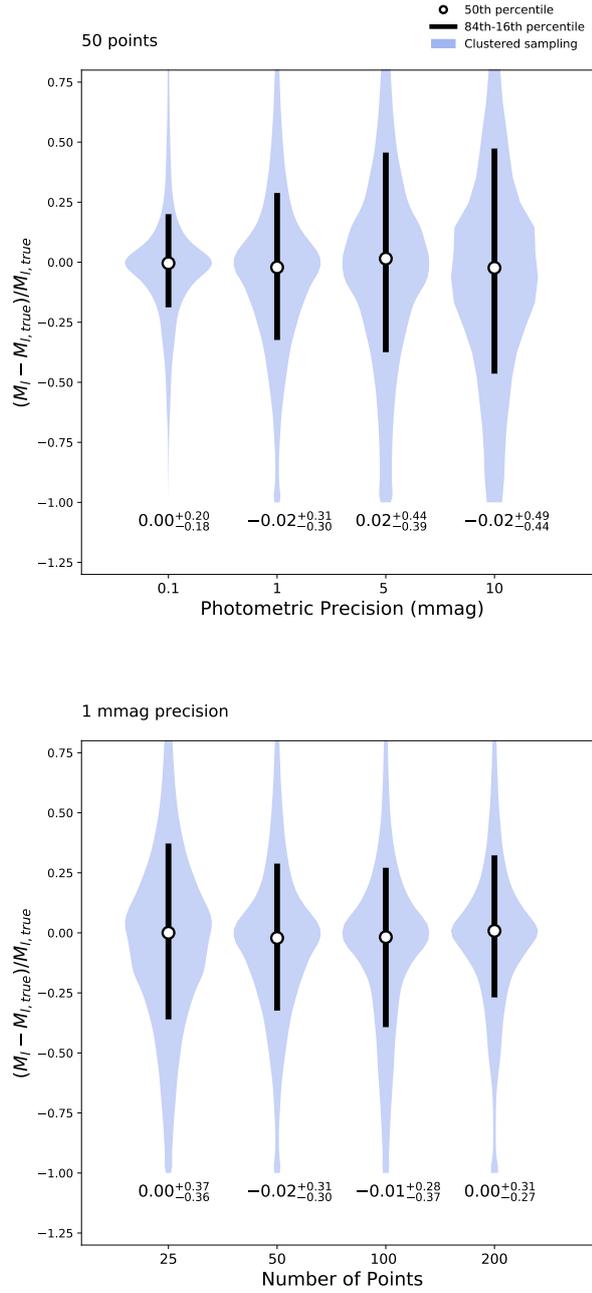


Figure 36: Violin plot for the stacked samples of the inferred mass distributions for candidate 2. In all cases, we distribute data points clustered around the maximum amplification and investigate varying degrees of photometric precision, noise, and number of data points. **Top:** 50 data points with varying degrees of photometric precision and noise. **Bottom:** 1 mmag photometric precision and noise for various numbers of data points. For example, through clustered sampling of 50 data points at 1 mmag photometric precision, we expect to recover the mass of L2 at $-0.02^{+0.31}_{-0.30}$ or $\approx 30\%$ precision.

B18, we find candidate 1 will only be borderline detectable around its unresolved maximum with $\sigma_{\text{max}}(0.57 \text{ mas}) \geq \sigma_{\text{AL}}(0.5 \text{ mas})$ with Gaia. We argue that the outlook for astrometric measurements with HST is not much better. Overall, we conclude that the mass of L1 is unlikely to be inferred from measurements of the predicted astrometric shifts.

For candidate event 2, also found independently by **K18b**, the epoch of maximum amplification occurs on June 3rd J2019 (± 4 days). The lens is a metal-poor M dwarf with mass $\approx 0.38M_{\odot}$. This is derived from SED fitting using Gaia, VISTA, and WISE photometry. The astrometric shifts for candidate 2, in both the resolved and unresolved regimes, are unlikely to be detectable by Gaia. With HST, we expect the resolved tails of this event will be borderline detectable and the unresolved shift will likely be undetected. Similarly to candidate 1, we conclude that the mass of L2 is unlikely to be inferred from measurements of the astrometric shifts by either HST or Gaia.

Rather than follow these events up astrometrically, we suggest that both are excellent candidates for targeted photometric campaigns, in the case of candidate 1, from space and of candidate 2, from the ground. This idea exploits the excellent Gaia astrometric solution to provide priors for all the lightcurve model parameters (with the exception of the lens mass). With current GDR2 data, we have shown that this can provide the lens mass L1 to $\approx 20 - 27\%$, and L2 to $\approx 30 - 42\%$ precision, and in both cases this precision will likely be limited by the photometry and not the priors provided by GDR2.

We note that for both candidates, there is an opportunity for greater precision if improved astrometric data from future Gaia data releases is used (e.g. GDR3 is expected \approx J2023). For candidate 1, as pointed out by **B18**, we would see a boosted signal by a careful choice of the photometric filter (factor of ≈ 1.3 in G_{BP}), leaving further room for improvement. For candidate 2, we will be able to obtain multi-band observations from several ground-based telescopes, and expect a boosted signal in the IR (for example a factor of ≈ 1.2 in 2MASS H), which will lead to improved constraints on the mass of L2.

Some further extensions of the method in this Chapter suggest themselves. For example, we may also be able to extract lens masses for photometric events that have already happened. We could identify now separated lens–source pairs that were detected in the past as an event in the OGLE data sets (**Udalski et al., 2015**). We could then exploit the Gaia astrometry to break the degeneracy of historical microlensing light curves and measure the mass of the lens (e.g.

Kozłowski et al., 2007; Batista et al., 2015). We may also be able to identify low-amplitude events that OGLE missed, but are in the OGLE data stream. This is because Gaia astrometry puts very strong priors on where the event maximum is, so we could possibly tease out very weak microlensing signals in archival data.

This method provides a new way to extract microlensing masses for forthcoming events that can be predicted from Gaia data releases. A natural extension to this paper would be to consider the photometric events found beyond the time baseline of Gaia (Bramich & Nielsen, 2018; Mustill et al., 2018). Moreover, with VIRAC providing a complementary source catalogue to Gaia of ≈ 300 million sources in the Galactic bulge and southern disc, predictive microlensing work is far from complete.

4.9 POST-CONCLUSION NOTE

Candidate 1 presented in this Chapter was followed up by HST (Sahu et al., 2020, Principle investigator Kailash Sahu, Co-investigators including myself). Analysis of this event is ongoing at the time of submission of this thesis and is being led by Kailash Sahu.

Candidate 2 was also followed up. High Cadence (\approx hourly) optical and near-infrared follow up was obtained via the Las Cumbres Observatory data over the event maximum. Additionally, lower cadence infrared follow-up (\approx daily) was obtained via Infrared service facility at the South African Astronomical Observatory. Unfortunately, both sets of follow-up data had light curve scatter at the ≈ 40 mmag level, dwarfing the predicted microlensing signal. Matters were complicated further by the discovery that the background source is a eclipsing binary system with a depth ≈ 0.4 mag in the optical, and period of ≈ 25 days. These two factors meant that the microlensing signal was unable to be detected or extracted.

It is worth mentioning that both the source and lens for both events are present in GEDR3 with updated astrometric solutions. This may improve the estimated mass inferences presented in this Chapter.

A PREDICTIVE MICROLENSING SEARCH BEYOND GAIA

Based on work originally published in [McGill et al. \(2019b\)](#).¹

We extend predictive microlensing event searches using the VVV and GDR2. We identify two events with maxima in J2019 that require urgent follow-up. First, we predict that the nearby M2 dwarf L 338-152 will align with a background source with a closest approach of 35_{-23}^{+35} mas on 2019 November 16_{-27}^{+28} days. This will cause a peak astrometric shift and photometric amplification of the background source of $2.7_{-1.5}^{+3.5}$ mas and $5.6_{-5.2}^{+143.2}$ mmag, respectively. This event should be astrometrically detectable by both HST and the Spectro-Polarimetric High-contrast Exoplanet REsearch (SPHERE) instrument on the VLT. Secondly, we predict that the likely K dwarf NLTT 45128 will lens a background source with a closest approach of $105.3_{-11.7}^{+12.2}$ mas on 2019 September 26_{-15}^{+15} days. This will produce a peak astrometric shift of $0.329_{-0.059}^{+0.065}$ mas. NLTT 45128 is only 3.6 mag brighter than the background source which makes it an excellent candidate for follow-up with HST. Characterization of these signals will allow direct gravitational masses to be inferred for both L 338-152 and NLTT 45128 with an estimated precision of $\approx 9\%$ and $\approx 13\%$, respectively.

5.1 INTRODUCTION

The only work to date to extend searches beyond GDR2 is [Nielsen & Bramich \(2018\)](#). They used a population of potential lenses consisting of nearby very low mass objects from PanSTARRS data release 1, and a background source population from GDR2. The rationale for doing this is to remedy GDR2's incompleteness for faint high proper motion objects. Using this complementary data set to GDR2, [Nielsen &](#)

¹ This work was carried out in collaboration with Leigh Smith (LS), Wyn Evans (WE), Vasily Belokurov (VB) and Philip Lucas (PL). The original project idea was entirely mine and I carried out the full search and analysis of events presented in this Chapter. LS provided the VVV Infrared Astrometric Catalogue reduction. WE and VB provided useful comments on early drafts of this Chapter. The referee for the published manuscript was Daniel Bramich, and PL suggested minor clarifying changes throughout the text.

[Bramich \(2018\)](#) found a further 27 predicted microlensing events over the next 50 years. In this study, we extend predictive microlensing by using a complementary background source population from the VVV ([Minniti et al., 2010](#)) with GDR2 lenses. This allows us to probe deeper (K_s -band ≈ 17 mag) into areas of high source density in the Galactic bulge and southern disc where microlensing events are most likely to happen (e.g. [Navarro et al., 2018](#)).

5.2 SEARCH FOR PREDICTED EVENTS

5.2.1 Selection and lens mass estimates

To select stars which have a high probability of lensing a background source, we take a high proper motion sample (> 150 mas/year) of stars from GDR2 that lie within the VVV footprint. To remove spurious high proper motion objects, we insist on high quality astrometric solutions. Specifically, we employ a magnitude dependent cut on the unit weight error of an object’s astrometric solution defined and suggested in appendix C, equation C1 of [Lindgren et al. \(2018\)](#). We also require that a lens has a positive parallax that is less than that of the closest star, Proxima Centauri ($0 < \varpi_L < 768.5$ mas). Finally, we require high significance parallax detection $\varpi_L/\sigma_{\varpi_L} > 10$, which ensures that a high precision lens mass can be extracted from the astrometric signal. This leaves a sample of 1915 lens stars ranging from ≈ 2 to 21 in G-band apparent magnitude, of which 1638 have measured $G_{BP} - G_{RP}$ colours. For the selection of the background source sample from VIRAC (described in Section 1.3.2), we require a strict quality cut. This is because spurious objects are likely to be found near bright objects. As we are selecting a source close to a likely bright high proper motion GDR2 lens, this unfortunately makes us prone to picking these spurious objects. We therefore require a source to be detected for > 6 consecutive calendar years in the best-seeing pointing covering the source each year. This leaves a sample of 436,903,339 source stars ($\approx 25\%$ of the total number of VIRAC sources).

We estimate the mass of each lens using the method suggested by [K18b](#). Specifically, we first classify each lens as either main sequence, red giant, white dwarf, or brown dwarf according to G , G_{BP} , G_{RP} magnitudes and parallax. If a lens has no G_{BP} , or G_{RP} measurements, it is assumed to be on the main sequence. For red giants, white dwarfs and brown dwarfs, we assume fixed masses of

$(1.0 \pm 0.5, 0.65 \pm 0.15, 0.07 \pm 0.03) M_{\odot}$, respectively. For main-sequence lenses, we use the [K18b](#)'s empirical mass–luminosity relations and assume a 10% error. After an event is found, we update the mass if a more accurate measurement derived using spectroscopy and stellar models is available from the literature.

5.2.2 Search method

To search for predicted events among the large number of possible pairings of GDR2 lenses and VIRAC sources, we only compute predicted signals for lenses that pass sufficiently close to background sources for detectable signals to be produced. To achieve this we use a method similar to [Proft et al. \(2011\)](#) and [K18b](#). Specifically, we match each lens with all sources that pass within a distance of 10 arcseconds of the predicted lens trajectory (as described in Section 2.4.4) over the next ~ 20 years (J2018–J2040). This leaves 7827 lens source pairs to be checked. The 10 arcseconds match distance chosen is more than sufficient to match all sources that are deflected by a detectable amount ($\delta_+ > 0.2$ mas - Eq. 32) by a typical nearby stellar lens with $\theta_E \approx 10$ mas.

Next, for each lens source pair, we find the time of minimum separation by numerically minimizing the angular separation between them (using Eq. (39) for the projected position of the lens and source). In order to find the global minimum, we first minimize the angular separation without parallax motion (which can be done analytically), and then use this as an initialization to the basin-hopping algorithm ([Wales & Doye, 1997](#)) implemented by SciPy (as described in Section 2.4.2). The basin-hopping algorithm combines global stepping with local minimization allowing the global minimum to be found. We set the temperature parameter for the basin-hopping algorithm (expected distance between local minima) to 0.5 years.

Using the estimated θ_E , we calculate the peak astrometric and photometric signals. Uncertainties are derived using Monte Carlo simulations. We draw 10^4 samples from Gaussian distributions consistent with the means and covariances reported by GDR2 and VIRAC for the astrometric parameters, and our lens mass estimates. VIRAC parallax measurement uncertainties are often at the level of a few mas, particularly in the vicinity of bright sources as is the case for all our candidates. Because sources in the VVV area are generally several thousand parsecs distant, many scatter below zero in measured par-

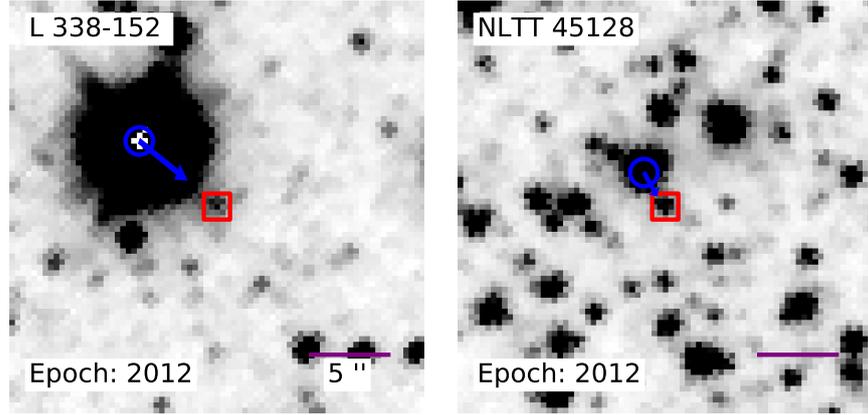


Figure 37: Stellar field around each of the candidate events. In each image, the blue circle and red square indicate the position of the lens and source at the image epoch, respectively. The blue arrow is the proper motion vector of the lens and length of 5 years of proper motion. The epoch of each image is indicated in the bottom left in Julian Years, and the SIMBAD lens name is in the top right. Both images are K_s -band cutouts from VVV.

allax. Hence, we draw from Gaussian distributions truncated at 0, which avoids unphysical, negative source parallax values.

For an event to be astrometrically detectable, we require a peak resolved astrometric shift $\delta_+ > 0.2$ mas (Eq. 32). For a photometric event, we require a resolved peak amplification of $A_+ > 1.0$ mmag (Eq. 19). This is in line with the current instrumentation detection capabilities (e.g. with HST, Gaia and SPHERE: S17; Kains et al., 2018; Rybicki et al., 2018; Z18). This leaves 14 candidate events with maximums in J2019. Visual inspection of Digitized Sky Surveys 2 and VVV K_s -band image data eliminated 12 events. In these cases, only the VIRAC source and not the lens could be definitively identified in the image. To avoid duplicating the efforts of other searches using GDR2 as background sources, we remove events where a VIRAC source lies within 1.0 arcseconds away from a GDR2 detection at epoch J2014.0. This cut has no effect, leaving two new candidate events.

5.3 CANDIDATE EVENTS

5.3.1 L 338-152

L 338-152 is a 7.6th magnitude (K_s -band) M2 dwarf (Rajpurohit et al., 2013) with a GDR2 parallax and total proper motion of 42.2 mas, 761 mas/year. We adopt the mass of $(0.51 \pm 0.06) M_\odot$ for L 338-152 esti-

mated by [Gaidos et al. \(2014\)](#). This mass was derived by fitting stellar models to L 338-152's spectrum and using empirical mass relations for M dwarfs. We note the mass obtained using [K18b's](#) method of $(0.43 \pm 0.04) M_{\odot}$ provides a reasonable estimate. We predict L 338-152 will lens a 15.6th magnitude (K_s -band) background source ([Fig. 37](#)). The closest approach of 35^{+35}_{-23} mas will occur on $J2019.872^{+0.076}_{-0.073}$ (or 2019 November 16^{+28}_{-27} days), with peak resolved shift and magnification of $2.7^{+3.5}_{-1.5}$ mas and $5.6^{+143.2}_{-5.2}$ mmag, respectively (see [Table 8](#) and [Fig. 38](#)).

5.3.2 NLTT 45128

NLTT 45128 is a little-studied, 10.8th magnitude (K_s -band) likely K dwarf star (identified from Gaia $G_{BP} - G_{RP}$ colour and G-band absolute magnitude). It has a GDR2 parallax and total proper motion of 16 mas and 327 mas/year respectively. We adopt a mass of $(0.29 \pm 0.03)M_{\odot}$ for NLTT 45128 derived from [K18b's](#) mass luminosity relations. With this in hand, we predict NLTT 45128 will lens a 14.5th (K_s -band) source ([Fig 37](#)) with a closest approach of $105.3^{+12.2}_{-11.7}$ mas on $J2019.732^{+0.040}_{-0.040}$ or (2019 September 26^{+15}_{-15} days). This will cause a resolved astrometric shift of the source of $0.329^{+0.065}_{-0.059}$ mas ([Table 8](#) and [Fig. 38](#)). As the closest approach of the lens and source is much larger than the Einstein radius ($u \approx 18 \gg 1$), there will be no detectable photometric signal.

5.3.3 Observational outlook

For both events, the background source is not present in GDR2. We thus consider other instruments that have achieved sub-mas single epoch astrometric precision and can observe these southern targets, namely HST and SPHERE.

[S17](#) demonstrated that, through single-epoch pointed imaging with the WFC3, HST can achieve a precision of $\sigma_{\text{HST}} \approx 0.13$ mas. This was attained for a background source ≈ 6.5 magnitudes fainter than the white dwarf lens. [S17](#) notes that the background source could only be resolved at separations > 450 mas for this lens source contrast ratio. The predicted minimum separation between Stein 2051B and the source was ≈ 203 mas, meaning that [S17](#) was only able to detect the tails of the deflection.

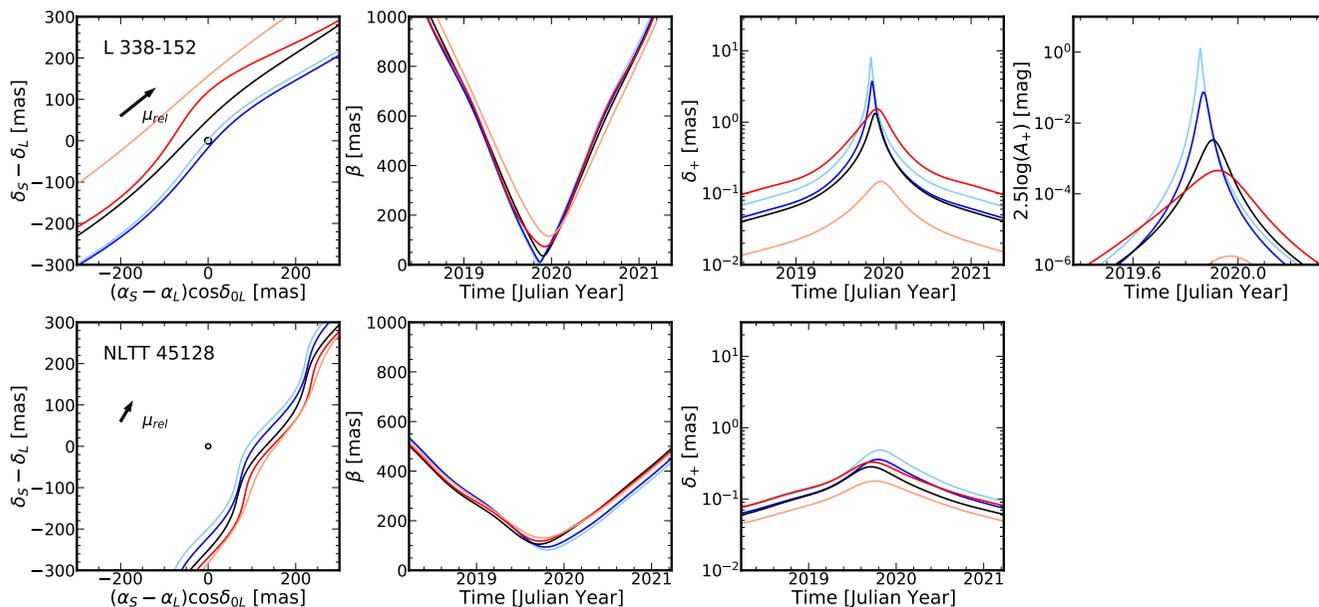


Figure 38: Predicted trajectories and signals for the microlensing event candidates. In each panel the light-blue, blue, black, red and light-red curves are example trajectories and signals corresponding to the the 2.3, 15.9, 50, 84.1, and 97.7 percentiles, respectively, of u_0 , calculated in the Monte Carlo simulations described in section 5.2.2. (Left to right) The first column shows samples of the predicted source trajectory in the lens rest frame. The median of the estimated angular Einstein ring is shown as a black circle. SIMBAD names for the lenses are also indicated for each event. The relative proper motion vector $\vec{\mu}_{rel} = \vec{\mu}_S - \vec{\mu}_L$ is shown with a length of 0.1 years of motion. The second column shows samples of the predicted separation of the lens and source around the event maximum. The third column shows samples of the predicted resolved astrometric shift of the source. Finally, the fourth column shows samples of the predicted photometric amplification for the only event with a detectable peak signal > 1 mmag.

Using the infrared dual imaging and spectrograph plus integral field spectrograph (IRDIFS) observing mode of SPHERE on the VLT, [Z18](#) achieved a single-epoch astrometric precision of $\sigma_{\text{VLT}} \approx 0.5$ mas. This was achieved for a background source ≈ 11 magnitudes fainter than the lens (Proxima Centauri). SPHERE was able to resolve Proxima Centauri and the background source over the duration of the event, for which the closest separation was ≈ 500 mas.

L 338-152 will lens a background source that is ≈ 8 mag fainter at a closest approach of ≈ 30 mas. This is a larger contrast ratio than the Stein 2051B event in [S17](#). Assuming the likely best-case with HST where L 338-152 and the background source could be resolved at ≈ 450 mas, the astrometric shift would be ≈ 0.24 mas which is $\approx 1.8\sigma_{\text{HST}}$. This means HST is only likely to detect the astrometric tails of this event. Detection of the photometric signal would require resolution of the lens and source at ≈ 30 mas separation, which, for this lens-source contrast ratio, is likely beyond HST's capabilities. However, L 338-152 is a bright enough target for good performance with SPHERE. The L 338-152 event has a more favourable lens source contrast ratio than the event successfully detected in [Z18](#) (8 compared to 11 mag difference). However, L 338-152 and the background source would need to be resolved at closer angular separations (≈ 30 compared to ≈ 500 mas). Recent work by [Claudi et al. \(2019\)](#), on the characterization of a faint low-mass companion of HD 142527, shows that for a contrast ratio of ≈ 8 magnitudes, IRDIFS observing modes can resolve separations of $\approx 50 - 100$ mas. Assuming that L 338-152 and the background source can be resolved at 75 mas, the astrometric deflection would be ≈ 1.41 mas, which is $\approx 2.8\sigma_{\text{VLT}}$. L 338-152 is not visible at low airmass (< 1.4) from the VLT at Cerro Paranal between September 2019 to the end of January 2020 (J2019.7-J2020.1). Unfortunately, this is at the maximum of the event at ≈ 2019 November 15th, and long enough that the short photometric signal will be missed. However, when L 338-152 becomes visible at low airmass after J2021.1, the background source will be separated by ≈ 100 mas with an astrometric shift of ≈ 1 mas which is $\approx 2\sigma_{\text{VLT}}$. We conclude that only the astrometric signal will be detectable with SPHERE.

NLTT 45128 will lens a background source only ≈ 3.6 mag fainter. This contrast ratio is considerably better than the event in [S17](#) (6.5 mag). This means HST will likely be able to resolve NLTT 45128 and the background source for separations $\ll 450$ mas. If this event is resolved over closest approach (≈ 105 mas), it will be detectable at the

Table 8: Details of the microlensing events. Lens and source astrometric parameters are on the IRCF and obtained from GDR2 and VIRAC, respectively. Lens and source initial positions are at epochs J2015.5 and J2014.0, respectively. K_s -band magnitude of the lens and source are from 2MASS and VIRAC respectively. Subscripts L and S represent parameters of the lens and source respectively. $T[\delta_+, A_+]$ are the Full-Width-Half-Maximums of the signals. () indicate uncertainty on the last digit. Median ± 34 percentiles are shown for the event parameters.

Lens	L 338-152	NLTT 45128
Type	M2	K
$M_L(M_\odot)$	0.51(6)	0.29(3)
$\theta_E(\text{mas})$	$10.4^{+1.8}_{-2.7}$	$5.94^{+0.20}_{-0.35}$
$\delta_+(\text{mas})$	$2.7^{+3.5}_{-1.5}$	$0.329^{+0.065}_{-0.059}$
$A_+(\text{mmag})$	$5.6^{+143.2}_{-5.2}$	-
$t_0(\text{Julian year})$	$2019.872^{+0.076}_{-0.073}$	$2019.732^{+0.040}_{-0.040}$
$\beta_0(\text{mas})$	35^{+35}_{-23}	$105.3^{+12.2}_{-11.7}$
u_{\min}	$3.5^{+4.0}_{-2.4}$	$17.9^{+2.7}_{-2.3}$
$T[\delta_+](\text{day})$	65^{+57}_{-36}	342.9^{+36}_{-32}
$T[A_+](\text{day})$	25^{+24}_{-15}	-
$\varpi_L(\text{mas})$	42.29(6)	16.47(5)
$\mu_{\alpha*L}(\text{mas/year})$	-586.9(1)	-158.5(1)
$\mu_{\delta L}(\text{mas/year})$	-484.55(5)	-286.42(7)
$\alpha_{\text{ref}L}(\text{deg})$	245.66683036(2)	264.32271595(1)
$\delta_{\text{ref}L}(\text{deg})$	-48.657646525(9)	-31.22567513(1)
$K_{sL}(\text{mag})$	7.64(2)	10.83(3)
$\text{id}_L(\text{GDR2})$	5941478335062267392	4055355782158028416
$\varpi_S(\text{mas})$	14(13)	-1(3)
$\mu_{\alpha*S}(\text{mas/year})$	18(6)	-4(1)
$\mu_{\delta S}(\text{mas/year})$	30(8)	-8(2)
$\alpha_{\text{ref}S}(\text{deg})$	245.665710(3)	264.3224710(7)
$\delta_{\text{ref}S}(\text{deg})$	-48.658284(5)	-31.2259846(9)
$K_{sS}(\text{mag})$	15.638(9)	14.452(2)

$\approx 2.5\sigma_{\text{HST}}$ level. NLTT 45128 is on the faint end of targets for good performance with SPHERE, and moreover, the maximum astrometric shift is $0.327 < \sigma_{\text{VLT}}$. We therefore conclude that WFC3 on HST is likely the best instrument for follow-up observation of the NLTT 45128 event.

To obtain rough estimates of the precision of the inferred mass from these events, we use Monte Carlo simulations. Inverting Eq. (32), the mass of the lens may be written in terms of δ_+ , β_0 , ϖ_L and ϖ_S . Assuming the deflection is measured at maximum for L 338-152 and NLTT 45128, we take δ_+ to be a Gaussian distribution centred on the median value reported in Table 8, with standard deviation of the measurement uncertainties of σ_{VLT} and σ_{HST} , respectively. We then take Gaussian distributions consistent with the GDR2 value for ϖ_L , and for β_{min} , ϖ_S , we centre on the medians reported in Table 8 but with standard deviations of 1 mas, and truncated at 0 for ϖ_S . Here, we assumed that the source’s astrometric solution can be refined by space-based observations (e.g. with HST at baseline after the event). Taking 10^5 samples for both events per epoch, neglecting covariances between the astrometric parameters, and assuming 8 epochs of data (as was achieved in S17) can be taken around maximum (with $1/\sqrt{8}$ error scaling over all epochs), we recover the mass of L 338-152 and NLTT 45128 to $\approx 9\%$ and $\approx 13\%$, respectively.

5.4 CONCLUSIONS

We have extended predictive microlensing searches beyond Gaia using the VIRAC catalogue derived from observations from the VVV. This allowed us to probe deep (K_s -band ≈ 17 mag) into regions of high source density in the Galactic bulge and southern disc and find events missed by previous searches. We identify two events with detectable signals, with maximums in J2019, that require immediate follow-up.

We predict that L 338-152, a M2 star, will lens a background source on 2019 November 16_{-27}^{+28} days. This will produce a resolved astrometric shift of $2.7_{-1.5}^{+3.5}$ mas and resolved photometric amplification of $5.6_{-5.2}^{+143.2}$ mmag. Despite the photometric signal of this event being undetectable with the currently available instrumentation, the astrometric shift of the background source is likely to be detectable by instruments on both HST and the VLT. We also predict that NLTT 45128, a likely K-dwarf, will lens a background source on 2019 September

26_{-15}^{+15} days. This will produce a maximum astrometric shift of the background source of $0.329_{-0.059}^{+0.065}$ mas. This event has a particularly favourable contrast ratio (the source is only 3.6 magnitudes fainter than NLTT 45128) and the shift will likely be detectable with HST. Characterization of these astrometric signals will likely allow direct mass determinations of L 338-152 and NLTT 45128 to $\approx 9\%$ and $\approx 13\%$ precision.

5.5 POST-CONCLUSION NOTE

To my knowledge, the two predicted microlensing events presented in this Chapter were not followed up. Despite this, the work in this Chapter has still provided useful and meaningful contributions. Specifically, this work has demonstrated searches solely using Gaia data for the background sources are likely incomplete for sources in the Galactic bulge and plane. Moreover, this work has demonstrated that VIRAC is of sufficient astrometric quality when combined with GDR2, to predict microlensing events with reasonable confidence. Overall, VIRAC could provide the means to fully open up the Galactic bulge and southern disc to predictive microlensing searches, a region of the sky off limits to Gaia for background sources.

IS THAT REALLY A PREDICTED MICROLENSING EVENT?

Based on work originally published in [McGill et al. \(2020\)](#).¹

Precision astrometry from GDR2 has allowed astronomers to predict 5,787 microlensing events, with 528 of these having maximums within the extended Gaia mission (J2014.5-J2026.5). Future analysis of the Gaia time-series astrometry of these events will, in some cases, lead to precise gravitational mass measurements of the lens. We find that 61% of events predicted during the extended Gaia mission with sources brighter than $G = 18$ are likely to be spurious, with the background source in these cases commonly being either a duplicate detection or a binary companion of the lens. We present quality cuts to identify these spurious events and a revised list of microlensing event candidates. Our findings imply that at least half of the predictable astrometric microlensing events during the Gaia mission have yet to be identified.

6.1 INTRODUCTION

The advent of astrometric data of an unprecedented precision and volume from GDR2 ([Gaia Collaboration et al., 2018](#)) reignited interest in predicting microlensing events. Searches by many studies resulted in precise predictions of 5,787 microlensing events occurring over the next century ([Bramich, 2018](#); [Mustill et al., 2018](#); [Klüter et al., 2018a,b](#); [Bramich & Nielsen, 2018](#); [McGill et al., 2019a](#)). In addition to predicting future microlensing events, events occurring over Gaia's observation baseline were presented with the view that they could be analyzed when Gaia releases time series astrometry (e.g. [Klüter et al., 2020](#)).

¹ This work was completed in collaboration with Andrew Everall (AE), Douglas Boubert (DB) and Leigh Smith (LS). The original project idea and initial investigation of spurious microlensing events was mine. AE produced Fig. 41, and DB did the event rate calculation in section 6.4, and produced Fig. 40. Understanding the case of G123-61 in section 6.3 was the result of discussions with Timo Prusti at the Gaia Helpdesk. The referee for the published paper Ulrich Bastian and LS provided useful suggestions clarifying parts of the text throughout.

Can we trust these predictions? How many of these events will Gaia observe? In this Chapter we examine these questions in detail.

6.2 THE PREDICTED EVENTS

We analyze the predicted microlensing events found by searches solely using GDR2 (Bramich, 2018; Mustill et al., 2018; Klüter et al., 2018a,b; Bramich & Nielsen, 2018; McGill et al., 2019a), giving us a total sample of 5,787 distinct events caused by 4,436 lenses. Although many of these studies predict some of the same events, there are key differences. Mustill et al. (2018) searched for photometric events caused by lenses within 100 pc over the next 20 years. Klüter et al. (2018a) presented two ongoing astrometric events which at the time required immediate follow up. Bramich (2018) presented a catalogue of photometric and astrometric events with maximums during the extended Gaia mission (J2014.5 - J2026.5). Klüter et al. (2018b) (hereafter K18b), presented a catalogue of predicted photometric and astrometric events with maximums between J2014.5 and J2065. Bramich & Nielsen (2018) presented a catalogue of photometric and astrometric events with maximums between the end of the extended Gaia mission and the end of the Century (J2026.5 - J2100.0). Finally, Chapter 4 (McGill et al., 2019a) presents two photometric events, which required immediate follow up in J2019. Each of these studies used different event detectability criteria, lens and source selection criteria, and lens mass estimates, which resulted in different sets of events. In most cases where two different studies found the same event, the predictions were consistent. Searches using GDR2 in combination with other catalogues have also occurred (Ofek, 2018; Nielsen & Bramich, 2018; McGill et al., 2019b) but we do not consider them further in this Chapter.

6.3 THE INTRIGUING CASE OF G123-61

At first glance, the predicted microlensing event by the lens G-123-61A (lens - GDR2 1543076475509704192, $G = 12.9$; source - GDR2 1543076471216523008, $G = 13.1$) looks like a promising candidate. This event, predicted by K18b, peaked on J2016.311 with a predicted maximum astrometric deflection of ≈ 0.45 mas. Klüter et al. (2020) conclude that, with time series astrometry from Gaia, this event should permit a mass determination of G-123-61A to 24% precision.

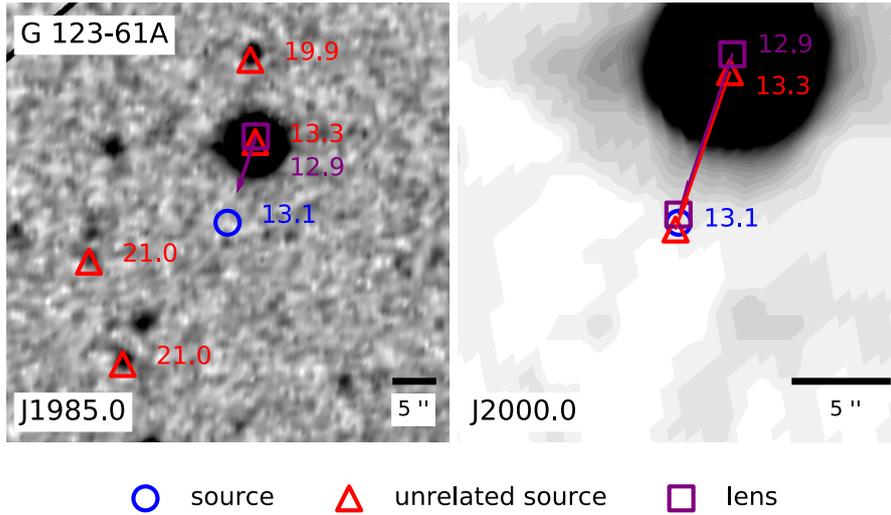


Figure 39: Images of the region surrounding the predicted microlensing event caused by the lens G123-61A showing that the source is not present. Arrow shows the direction of the proper motion of the lens. Annotated text gives the GDR2 G magnitudes. North is in the upwards vertical direction in both images. **Left:** DSS-blue image (epoch \approx J1985.0) with positions projected to the image epoch if the source has a GDR2 proper motion, otherwise the position is at the GDR2 reference epoch of J2015.5. **Right:** Two Micron All Sky Survey (2MASS) K_s -band image (epoch \approx J2000.0) of the event with positions shown at both the GDR2 reference epoch of J2015.5 and the 2MASS image epoch.

The quality of this estimated constraint is largely due to the high apparent brightness of the source. Crucially, this would allow Gaia to obtain high precision single epoch measurements of the astrometric deflection (Rybicki et al., 2018). Moreover, the apparent small lens-source contrast ratio provides favourable conditions to resolve the lens and source around their closest alignment (≈ 150 mas), which may help with the deflection detection.

Fig. 39 (left) shows a Digitized Sky Survey 2 (DSS2) cutout image of the field around the G-123-61A event. The positions are the projections of the GDR2 astrometric solution back to the epoch of the DSS2 image (\approx J1985). The source in the event does not have a GDR2 proper motion and therefore cannot be projected back to the DSS2 epoch – its position is at the GDR2 epoch of J2015.5. It is clear from Fig. 39 that a thirteenth magnitude source is not present in the image at the GDR2 J2015.5 position which was used in the event’s prediction, nor is there an unaccounted thirteenth magnitude source elsewhere in the image.

K18b predicts a second microlensing event with GDR2 15430764-71216523008 as the source by the lens G-123-61B (GDR2 15430764755-14008064, $G = 13.3$). This event peaked on J2014.76 with a predicted

maximum astrometric deflection of ≈ 1 mas. Klüter et al. (2020) predict this will enable Gaia to measure G-123-61B's mass to 37% precision. The lens stars for these two predicted events (G-123-61A and G-123-61B) have similar GDR2 positions, proper motions and parallaxes, and have consequently been assigned to a binary system by SIMBAD. Whilst there is no reference in the literature to G-123-61 being a binary prior to GDR2, there is some weak indication of it. The radial velocity from Gaia Collaboration et al. (2018) of G-123-61A is -28.69 ± 2.80 km s⁻¹ based on eight measurements, whereas Reid et al. (1995) measures the radial velocity to be 3.5 ± 10 km s⁻¹. These values are mildly inconsistent and provide weak evidence that G-123-61 is indeed a binary, and suggests that the two GDR2 lens detections could be genuine. Regardless of this, Fig. 39 shows that for both of these events, the source star is suspect. G-123-61B can be seen in Fig. 39 as the 13.3 magnitude GDR2 source marked as unrelated to the G-123-61A event.

A troubling aspect of these two events is that at J2015.5, the source lies in between the two suggested lenses G-123-61A and B. Furthermore, both the lenses and the source are similar magnitudes and the source ($G_{BP} - G_{RP} = 2.67$) has an almost identical colour to the lens G-123-61A ($G_{BP} - G_{RP} = 2.66$, the other lens does not have GDR2 colour photometry). The rest of this paragraph is an abridged explanation provided by the Gaia Helpdesk in response to our queries. Using the scanning law from Boubert et al. (2020) and the astrometric gaps during GDR2 provided by the Gaia Data Processing and Analysis Consortium (DPAC)², we were able to predict that this system should have been observed 65 times during the 22 months of GDR2. Each of these observations should have resulted in a detection of each component of the thirteenth magnitude binary (the detection probability for thirteenth magnitude stars is close to 100%, Boubert & Everall, 2020), giving a total of 130 detections. It is not a coincidence that the total detections (MATCHED_OBSERVATIONS³) of G-123-61A (54), G-123-61B (42) and the source (34) also sums to 130. We deduce that when the Gaia DPAC clustering algorithm was merging detections into stars, a fraction of the 65 detections of the two components of the binary were mistakenly combined into a spurious third star. It is likely that the proper motion of the system being in the same direction as the

² <https://www.cosmos.esa.int/web/gaia/dr2-data-gaps>

³ Text in TYPEWRITER font in this Chapter refer to specific column names in the GDR2 data model which can be accessed at https://gea.esac.esa.int/archive/documentation/GDR2/Gaia_archive/chap_datamodel/sec_dm_main_tables/sssec_dm_gaia_source.html

separation of the two components made this a difficult system for the clustering algorithm. The donated astrometric detections from the two components of the binary will lead to a problematic astrometric solution for the spurious source, thus explaining why the source does not have an astrometric solution despite being a thirteenth magnitude star. We can thus conclude that the source is spurious and that these two predicted microlensing events did not occur.

While we expect cases like that of G-123-61 to be rare, the trouble does not end there. Inspection of imaging around several other events revealed further missing sources, with one example being the event caused by the lens LP 701-45 (lens - GDR2 2610954226042154624, $G = 14.8$; source - GDR2 2610954226041533696, $G = 17.3$) predicted by [K18b](#) to peak on J2022.36 with a deflection ≈ 0.1 mas. The Gaia uncertainties on the right ascension and declination of the source are highly degenerate with a correlation of 99.8%, suggesting that the 2D astrometric pipeline was attempting to fit points lying along a line. We suspect that the source in this event is truly a binary companion of the lens, explaining why it does not appear in legacy DSS2 imaging. In many cases we were not able to tell by visual inspection of the DSS2 imaging whether the source was visible at the J2015.5 position because the lens was not well enough separated at \approx J1985, but our suspicions were sufficiently raised that we decided to investigate further.

6.4 DIAGNOSTICS FOR MICROLENSING CANDIDATES

Typical predictable astrometric microlensing events are caused by nearby, and therefore bright and high proper motion, lenses and more distant, and therefore usually fainter, background sources. This is because small observer-lens distances cause larger astrometric signals ([Dominik & Sahu, 2000](#)), and high proper motion objects are more likely to align with a source over a given time. With this in mind our naive expectation of a source-lens pair is that the lens should be a bright star and that the source should be a randomly picked star in the background (and thus resemble a ‘typical’ star in GDR2). If the properties of the source are unusual or are similar to that of the lens, then we should question whether the predicted microlensing source-lens pair is real.

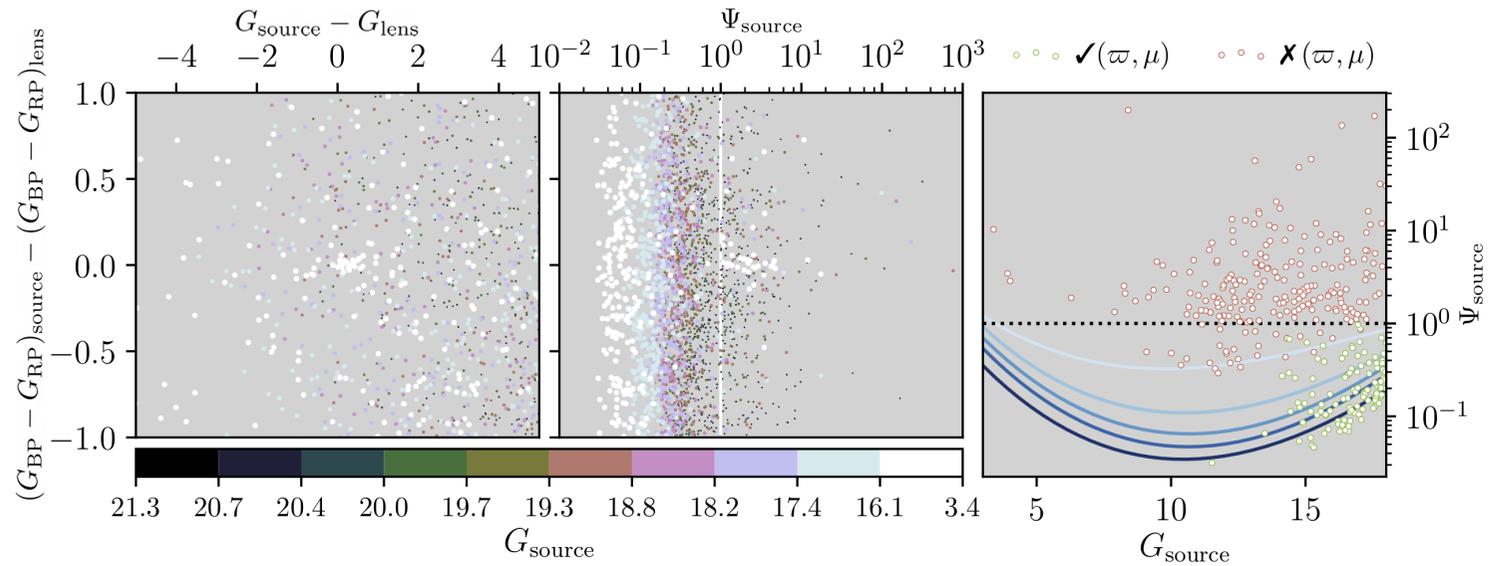


Figure 40: Many of the predicted microlensing events have sources with suspect properties. **Left:** The magnitude and colour difference between the source and lens for the microlensing events predicted with GDR2 astrometry, with the size and colour of the marker indicating the source magnitude. There is a cluster of photometrically identical lens-source pairs. **Middle:** Same schema as the left panel but showing the astrometric quality indicator Ψ for the source, as defined by Eq. 66, with $\Psi \leq 1$ being required for the source to have a parallax and proper motion in GDR2. The sources in the photometric cluster are all missing these quantities. **Right:** For events peaking prior to J2026.5, we show the magnitude of the source versus Ψ , with those sources having/missing GDR2 parallaxes and proper motions circled in green/red. The blue lines gives the 50%, 80%, 90%, 95% and 98% percentiles of Ψ for all stars in GDR2 binned by G magnitude (smoothly interpolated for illustrative purposes). Most of the sources in predicted microlensing events peaking before J2026.5 lie above the 98% line.

In Fig. 40 (left) we show the difference in colour between the source and lens versus the difference in their magnitudes. Surprisingly, there is a cluster of lens-source pairs with bright sources where the source colour is within 0.1 mag of the lens colour. Given that the most observable microlensing events are those with bright sources, we decided to investigate the astrometric properties of these sources to see if they are consistent with being genuine stars.

Stars only have reported parallaxes and proper motions in GDR2 if they satisfy the three criteria given by [Lindgren et al. \(2018\)](#):

$$\begin{cases} \text{mean magnitude } G & \leq 21, \\ \text{VISIBILITY_PERIODS_USED} & \geq 6, \\ \text{ASTROMETRIC_SIGMA5D_MAX} & \leq (1.2 \text{ mas}) \times \gamma_{18}(G), \end{cases} \quad (65)$$

The function $\gamma(G) = \max[1, 10^{0.2(G-18)}]$ is flat for $G \leq 18$ and then transitions to exponential growth. The quantity `VISIBILITY_PERIODS_USED` is the number of time-resolved clusters of detections used in the astrometric pipeline and is required to be at least six to ensure a long enough baseline for the astrometric solution. The quantity `ASTROMETRIC_SIGMA5D_MAX` is the square root of the largest singular value of the scaled 5×5 covariance matrix of the astrometric parameters, and so is comparable to the semi-major axis of a position error ellipse ([Lindgren et al., 2018](#)). The third cut can be interpreted as requiring that the astrometric uncertainty should not be unusually large for a star of that magnitude. We define the quantity

$$\Psi = \frac{\text{ASTROMETRIC_SIGMA5D_MAX}}{(1.2 \text{ mas}) \times \gamma(G)}, \quad (66)$$

such that if $\Psi > 1$ then the source will fail the third cut. We show Ψ versus the source-lens colour difference in the middle panel of Fig. 40. The cluster of source-lens pairs with identical photometry in the left panel stands out in the middle panel, with all of these sources having $\Psi > 1$. This is highly unusual for sources brighter than $G = 16.1$ – in Fig. 40 most of these sources have $\Psi < 0.1$ – and so we can conclude that these sources are atypical and thus concerning. All of the events in the cluster were identified by [K18b](#) and were predicted to have their peak prior to J2026.5. We emphasise that the left and middle panels of Fig. 40 only show the sources with colour photometry. There is a few-fold larger group of sources without colour photome-

try that have $G_{\text{source}} < G_{\text{lens}}$, $\Psi \gtrsim 0.3$ and a peak prior to J2026.5, and thus the cluster in Fig. 40 is only a subset of the phenomenon. We conjecture that the reason only those sources in the extended cluster with $G_{\text{source}} \approx G_{\text{lens}}$ have colour photometry is the colour excess cut applied by DPAC, $E = (F_{\text{BP}} + F_{\text{RP}})/F_{\text{G}} < 5$. The colour photometry measured for these sources is dominated by the flux from the nearby lens and so $F_{\text{BP/RP,source}} \approx F_{\text{BP/RP,lens}}$, and thus $E_{\text{source}} \gg 5$ unless the source and lens have similar magnitudes.

In Fig. 40 (right) we show the sources with $G < 18$ and a predicted astrometric microlensing signal peak before J2026.5. Most of these sources are astrometrically unusual compared to typical GDR2 sources, having an `ASTROMETRIC_SIGMA5D_MAX` in the top 2% of stars at that magnitude. Those sources with extreme astrometric errors are mostly those without a 5D astrometric solution⁴, while those with 5D astrometry are representative of the GDR2 source catalogue. We propose that events with sources brighter than $G=18$ should only be considered reliable if the source has a published 5D astrometric solution.

⁴ A source with 5D astrometry has positions, proper motions and parallax measured. A source without 5D astrometry just has a position measured.

Two factors determine whether a bright source will have a published 5D astrometric solution. Firstly, fewer detections will reduce the likelihood of having the requisite six visibility periods as well as making it more likely that $\Psi > 1$, because fewer measurements constraining the astrometric solution will increase the astrometric uncertainty. For each source with $G \leq 18$ and a microlensing peak prior to J2026.5, we calculated the ratio of astrometric detections k (`ASTROMETRIC_MATCHED_OBSERVATIONS`) to the predicted number of astrometric observations n computed using the scanning law of [Boubert et al. \(2020\)](#) and the published gaps from DPAC. The 1σ interval $k/n \in (22.9, 60.0)\%$ for the sources without 5D astrometry is significantly lower than the 1σ interval $k/n \in (65.0, 95.8)\%$ for the sources with 5D astrometry. Secondly, an increase in either the astrometric uncertainty of the centroiding of individual detections or the scatter between the centroids around the single source astrometric fit will increase the reported astrometric error. Under the assumption of independent Gaussian errors with the size of Gaia’s AL precision σ_{AL} , the reported inverse covariance matrix (Λ) for a source with a two parameter solution is approximately ([Everall et al., 2021](#))

$$\Lambda \approx \sum_{i=1}^{\nu} \frac{1}{\sigma_{\text{AL}}^2} \begin{bmatrix} \sin^2 \phi_i & \sin \phi_i \cos \phi_i \\ \sin \phi_i \cos \phi_i & \cos^2 \phi_i \end{bmatrix}. \quad (67)$$

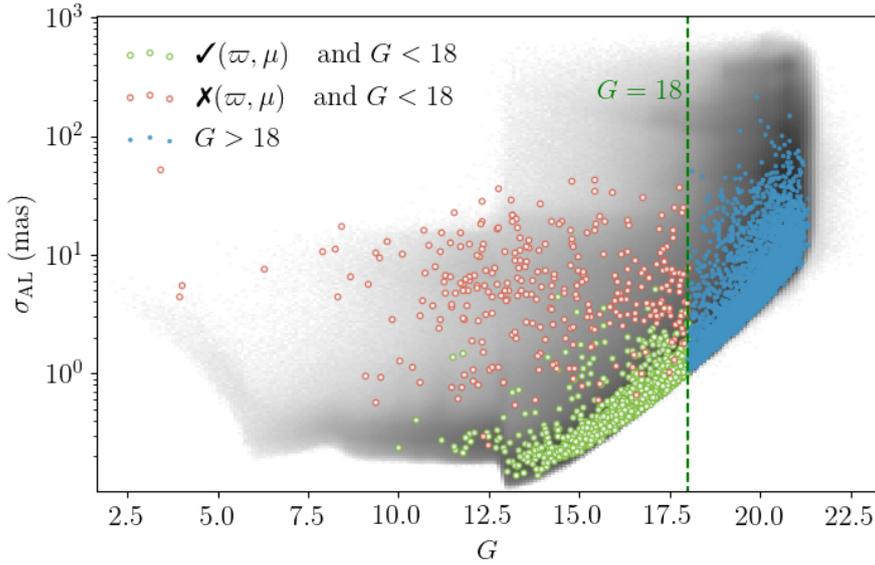


Figure 41: The well-behaved microlensing candidate sources brighter than $G = 18$ (green points) have AL astrometric errors per observation that are typical of the GDR2 population (grey-scale background, log-normalized) whilst spurious source candidates (red points) have systematically higher AL error. At dimmer magnitudes, the expected position error of well-behaved sources is high enough that the all event sources (blue points) trace the bulk distribution.

Here ν is `ASTROMETRIC_N_GOOD_OBS_AL`, and ϕ_i is the scan angle for a given measurement defined from the local north direction. Under the further assumption that $\Lambda = \text{diagonal}(1/\sigma_\alpha^2, 1/\sigma_\delta^2)$ is a diagonal matrix, and σ_α and σ_δ are the reported errors on the right ascension and declination of the source, we can estimate σ_{AL} for a source given the reported error on its position. This is achieved by summing the diagonal elements of Eq. (67). Rearranging for σ_{AL} , the estimated typical centroiding uncertainty is the harmonic mean of the equatorial positional uncertainties,

$$\sigma_{\text{AL}} = \sqrt{\nu / (1/\sigma_\alpha^2 + 1/\sigma_\delta^2)}. \quad (68)$$

Fig. 41 shows that sources without a 5D astrometric solution have enhanced centroid error compared to both the predicted microlensing sources with 5D astrometry and to the bulk of GDR2. We note some of the sources at magnitudes dimmer than $G = 18$ are likely to be spurious events. Nevertheless, these are difficult to distinguish from the main population as astrometric error increases with magnitude.

Requiring 5D astrometry resolves another troubling property of the current sample of predicted microlensing events: the microlensing

rate appears to peak at present day. In Fig. 42 we show the number of microlensing events with $G_{\text{source}} < 18$ predicted per year by K18b. We restrict ourselves to K18b because it is easier to interpret the rate if we only need to consider one set of selection cuts. Our expectation is that there should be a fairly constant rate of microlensing events, but, as noted by K18b, there should be a deficit of events around the GDR2 epoch of J2015.5 because the lens and source are less likely to be resolved by Gaia at their point of closest approach. Paradoxically, the rate of predicted microlensing events peaks either side of a dip at J2015.5. This is strong evidence to suggest that these events are unlikely to be real. If we remove sources without 5D astrometry then the microlensing event rate matches our expectations, with that cut predominantly removing sources with peaks near to J2015.5. The likelihood of observing N microlensing given an underlying rate parameter λ is given by the Poisson distribution,

$$p(N|\lambda) = \frac{\lambda^N e^{-\lambda}}{N!}. \quad (69)$$

We use the number of events per year between J2034.5 and J2055.5 to infer that the posterior Poisson rate (λ) of K18b-type astrometric microlensing events is $16.0^{+0.9}_{-0.8}$ events per year. Here we assumed the likelihood in Eq. (69) and an inverse square root Jeffrey's prior⁵ on λ , and obtained posterior samples using the ensemble sampler (emcee Foreman-Mackey et al., 2013; Goodman & Weare, 2010). Marginalising over this rate, we find that there should be 191^{+18}_{-17} events during the twelve year horizon of the extended Gaia mission, only 85 of which found by K18b meet our criteria for reliable identification. Over half of the astrometric microlensing events that will be detected by Gaia are yet to be identified.

Not all of the events during the Gaia mission, predicted by K18b, will result in useful mass measurements of the lens. Filtering events by our new astrometric cut changes the outlook for astrometric lensing events with Gaia. Of the 513 events predicted by K18b to peak between J2014.5 and J2026.5, only 260 have a source with 5D astrometry. Bright sources suffer an even higher attrition rate with only 85 of the 227 events with $G_{\text{source}} < 18$ surviving. This will impact the prospects of high-precision lens mass measurements with Gaia. Klüter et al. (2020) identified that 62 of their predicted single lens-source events could lead to mass measurements of the lens with precision better than 100%, but we find that only 28 of these are likely to occur. All

⁵ A Jeffrey's prior is an uninformative prior which is calculated as the square root of the determinant of the fisher information matrix. For the Poisson likelihood in Eq. 69 the Jeffrey's prior is $p(\lambda) \propto \frac{1}{\sqrt{\lambda}}$.

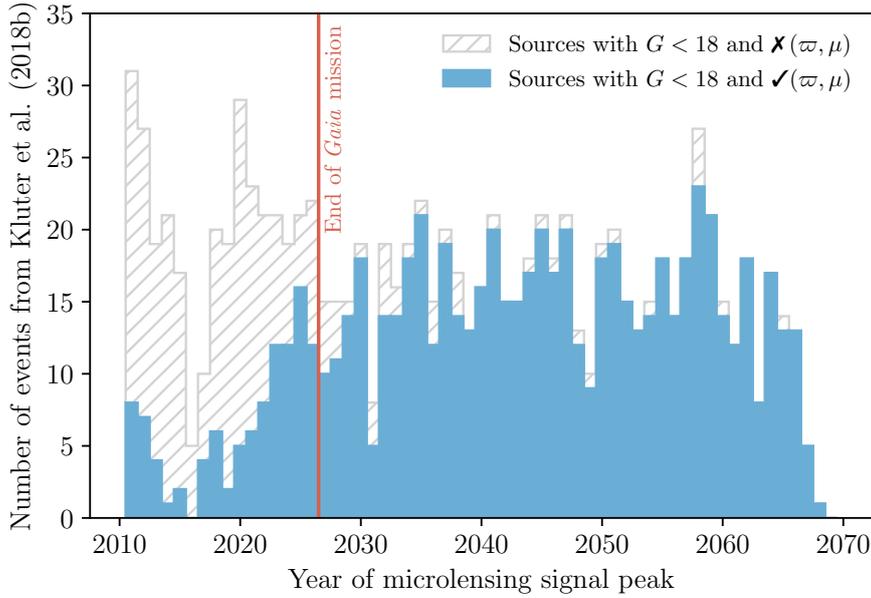


Figure 42: The number of microlensing events predicted per year by [K18b](#). If we do not apply any additional quality cuts then the rate peaks near present day. Applying our astrometric quality cut $\Psi \leq 1$ removes most of the predicted events during the horizon of the extended Gaia mission.

seven events with an estimated precision on the mass of the lens below 10% with the full ten year Gaia data are unlikely to occur. Notably the 5.4% mass estimate of 75 Cancri is unlikely to be realized. This spectroscopic binary system already has mass estimates of the two components accurate to three significant figures ([Konacki et al., 2010](#)), so would have been an ideal test bed of microlensing techniques. We conclude that many of these spurious events are likely caused by the source star being a detection of a binary companion of the lens.

The reliability of [K18b](#)'s predictions increases significantly beyond J2026.5. Requiring 5D astrometry for sources with $G < 18$ only eliminates 46 of the total 3,221 events. All of the events predicted by [Klüter et al. \(2018a\)](#), [Mustill et al. \(2018\)](#) and [McGill et al. \(2019a\)](#) (Chapter 4) survive. Only 2 of the 76 events predicted to happen over the extended Gaia mission by [Bramich \(2018\)](#) are eliminated. Finally, 37 of the 2,509 events predicted to happen between J2026.5-J2100.0 by [Bramich & Nielsen \(2018\)](#), are eliminated by our cuts. The bulk of the likely spurious events occur around the GDR2 reference epoch (J2015.5) and therefore make up a large fraction of the events predicted to happen during the extended Gaia mission.

Inspection of the DSS2 imaging data of the two events, predicted by [Bramich \(2018\)](#) but eliminated by our cuts, reveal that the source

and lens are clearly present and thus that these events are likely real. To ensure that this is a rare occurrence and our proposed quality cuts do not eliminate many real events, three of the authors independently eyeballed the 231 events with $G < 18$ and a predicted peak between J2014.5 and J2026.5. We used the ESAsky⁴ online observation tool with images from DSS2, 2MASS, Sloan Digital Sky Survey DR9 and AllWISE to determine whether an event is plausible. Of the 89 events for which the source has a 5D astrometric solution, we find that 87 are plausible, implying that the cut has a false positive rate of 2%. We classify as plausible 17 of the 142 events which fail the cut, implying a false negative rate of 12%. If a pure sample is required, we additionally recommend only considering events where both the source and lens have 2MASS detections. This leaves no false positives whilst increasing the false negative rate to 27%. We note that for the two events we eliminated from Bramich (2018), both the source and lens have 2MASS detections.

6.5 CONCLUSIONS

We critically analyze the fidelity of predicted microlensing events extracted from GDR2. We find that a significant portion of the bright events ($G_{\text{source}} < 18$) which are promising candidates for detection with Gaia are likely not genuine. This is demonstrated with a case study of G123-61, a high quality candidate for which two lenses are predicted to pass over the same source. Comparing with DSS and 2MASS observations, we find that the source in both events is almost certainly not real and is generated in Gaia from misclassified observations of the two lens objects. We propose that sources brighter than $G = 18$ should only be considered if they have a published 5D astrometric solution. We demonstrate that this cut increases the reliability of the sample of microlensing events and significantly changes the outlook for measuring precise lens masses with future Gaia data. We recommend this cut for all microlensing event searches with Gaia. In Table 9 we list all of the events predicted using GDR2 by each of the studies we considered, with results of the visual inspection described above, 2MASS lens and source IDs where available and Gaia 5D astrometry flags. On the positive side, our findings imply that at least half of the astrometric microlensing events during the Gaia extended mission are yet to be identified.

⁴ <https://sky.esa.int/>

Table 9: Diagnostic data for all of the microlensing events predicted using GDR2 astrometry, sorted by the magnitude of the lens. The full table along with column descriptions are available in the supplementary material associated with this thesis.

Lens _{GDR2} ID	Source _{GDR2} ID	ψ	Visually confirmed	Study	t_{\min} [Julian year]	G_{source} [mag]	...
5826168461827543168	5826168461827531520	0.238	N/A	K18b	2038.690	15.714	...
4269932382606282112	4269932382603752192	17.347	N/A	K18b	2015.465	14.046	...
2195115561163064960	2195117038631812736	0.051	N/A	K18b	2057.307	15.377	...
425040000951479424	425039245037744256	1.548	N/A	K18b	2052.633	20.103	...
425040000951479424	425039245037256448	0.508	N/A	K18b	2053.394	19.148	...
425040000951479424	425039245037886208	0.440	N/A	K18b	2061.172	20.418	...
425040000951479424	425039245037254016	3.200	N/A	K18b	2050.634	20.063	...
425040000951479424	425039245037258880	0.257	N/A	K18b	2057.127	19.802	...
4529285391522200320	4529285391531266304	10.295	N/A	K18b	2014.240	3.395	...
643819484616249984	643819484615911552	197.693	N/A	K18b	2014.823	8.402	...
1964791583368232832	1964791583375277312	1.886	N/A	K18b	2013.663	6.287	...
1193030490492925824	1193030486200630912	1.317	N/A	K18b	2025.656	16.779	...
2022881668933780096	2022881668925464704	2.526	N/A	K18b	2033.608	11.135	...
436648129326142336	436647751365040512	0.705	N/A	K18b	2055.237	19.734	..
2198483880674251136	2198483876385138816	0.098	N/A	K18b	2063.640	16.38	...
2198483880674251136	2198483880687228416	0.861	N/A	K18b	2045.119	17.974	...
...

A MICROLENSING SEARCH OF 700 MILLION VVV LIGHT CURVES

Based on work originally published in Husseiniova et al. (2021).¹

The VVV survey and its extension have been monitoring about 560 square degrees of sky centred on the Galactic bulge and inner disc for nearly a decade. The photometric catalogue contains of order 10^9 sources monitored in the K_s band down to 18 mag over hundreds of epochs from J2010 to J2019. Using these data we develop a decision tree classifier to identify microlensing events. As inputs to the tree, we extract a few physically motivated features as well as simple statistics ensuring a good fit to a microlensing model both on and off the event amplification. This produces a fast and efficient classifier trained on a set of simulated microlensing events and cataclysmic variables together with flat baseline light curves randomly chosen from the VVV data. The classifier achieves 97% accuracy in identifying simulated microlensing events in a validation set. We run the classifier over the VVV data set and then visually inspect the results which produces a catalogue of 1959 microlensing events. For these events, we provide the Einstein timescale via a Bayesian analysis. The spatial dependence on recovery efficiency of our classifier is well characterized, and this allows us to compute spatially resolved completeness maps as a function of Einstein timescale over the VVV footprint. We compare our approach to previous microlensing searches of the VVV. We highlight the importance of Bayesian fitting to determine the microlensing parameters for events with surveys like VVV with sparse data.

¹ This work was completed in collaboration with Andrea Hussuenova (AH), Leigh Smith (LS) and Wyn Evans (WE). The original project idea was mine. AH investigated an early prototype version of the machine learning classifier as part of a summer project. The machine learning classifier used and applied in this Chapter was developed by myself. LS provided the maps in Figs. 49 and 50. WE provided some of the introductory text. The anonymous referee for the original manuscript suggested clarifying changes to the text throughout. Overall, I did $\approx 80\%$ of the work presented in this Chapter.

7.1 INTRODUCTION

Historically, microlensing projects were amongst the first to acquire large photometric data sets and systematically search through them for very rare events (e.g. [Aubourg et al., 1993](#); [Alcock et al., 2000b](#); [Calchi Novati et al., 2005b](#)). The probability that a star is photometrically microlensed, or the all-sky source-averaged optical depth to microlensing is very low $\sim 10^{-7}$ (e.g. [Belokurov & Evans, 2002](#)), and so the events are much less numerous than stellar variability.

Some microlensing events deviate from the standard form, or can be predicted ahead of time enabling degeneracies to be wholly or partially lifted (e.g., [Gould, 1994](#); [Bennett, 1998](#); [Evans, 2003](#); [McGill et al., 2019a](#)), though these exotica can be hard to pick out. In this Chapter, we direct our attention exclusively to standard PSPL photometric events. Specifically, photometric events described by the rectilinear trajectory in Eq. (20), and that have amplification,

$$A_{\text{obs}}(t) = f_{\text{bl}}A(t) + (1 - f_{\text{bl}}). \quad (70)$$

Here $f_{\text{bl}} = F_S/(F_B + F_S) = 1/(1 + g)$, is the blending parameter.

As microlensing traces the mass of the Galaxy rather than the light, it provides a unique and complementary probe to other surveys of the Milky Way. Given substantial numbers of PSPL events, properties of the source and lens populations can be inferred, providing the selection function of the survey is well characterized.

Bulge microlensing surveys in the optical hark back to the very beginnings of the subject ([Udalski et al., 1992](#); [Alcock et al., 1995](#)). The high number of background stars provides ample sources whilst the density of disc and bulge stars means that the optical depth to microlensing is high at $\sim 10^{-6}$. OGLE ([Udalski et al., 1997, 2015](#); [Wyrzykowski et al., 2015](#)) has heroically maintained a decades long campaign to monitor events towards the centre of the Galaxy in optical passbands at high cadence. The most recent results from the OGLE-IV campaign provide measurements of the optical depth and event rate across the Galactic bulge and disc based on nearly 9000 events ([Mróz et al., 2019, 2020a](#)). However, there are parts of the bulge that are heavily obscured by extinction, hampering searches in optical passbands. It was realized early that photometry in the NIR overcomes this obstacle and can provide for many more locations throughout the bulge to search for events ([Gould, 1994](#)).

In order to search large surveys, automatic classification of microlensing events is required. An early attempt was made by [Belokurov et al. \(2003\)](#) using neural networks, but the subject has come of age in recent years. First, [Price-Whelan et al. \(2014\)](#) drew attention to the usefulness of the von Neumann statistic which is the mean square of successive differences divided by the variance of data points in the light curve. This is very fast to compute, especially compared to the process of fitting a PSPL light curve. They found that it gave excellent performance in identifying simulated microlensing events injected into real light curves from the Palomar Transient Factory ([Ofek et al., 2012](#)), which have irregular temporal sampling. Then, [Wyrzykowski et al. \(2015\)](#) used a random forest classifier to extract microlensing events for OGLE-III, finding 3700 unique events from over 150 million light curves in data taken between J2001-J2009. After preliminary cuts removing obvious variable stars and other contaminants that did not possess single-humped light curves, [Wyrzykowski et al. \(2015\)](#) were left with 48112 light curves for further classification. They then used a random forest classifier for the last stage. This took as input 27 features extracted from the primarily I-band, but also some V-band, light curves. This final step extracted the 3700 unique events – including 1409 events that had not been detected before in real time by the OGLE early warning system. This was the first study to harness the power of machine learning algorithms to find rare microlensing phenomena and to derive the detection efficiency. Then [Wyrzykowski et al. \(2016\)](#) used a random forest classifier to identify microlensing events showing annual parallax signals also from OGLE-III.

[Godines et al. \(2019\)](#) again used a random forest classifier, but now with 47 statistical features extracted from light curves. They included variable stars, cataclysmic variables (CVs) as well as constant light curves in their training set so that the machine learning algorithm is tasked with all the work and there is no preliminary filtering. They trained the classifier using the OGLE-II microlensing data set, and tested it on Palomar and Zwicky Transient Factory survey data. Finally, we note that the problem of automatic identification of microlensing events in which the system is not PSPL, sometimes showing substantial deviations from the standard [Paczynski \(1986b\)](#) curve, has also begun to receive attention ([Mróz, 2020](#); [Zhang et al., 2021](#); [Khakpash et al., 2021](#)).

This Chapter is arranged as follows. In Section 7.2, we provide an overview of the characteristics of the simulated light curves that are used to populate the training set and build completeness maps. Section 7.3 describes the construction of the classifier, the choice of the features used to train and tune it as well as its performance on the validation data. The results obtained when applied to the VVV data, together with the completeness maps, are given in Section 7.4. The catalogue of 1959 microlensing events, together with their fitted parameters, is presented in Table 13. Finally, Section 7.5 compares the properties of our sample of microlensing events with the earlier work of Navarro et al. (2018, 2020a,b,c), and a brief outline of the inference on t_E , for our sample of events, is presented in Section 7.6.

7.2 DATA

We used the photometric time-series data from the VVV as described in Section 1.3.2.

7.2.1 *Simulated light curves*

In order to evaluate an algorithm for discrimination between microlensing and not microlensing, we adopt a simulation and data-driven approach. This requires large samples of microlensing, not microlensing, variable, and non-variable light curves.

A sample of non-variable events is straightforward to produce as a random selection of light curves contains only a small fraction of sources with significant real variability in VVV data. The production of sufficiently large samples of variable events requires simulation. Godines et al. (2019) identified CVs, RR Lyrae, and Cepheid variables as likely contaminants in searches for microlensing events. Due to the long baseline of our data, and hence good phase coverage, we do not anticipate significant contamination from RR Lyrae or Cepheid variable stars in our search; there is, however, potential for contamination by CVs due to the irregular observing pattern and large gaps in coverage between observing seasons. We therefore restrict our simulation of not microlensing variables to CVs only.

It is important that our simulated events adopt realistic observing cadence, measurement uncertainty, outlier incidence, and scale, spatial, and brightness distributions. These are necessary for an accurate evaluation of both the performance of our selection criteria and the

overall completeness, which we expect to vary as a function of sky position and brightness due to the variable nature of the underlying survey.

The most straightforward approach to meeting the above requirements is to perform a catalogue-level simulation (e.g. Mróz et al., 2019; Wyrzykowski et al., 2015, 2009). Specifically, we take a random selection of real light curves and inject signals into them. This naturally adopts the true observing cadence, spatial and brightness distributions, and with some care outlying data points can also be preserved. The response of the brightness measurement uncertainty to the signal amplification is the only remaining detail to be addressed. To deal with this, we first select 10^8 detections at random and reject the $\sim 14\%$ of these that are either obvious blends or non- K_s observations. In the range $K_s = [10, 18)$ with a step size of 0.1 mag, we select from this detection pool those which are within 0.1 mag of each interval and measure the integer percentiles of their photometric uncertainties in the range $[0, 100]$. This produces a photometric uncertainty lookup table of 101 percentile measurements at 80 K_s -band magnitude points, from which we can select appropriate photometric uncertainties at a given magnitude and relative observation quality.

The process of simulating a single light curve begins with a selection of a random real light curve as a ‘seed’. The seed light curve sets the observing sequence, and its median dictates the baseline magnitude m_0 . The magnitude at a given time, m_t , is then drawn from the appropriate model and random parameter distribution (see Sections 7.2.1.1 and 7.2.1.2).

An important detail is that the observing sequence is not only the epochs of the observations, but also their quality. To assess the quality of each real observation from the seed light curve, we find the percentile, P_t , corresponding to the measured magnitude uncertainty for the measured magnitude from our photometric uncertainty lookup table using linear interpolation. Finally, for each observation with a model magnitude, m_t , we select an appropriate magnitude error, σ_{m_t} , through linear interpolation of our photometric uncertainty lookup table as a function of m_t and P_t , and scatter the model, m_t , by a magnitude drawn from $\mathcal{N}(0, \sigma_{m_t})$.

The unique identifier of the seed light curve is preserved throughout this process so that we may later recover location information. Fig. 43 shows a representative seed light curve and examples of simulated microlensing and CV light curves using this seed.

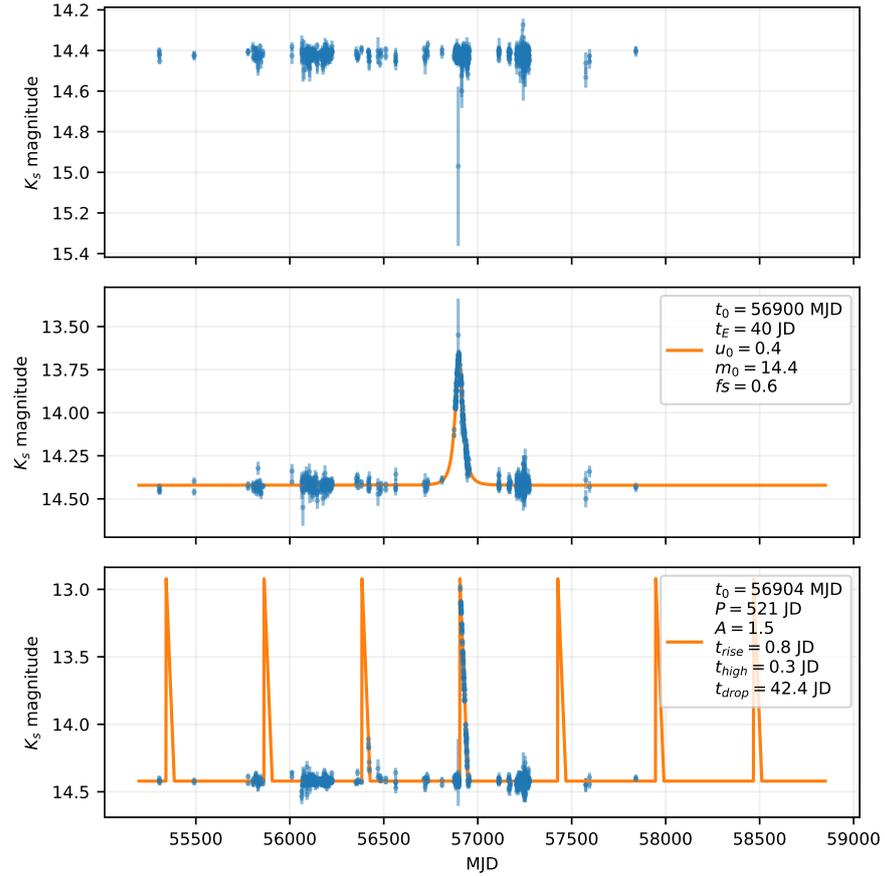


Figure 43: **Top:** The seed light curve used for the simulations in the lower panels. **Middle:** A simulated PSPL light curve and model. **Bottom:** A simulated CV light curve and model. All model parameters were hand-selected from the simulation distributions to best highlight the response of the photometric uncertainties to the variability.

7.2.1.1 PSPL simulation model

To simulate a microlensing event, we randomly assign the remaining PSPL model parameters by drawing from uniform distributions ($\mathcal{U}(\text{lower limit}, \text{upper limit})$) on a linear scale as follows: $t_0 \sim \mathcal{U}(2009, 2021)$ Julian year, $t_E \sim \mathcal{U}(10 \text{ minutes}, 1000 \text{ days})$, $u_0 \sim \mathcal{U}(10^{-3}, 5)$, $f_{bl} \sim \mathcal{U}(0, 1)$. The microlensing light curve model magnitudes for each seed observation epoch and the PSPL parameters are generated using Eq. (70).

7.2.1.2 CV simulation model

To simulate a CV light curve, we adopt the same approach as [Godines et al. \(2019\)](#) but with slightly tweaked parameter distributions. The original (optical) data source that inspired the choices is [Robinson](#)

(1976). To summarise, CV light curves are characterized by periodic outbursts with a relatively sharp rise in flux to m_{\min} (where $m_{\min} = m_0 - A$ and A is the outburst amplitude), a short period of high flux at m_{\min} , and a slower drop in flux back to the baseline level of m_0 . For each simulated light curve, we randomly select parameters as follows: Period, P , from $\mathcal{U}(100, 200)$ days, truncated on the low end at 30 days; Duration, d , from $\mathcal{U}(3, 0.1P)$ days; Amplitude, A , from $\mathcal{U}(0.5, 5.0)$ mag; Rise time, t_{rise} , from $\mathcal{U}(0.5, 1.0)$ days; Time at high flux, t_{high} , from $\mathcal{U}(0.4, 0.2)$ days, truncated on the low end at 0 days; Drop time, t_{drop} , is simply $t_{\text{drop}} = d - t_{\text{rise}} - t_{\text{high}}$.

The truncation on the period and t_{high} parameter distributions are not mentioned by [Godines et al. \(2019\)](#), but are necessary purely to ensure times are always positive and that the gradient of the flux drop is always negative. CVs with periods shorter than this are likely to be reasonably well sampled in phase space by our long baseline data and therefore not cause significant contamination anyway. The gradient of the rising portion of the outburst is therefore $g_{\text{rise}} = -A/t_{\text{rise}}$ mag day $^{-1}$, and the gradient of the dropping portion is $g_{\text{drop}} = A/t_{\text{drop}}$ mag day $^{-1}$.

The magnitude of a simulated CV light curve at time since the start of a given outburst, t' , is:

$$M(t) = \begin{cases} m_0 + t' g_{\text{rise}} & \text{if } t' \leq t_{\text{rise}} \\ m_0 - A & \text{if } t_{\text{rise}} < t' \leq (t_{\text{rise}} + t_{\text{high}}) \\ m_0 - A + g_{\text{drop}}(t' - t_{\text{rise}} - t_{\text{high}}) & \text{if } (t_{\text{rise}} + t_{\text{high}}) < t' \leq d \end{cases} \quad (71)$$

7.3 SEARCH ALGORITHM

In our search algorithm, we opt for a feature extraction based approach to distinguish microlensing events from other light curves (c.f. [Price-Whelan et al., 2014](#); [Wyrzykowski et al., 2015](#); [Godines et al., 2019](#)). However, instead of computing many general summary statistics of the light curve, we focus on a few interpretable features based on microlensing model fitting and checking.

7.3.1 Maximum likelihood estimates

This first feature captures how well a microlensing model fits the data compared to a constant magnitude model. A full Bayesian analysis of this problem obtaining posterior samples via MCMC methods for both the microlensing and constant light curve models for all $\sim 10^9$ sources is computationally infeasible. Even optimized, gradient-based, Hamiltonian MCMC samplers², typically take ~ 10 seconds to converge on light curves with hundreds of data points (e.g. [Golovich et al., 2020](#)). Therefore, we are restricted to use maximum likelihood fits which are much cheaper to compute (typically 10^{-1} seconds per source for PSPL microlensing models).

Let a data point be the set of time, magnitude, and measurement error $D_i = \{t_i, m_i, \sigma_{m_i}\}$, and a light curve be the set of n data points $\mathcal{D} = \{D_i\}_{i=1}^n$. Let $m_{\mathcal{M}}(t_i; \vec{\theta})$ be the predicted magnitude at time, t_i , for model \mathcal{M} with parameters $\vec{\theta}$. Under the assumption that we know t_i exactly and the magnitudes are corrupted with independent Gaussian noise with variance $\sigma_{m_i}^2$, the log likelihood of the light curve given the model for a set of parameter values is,

$$\ln p(\mathcal{D}|\vec{\theta}, \mathcal{M}) = -\frac{1}{2} \vec{r}^T \mathbf{K}^{-1} \vec{r} - \frac{1}{2} |\mathbf{K}| - \frac{n}{2} \ln 2\pi \quad (72)$$

Here $\vec{r} = [m_1 - m_{\mathcal{M}}(t_1; \vec{\theta}), \dots, m_n - m_{\mathcal{M}}(t_n; \vec{\theta})]^T$ is the residual vector between the model and data of length n , and $\mathbf{K} = \text{diag}(\sigma_{m_1}^2, \dots, \sigma_{m_n}^2)$ is an $n \times n$ diagonal covariance matrix. The maximum likelihood estimate (MLE) for a given model, $\vec{\theta}_{\text{mle}}$, is obtained by maximizing Eq. (72; or minimizing $-\ln p(\mathcal{D}|\vec{\theta}, \mathcal{M})$). The corresponding maximum likelihood value is $\ln p(\mathcal{D}|\vec{\theta}_{\text{mle}}, \mathcal{M})$.

The first model fitted is a constant magnitude model, specifically, $m_{\text{const}}(t; C) = C$ where $\vec{\theta} \equiv C$. In this case, the maximum log likelihood solution can be obtained analytically and cheaply. Let $\vec{1}$ be a vector of ones $[1, \dots, 1]^T$ with length n , and \mathbf{m} be the vector of observed magnitudes from the light curve $[m_1, \dots, m_n]^T$. Then, the maximum likelihood solution for the constant magnitude model is,

$$C_{\text{mle}} = [\vec{1}^T \mathbf{K}^{-1} \vec{1}]^{-1} \vec{1}^T \mathbf{K}^{-1} \mathbf{m} = \frac{\sum_{i=1}^n m_i \sigma_{m_i}^{-2}}{\sum_{i=1}^n \sigma_{m_i}^{-2}}. \quad (73)$$

This is the inverse variance weighted average of the light curve magnitudes.

² e.g. <https://caustic.readthedocs.io/>

The second model is the standard Paczynski (1986b) PSPL model for photometric microlensing introduced in Chapter 2. Assuming some non-zero amount of blended light not from source, then the PSPL model is,

$$m_{\text{PSPL}}(t; \vec{\theta}) = m_0 - 2.5 \log_{10}[f_{\text{bl}}A(t) + 1 - f_{\text{bl}}]. \quad (74)$$

Here, m_0 is source baseline magnitude and $\vec{\theta} \equiv [t_0, t_E, u_0, m_0, f_{\text{bl}}]$. Unlike the constant model, the maximum likelihood solution for the PSPL model has to be obtained numerically. We used the differential evolution optimization algorithm (Storn & Price, 1997) implemented by the SciPy Python package (Virtanen et al., 2020) to find the MLE.

Differential evolution is a non-gradient based global optimization algorithm. It works by stochastically evolving a population of candidate solutions in the parameter space according to a chosen strategy. We used the "best1bin" strategy which picks and assesses trial solutions from the population of candidate solutions using the binomial distribution. We chose a population size of 15, a maximum number of iterations of 1000, a convergence tolerance of 0.01, a differential weight of 0.5, and a recombination constant of 0.7. Differential evolution also requires the parameters are bounded. The bounds (lower, upper) used for t_E, u_0 , and f_{bl} were fixed across all sources at (1.0, 10^3) days, (10^{-3} , 3.0), and (0.0, 1.0) respectively. The bounds for t_0 and m_0 were set dynamically for each source between the minimum and maximum values of observation epoch and magnitude in the light curve, respectively.

We note that it is possible to compute gradients of the likelihood with respect to the model parameters in the PSPL model which permits the use of gradient based optimizers (e.g. the BFGS algorithm, Broyden, 1970). Although these are typically faster than differential evolution because they require substantially less likelihood evaluations, we found differential evolution reliably found the global solution, whereas the gradient base optimizers often got stuck in local minima. This was likely due to the optimization problem being difficult because of the highly non-linear PSPL model and typically sparse and noisy VVV data. Despite differential evolution being comparatively slow, it is still feasible to run the optimization for all sources in the VVV survey. The PSPL maximum likelihood optimization took 253 ms per light curve per CPU core on AMD EPYC 7452 chips. It was run on a total of 669 244 443 light curves in the VVV on 1024 CPU cores, totalling 47 033 CPU hours of computation.

7.3.2 Model comparison

With the MLE of both the constant and PSPL models in hand for every source, we can assess the relative quality of their fits to the data. We choose the Akaike Information Criterion (AIC; Akaike, 1981) to compare the models. AIC is an estimator of the in-sample predictive performance of a model on a data set and is defined as

$$\text{AIC}_{\mathcal{M}} = 2k_{\mathcal{M}} - 2 \ln p(\mathcal{D}|\vec{\theta}_{\text{mle}}, \mathcal{M}), \quad (75)$$

where $k_{\mathcal{M}}$ is the number of parameters being estimated in the model \mathcal{M} . AIC is the log predictive density given the MLE with a bias correction term to account for how much fitting a model of k parameters will increase the predictive accuracy alone (Gelman et al., 2014). The underlying assumption that permits adding the number of model parameters is that the posterior is a multivariate Gaussian distribution. This is generally not true for the PSPL model meaning Eq. (75) is an approximation. The main reason for using AIC is that it is readily calculable from the MLE. We are unable to calculate more accurate measures of model predictive performance, for example, the Deviance information criteria (Spiegelhalter et al., 2002), or cross-validation scores (e.g. Vehtari et al., 2017, Chapter 3), as they require samples from the posterior distribution which are too expensive to compute. The difference in AIC scores can then be computed for the PSPL and constant models as

$$\Delta\text{AIC} = \text{AIC}_{\text{const}} - \text{AIC}_{\text{PSPL}}, \quad (76)$$

where a positive number indicates the PSPL model, better explains the data, approximately accounting for the increased complexity of the PSPL model.

7.3.3 Model checking

The ΔAIC score only assesses the relative model performance between the PSPL and constant models. Therefore, we are still blind to cases where both models have poor fits to the data. To remedy this, we compute quality cuts and features based on the PSPL MLE solution. The cuts are applied to all light curves, both real and simulated. This serves two purposes. In the case of simulated microlensing events, it ensures the simulations produce events that were reason-

ably able to be detected by the survey. In the case of real data, the cuts eliminate likely not microlensing events or events not well detected in the survey. Features are used to train the classifier to extract microlensing events.

7.3.3.1 *Quality cuts*

The first quality cut aims to ensure that the microlensing events are well covered by the data. This eliminates events for which the PSPL fit is being driven by only a few data points on the amplification. Moreover, it aims to ensure we extract microlensing events with sufficient coverage that they can have useful inferences drawn about the PSPL parameters. Using the PSPL maximum likelihood parameters for each source, we require the number of data points, before and after two t_E of t_0 , to be greater than 10 ($n_{\text{before}} > 10$ and $n_{\text{after}} > 10$). We also require more than five data points within one t_E of t_0 ($n_{\text{amp}} > 5$).

Next we make two cuts that ensure that the microlensing amplification is significant compared to noise in the baseline. For both of these cuts, we define points on the baseline to be at least two t_E away from t_0 and points on the amplification to be within two t_E of t_0 . First, we require at least three data points on the amplification above three standard deviations of the mean of the data points on the baseline ($n_{3\sigma} > 2$). Specifically, we take the mean, μ_b , and standard deviation, σ_b , of the data points on the baseline (ignoring the reported error-bars) and require three points $> \mu_b + \sigma_b$. Secondly, we require that the predicted amplification of the PSPL MLE solution at the time of the data point closest to t_0 , to be at least 3 times the standard deviation ($> 3\sigma_b$) of the baseline ($\text{Amp}_{\text{sig}} > 3$).

For the PSPL MLE, optimization bounds for the PSPL parameters had to be chosen (see Section 7.3.1). Naturally, because all the light curves in the survey are being fitted with the PSPL model, most of which are not microlensing, some of the optimizations fail. In some cases, this is indicated by a solution near or on the optimization boundary. Fig. 44 shows the distribution of the PSPL MLE parameters (u_0 , t_E , and f_{bl}) close to the lower end of the optimization boundaries for the 513 047 sources that met all the other quality cuts described in this section. In each case, a build-up at the lower bound is seen. We remove these sources by requiring the PSPL MLE parameters $u_0 > 0.0012$, $t_E > 2.5$ days, and $f_{\text{bl}} > 0.004$. We note that no build-up at the optimization bounds was observed for the m_0 and t_0 which were set dynamically for each light curve (see Section 7.3.1), or

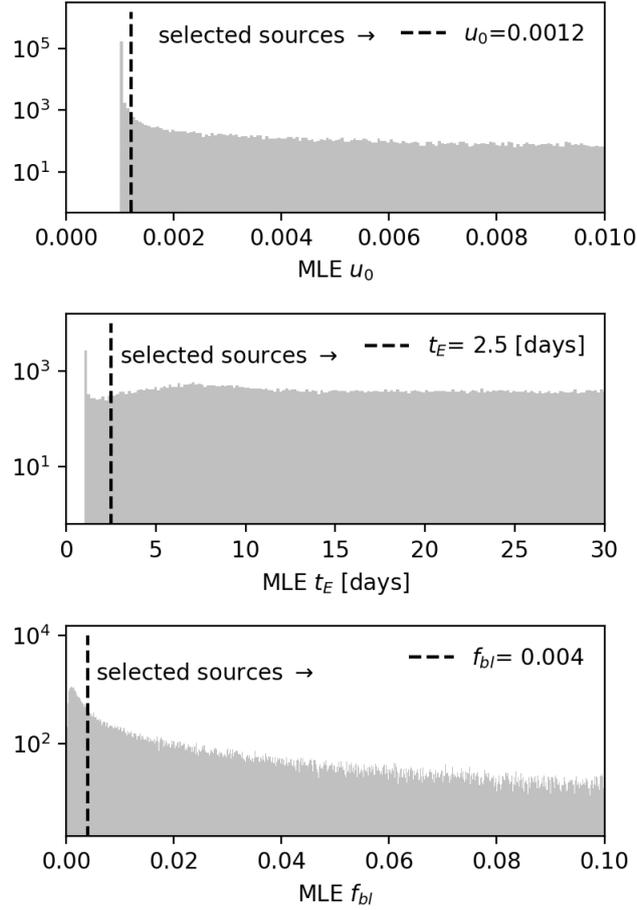


Figure 44: Cuts applied to the MLE parameters for each source. The upper panel is the normalized impact parameter u_0 and the lower panel is the Einstein timescale t_E . A build up at the lower optimization boundaries (10^{-3} for u_0 , 1.0 day for t_E) are observed. The sources at the boundary are removed with the cuts indicated by the dashed lines. The MLE optimization is highly likely to have failed in these cases.

at the upper bounds for u_0 or t_E . An accumulation of sources at the upper bound of $f_{bl} = 1$ was also observed. However, we do not eliminate these sources with a quality cut. This is an expected outcome as $f_{bl} = 1$ is a reasonable physical solution indicating no blending.

The quality cuts described in this section and the number of sources from the VVV survey that survive them, are summarised in Table 10.

7.3.3.2 Features

Next we compute model checking features to be used to train the classifier to identify microlensing events. The features are computed from the distribution of standardized residuals between the data and the

MLE solution of the PSPL model. Specifically, a standardized residual for a given data point D_i is defined as,

$$\hat{r}_i = \frac{m_i - m_{\text{PSPL}}(t_i; \vec{\theta}_{\text{MLE}})}{\sigma_{m_i}}. \quad (77)$$

Under the assumption that the PSPL MLE solution is the true model and the noise assumptions underpinning the likelihood function in Eq. (72) are true, the distribution of standardized residuals over all the data should be a Gaussian with mean zero and unit standard deviation (Andrae et al., 2010).

We compute the standard deviation of the standardized residuals for data points both inside and outside of two t_E of t_0 . This gives two features, σ_{amp} and σ_{base} , on the amplification and baseline, respectively. The features allow the performance of the PSPL model and its noise modeling assumptions to be evaluated both on and off the microlensing amplification.

7.3.4 The von Neumann ratio

We also compute the von Neumann ratio η (von Neumann, 1941, 1942) which is defined as the ratio of the mean square of adjacent differences of data points to the variance of all the data points on the light curve. Specifically, it is

$$\eta = \frac{\sum_{i=1}^{N-1} (m_{i+1} - m_i)^2 / (N-1)}{\sum_{i=1}^N (m_i - \bar{m})^2 / (N-1)}, \quad (78)$$

where N is the number of data points on the light curve and \bar{m} is the mean magnitude. Although η was derived as a trend indicator under the assumption that the data points are independent draws from a population level Gaussian distribution, it has been demonstrated to be useful in identifying microlensing events (e.g. Price-Whelan et al., 2014; Wyrzykowski et al., 2015, 2016; Godines et al., 2019).

Table 10: Initial cuts applied to the simulated events and VVV K_s -band light curve data. The sources remaining are the number of sources surviving the accumulation of the rows above it.

Cut	Description	sources remaining
none	...	669 244 443
$n_{\text{before}} > 10$	number of data points before $t_0 - 2t_E$	375 837 359
$n_{\text{after}} > 10$	number of data points after $t_0 + 2t_E$	159 192 879
$n_{\text{amp}} > 5$	number of data points inside $2t_{E,\text{mle}}$ of $t_{0,\text{mle}}$	92 734 005
$n_{3\sigma} > 2$	number of data points on the amplification above three standard deviations of the baseline	6 115 025
$\text{Amp}_{\text{sig}} \geq 3$	ratio of the PSPL maximum likelihood solution prediction at the data point time closest to t_0 and the standard deviation of the baseline.	513 047
$u_0 > 0.0012$	maximum likelihood solution for u_0 is above the lower optimization bound	241 648
$t_E > 2.5$ [days]	maximum likelihood solution for t_E is above the lower optimization bound	228 985
$f_{\text{bl}} > 0.004$	maximum likelihood solution for f_{bl} is above the lower optimization bound	217 317

This is because even if the data are not Gaussian distributed, η still identifies smooth trends as long as the distribution is symmetric or nearly so (Sokolovsky et al., 2017; Lemeshko, 2006; Strunov, 2006). In the case of the PSPL model with no noise, the data are not Gaussian distributed, but they are symmetric suggesting η should identify PSPL microlensing events well in high signal-to-noise regimes. Moreover, η is trivially vectorizable and so is cheap to compute for all sources in the VVV survey.

7.3.5 Decision tree classifier

With our computed features in hand (ΔAIC , σ_{amp} , σ_{base} , and η), we train a decision tree classifier to extract microlensing events from the VVV survey. Decision trees are non-parametric, supervised, machine learning algorithms that classify objects based on computed features and were first proposed by Breiman et al. (1984). Given a set of objects with computed features and known classes (the training set), a decision tree can be built or trained to predict the class of new data in many ways (see e.g. chapter 9.2 in Hastie et al., 2009).

We use a training algorithm similar to C4.4 (Quinlan, 2014) which is implemented in the scikit-learn Python package (namely DecisionTreeClassifier; Pedregosa et al., 2011). Let Q_i represent the set of features and known class label (microlensing or not microlensing) for one example in the training data. Then the set of all N_T training examples is $\mathcal{Q} = \{Q_i\}_{i=1}^{N_T}$. In this algorithm, at the root node in the tree, there is a binary split in one feature which best separates the training data into their target classes in the two generated left and right child nodes. This split is chosen to minimize the total Gini impurity measure, I , across both child nodes. Specifically, for a given split in the chosen feature, we have

$$I = \frac{N_{\text{left}}}{N} G(Q_{\text{left}}) + \frac{N_{\text{right}}}{N} G(Q_{\text{right}}). \quad (79)$$

Here, Q_{left} and Q_{right} are the subsets of \mathcal{Q} that end up in the left and right child nodes after the split, containing N_{left} and N_{right} training examples, respectively. Initially, $N_T = N_{\text{left}} + N_{\text{right}}$ is the total number of training samples being considered at the parent node. G is the Gini impurity measure computed for the data in each child node,

$$G(Q_s) = p_m(1 - p_m) + p_{nm}(1 - p_{nm}). \quad (80)$$

Here, p_m and p_{nm} are the proportion of light curves that are labelled as microlensing and not microlensing, respectively, in subset Q_s . This procedure is recursively repeated for each new child node containing a subset of the data and the tree grows in size until some stopping criteria is reached. The stopping criteria can be a maximum tree depth, or a minimum number of training samples or a minimum threshold impurity decrease in each of the proposed child nodes.

Once the stopping criteria has been reached, all nodes in the tree without children are called leaf nodes. The trained decision tree can then be used to predict the class of an object from the leaf node the object ends up in. For each leaf node, the probability that the object is a particular class is the proportion of that class that ended up in that leaf node during training. In our case, for each leaf node in the tree, p_m is the probability that the object is microlensing and p_{nm} not microlensing. For an object to be initially classified as microlensing, its microlensing class probability has to be $> P_{\text{thresh}} = 0.5$. It is worth noting that since the training set does not contain realistic proportions of simulated events and contaminants, in addition to a pre-selection with quality cuts, the probabilities extracted are not a true measure of the probability of a particular class. However, the values are still useful because we expect leaf nodes with higher probabilities to contain more genuine microlensing events than leaf nodes with lower probabilities.

The advantages of using a decision tree classification algorithm are that it does not require pre-processing of the input features; once trained it is fast and scalable to run (prediction has logarithmic cost in the number of data points used to train the tree); and it is easy to interpret (prediction can be explained by Boolean logic) when compared with other classification algorithms such as neural networks. Disadvantages are that it tends to overfit training data unless regularized or pruned in some way; it is also poor at interpolating and does not form smooth decision boundaries.³

For our training data, we used 9000 simulated microlensing events and 2250 simulated CVs from the simulation procedure detailed in Section 7.2 that pass all the quality cuts in Table 10. We also take 6750 random light curves from the VVV data (which are vanishingly unlikely to be true microlensing events), which also pass all the quality cuts. We label both the CV simulations and random selection of light

³ See the scikit-learn documentation on decision trees (<https://scikit-learn.org/stable/modules/tree.html>) and references therein for a detailed explanation of the advantages and drawbacks of decision trees.

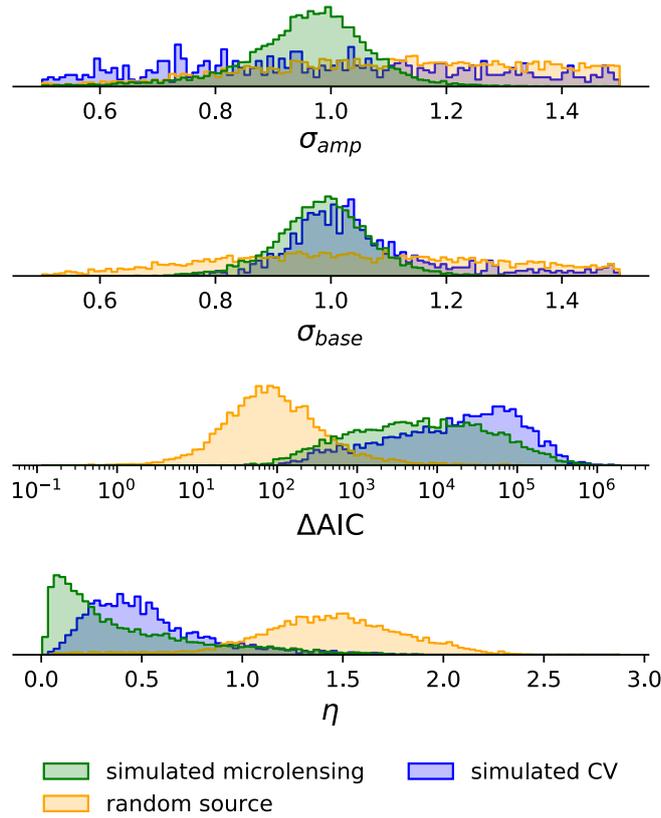


Figure 45: Feature distributions of the training data as input to the decision tree classification model shown as normalized histograms. The simulated microlensing events are well separated from the random sources in most features. As expected, simulated CVs are less well separated indicating that they could be a source of confusion for a classifier trained with these features.

curves to be members of the not microlensing class. Fig. 45 shows the distribution of the training data in the features with the classes overlaid. The microlensing simulations are well separated from the random light curves in most of the features. However, the CVs are less well separated, indicating they will likely be a source of contamination.

There are many possible strategies to train and tune the decision tree classifier. Typically, stopping criteria (e.g. maximum tree depth or minimum number of samples required to be in a leaf node) are tuned as hyper-parameters of the classifier and are chosen to maximize the classification accuracy. This can be done by an expensive grid search over all of the hyper-parameters (e.g. [Godines et al., 2019](#)) to find an optimal combination. We found that this method tended to choose large trees, (depths of 20 – 30 levels) with many parts that only slightly increase classification performance.

In order to keep the decision tree small and scalable, and to prevent over-fitting on the training data whilst keeping its accuracy high, we adopt a two-stage training procedure. We first train the decision tree with fixed hyper-parameters that do not limit the size of the tree. Specifically, we do not limit the maximum tree depth, we don't require any threshold minimum impurity decrease at a split which only needs two training examples, and we only require one training example to be present in each leaf node. We then prune the tree and remove nodes that don't significantly affect the classification accuracy using minimal cost complexity pruning (Breiman et al., 1984).

Minimal cost complexity pruning aims to find the subtree of the unpruned tree T that minimizes the cost complexity measure

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|. \quad (81)$$

Here $R(T)$ is the total sample weighted impurity of all the leaf nodes in T , $\alpha \geq 0$ is the cost complexity parameter, and $|\tilde{T}|$ is the total number of leaf nodes in T . The cost complexity of a single node t in the tree is similarly defined as $R_\alpha(t) = R(t) + \alpha$, where $R(t)$ is the impurity measure at the node. For a given branch of the tree T_t with root node t the effective cost complexity α_{eff} is defined by setting $R_\alpha(T_t) = R_\alpha(t)$,

$$\alpha_{\text{eff}} = \frac{R(t) - R(T_t)}{|\tilde{T}_t| - 1}. \quad (82)$$

For each node in the tree, excluding leaf nodes, this quantifies how well the subtree below the node decreases the classification impurity, regularized for the size of the subtree. Nodes with a high α_{eff} are the top of subtrees which decrease the classification impurity efficiently for their subtree size and should be kept. Pruning is performed by choosing a threshold value for α_{eff} for which all nodes in the tree that have α_{eff} value lower than the threshold are removed. Overall, pruning decreases the size of the tree by removing the weakest parts which do not contribute significantly to discriminating between classes for their size.

Fig. 46 (left panel) shows the decision tree average classification accuracy over the microlensing and not microlensing classes as a function of the threshold value of α_{eff} (the amount of pruning). Accuracy is calculated over a five-fold cross-validation on the 18000 training samples. For each fold, the accuracy on the held out data (test) and on the training data (train) is evaluated. A value of training accuracy

that is higher than test accuracy suggests that the decision tree is over-fitting. We therefore choose the value of threshold $\alpha_{\text{eff}} = 0.00135$ which corresponds to the point where the training accuracy becomes significantly higher than the test accuracy. Specifically, over the five-fold cross-validation, this is the point where the mean train accuracy is greater than one standard deviation above the mean test accuracy. This pruning results in a significantly smaller and simpler tree (< 10 levels deep compared with the unpruned 20-30 levels) that is not over-fitting, and retains a high accuracy $\sim 96\%$ on the test data. The threshold $\alpha_{\text{eff}} = 0.00135$ is fixed and the tree is trained on all 18000 training samples to produce the final trained classifier.

Fig. 46 (middle panel) shows the precision recall curve for the decision tree classifier. Precision and recall are defined as,

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}, \quad (83)$$

$$\text{Recall} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}. \quad (84)$$

Here, a true positive is the correct classification of a microlensing event. The middle panel of Fig. 46 shows how these quantities vary as the P_{thresh} is increased above 0.5. Requiring higher P_{thresh} trades off a purer sample of microlensing events (high precision) against a more complete sample (high recall).

The precision recall curve in Fig. 46 suggest both a high level of completeness and purity can be achieved with the classifier on the training data by choosing a P_{thresh} corresponding to a point in the top right corner of the plot. We will however, return to this point when we apply the classifier over the real data set in Section 7.4.1.

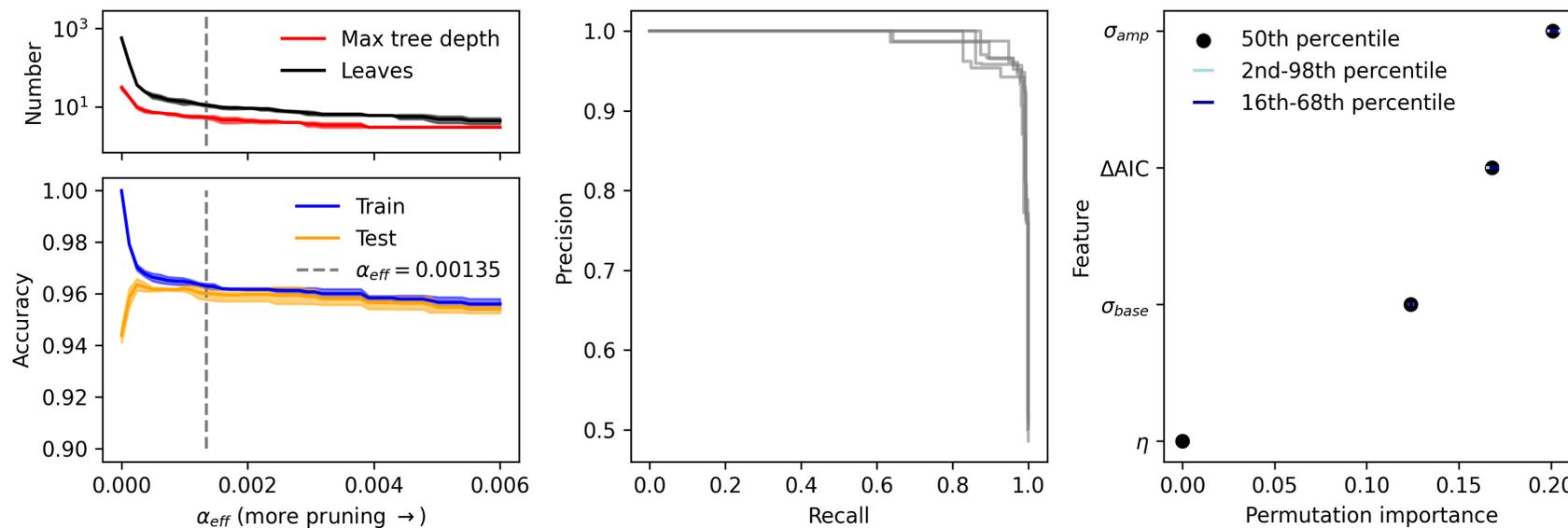


Figure 46: **Left:** The decision tree accuracy and size as function of how much pruning is performed (increasing α_{eff}). In both panels, coloured bands show the mean and standard deviation of each metric over a five-fold cross validation. $\alpha_{eff} = 0.00135$ (dashed grey line) is the smallest value chosen such that tree performance on the training data (blue) is at least one standard deviation above the tree performance on the test data (orange). If there is less pruning than this threshold, the tree starts to over-fit on the training data. **Middle:** Precision recall curve of the pruned decision tree. Grey lines show the curve for each fold of a five-fold cross-validation. **Right:** Permutation importance of each of the features. This shows how much the tree relies on each feature to classify microlensing events. The σ_{amp} feature is the most important discriminator followed by ΔAIC .

The right panel of Fig. 46 shows the permutation importance of each feature. This is the change in classification accuracy of the decision tree when the feature is corrupted by random permutation (Breiman, 2001). This is repeated a set number of times (10 in this work) for each feature giving rise to a distribution of permutation scores. A feature with a high permutation importance is relied on more by the decision tree to produce accurate classifications. Fig. 46 shows that σ_{amp} is most relied on by the decision tree. We note that η has a permutation importance of zero, but this does not mean that the decision tree does not use it at all. The inclusion of η does not increase the average classification accuracy when considering purely a crossing of the threshold $P_{\text{thresh}} = 0.5$, but we found it did increase the purity of the $p_m > 0.5$ bins.

7.3.6 Validation

To evaluate the performance of the trained classifier, it is applied to a previously unseen validation data set distinct from the training data used to train and tune the classifier in Section 7.3.5. This data set contains 2000 examples in the same proportions and labelling as the training data (1000 simulated microlensing events, 333 simulated CVs, and 667 random light curves, or 1000 not microlensing events).

Fig. 47 shows two-dimensional projections of the decision boundary. The lower left panels coloured according to the verdict of the classifier and the upper right panels coloured according to the known classification. The diagonal panels show the one-dimensional distributions of these classes for each feature. The fact that the red and green lines and data points, actual microlensing and claimed microlensing, respectively, shadow each other closely in both the one and two-dimensional distributions is very encouraging. The decision boundary in feature space is clearly complex, but there are prominent regions even in the projected spaces in which the microlensing events clearly predominate.

The performance of the classifier on the validation set can be made quantitative by computing the confusion matrix, as in Table 11. The classifier achieves 97% accuracy on identifying microlensing events, and 94% on identifying not microlensing. This is very respectable; however, the rate of $\sim 6\%$ of not microlensing samples misclassified as microlensing may yet prove troublesome. If the actual data had the same proportions of microlensing events to CVs (the principal con-

True class	Predicted class	
	Microlensing	Not microlensing
Microlensing	97%	3%
Not microlensing	6%	94%

Table 11: Confusion matrix on the previously unseen held out validation data of 2000 examples (1000 microlensing, 1000 not microlensing). Values in the confusion matrix are the fraction of validation examples normalized by row.

taminants) as the validation set, then this would still yield substantial numbers of false positives on large data sets.

The frequency of CVs can be estimated from the Gaia Catalogue of Nearby Stars (Gaia Collaboration et al., 2021b), which contains $\approx 320,000$ stars. The local space density of CVs is $\sim 2 \times 10^{-5}$ (Ak et al., 2008) which suggests that their frequency is $\sim 6 \times 10^{-5}$, or roughly an order of magnitude greater than microlensing. This though is an overestimate as to the seriousness of the contamination problem in VVV data because (i) the definition of CVs includes many classes of dwarf novae (such as SS Cygni stars) which repeat on timescales of a few weeks and so would not be mistaken for microlensing, and (ii) the amplitude of the outburst decreases with increasing wavelength and so the problem is less severe in the infrared VVV as opposed to optical surveys like OGLE. The troublesome CVs have rather specific characteristics, they must rise unusually slowly or the data points corresponding to the rise must be missing. Such troublesome CVs are very much less frequent than microlensing.

7.4 RESULTS AND COMPLETENESS ANALYSIS

7.4.1 Results

Once the classifier has been trained, it is applied to the 217,317 VVV light curves that passed the quality cuts in Table 10. This yields 6677 microlensing candidates with a decision tree microlensing class probability > 0.5 . Many of these are not microlensing, but this is not surprising. The training set consists only of the order of 10^4 examples, so there are bound to be cases of contamination and noise on which the classifier has not been trained. Additionally, a rate of $\sim 6\%$ of not microlensing training samples misclassified as microlensing (see the

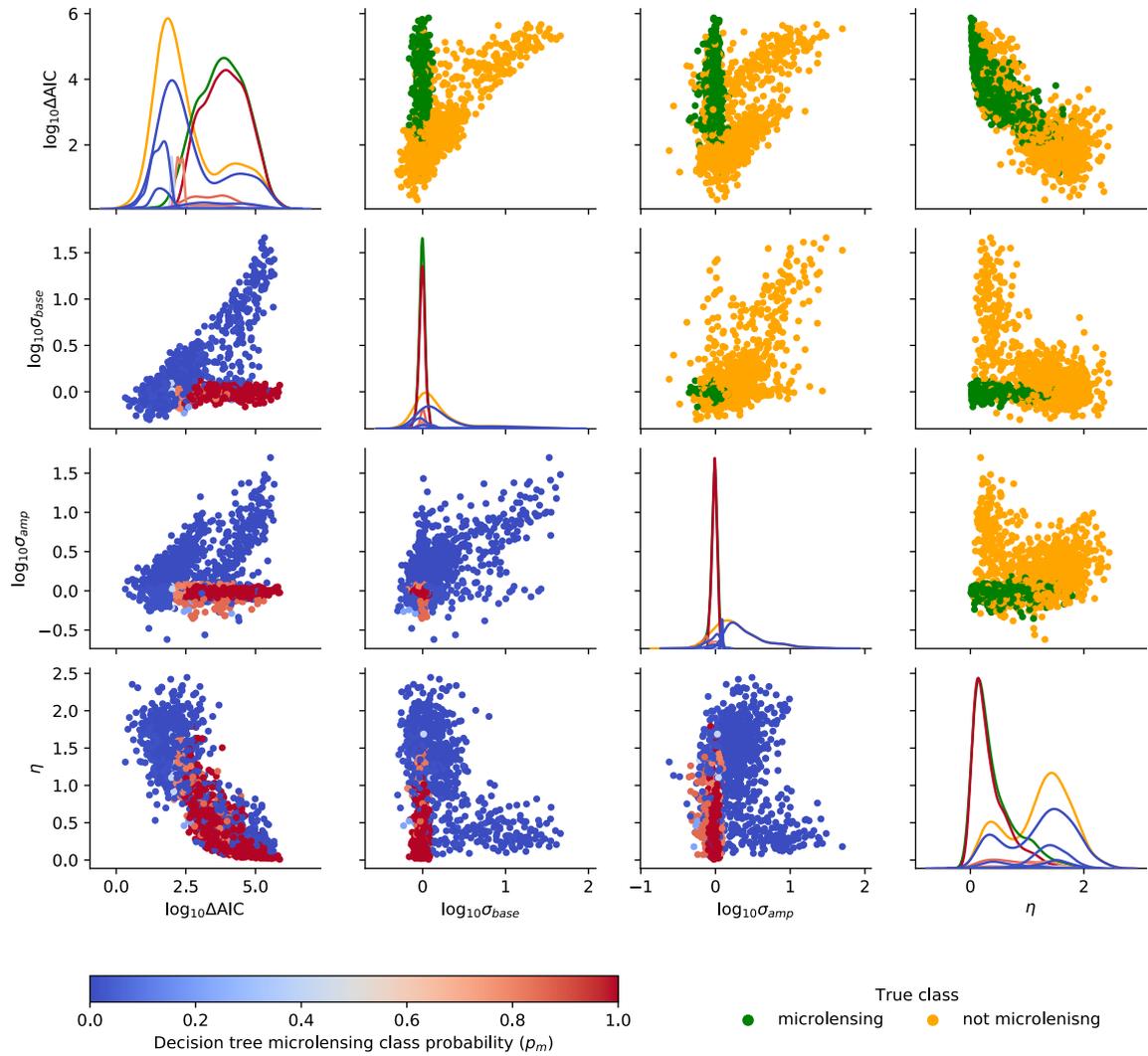


Figure 47: Projected decision surface for the trained decision tree evaluated on the validation data set. Lower scatter plots show the validation data coloured by the microlensing class probability they are assigned by the trained decision tree (p_m) in all two dimensional projections of the feature space. Upper scatter plots show the same two dimensional projections in the feature space, but the sources are coloured by their true classes.

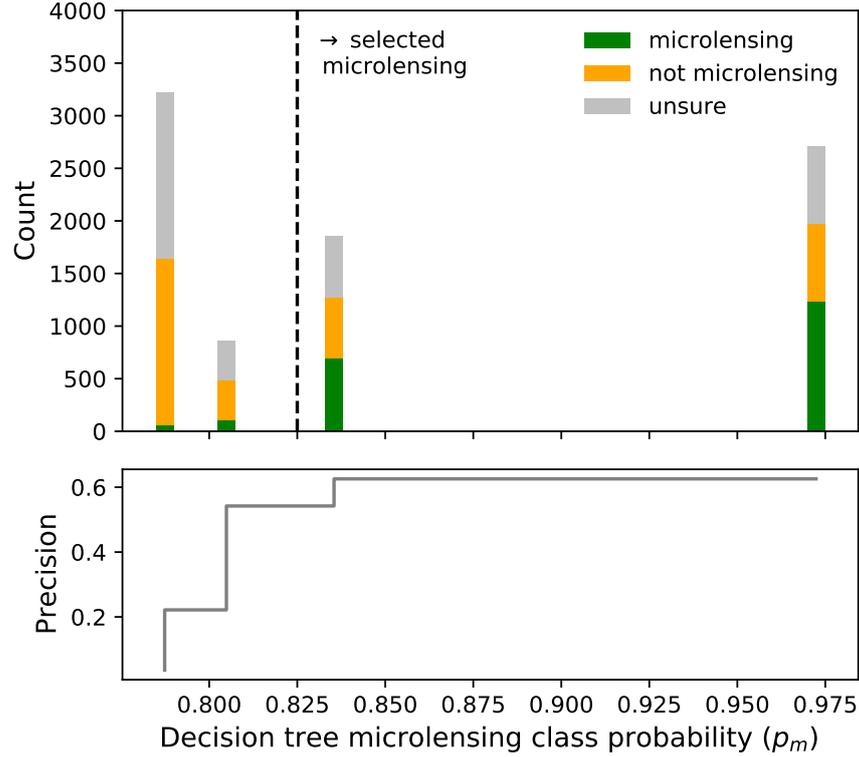


Figure 48: Actual purity of the sources with microlensing class probability $p_m \geq 0.5$ which are classified by visual inspection of the light curves. Dashed line indicates the threshold probability ($P_{\text{thresh}} = 0.825$) used to select microlensing events for the final sample. Values of precision in this plot are calculated ignoring sources with an unsure visual classification.

confusion matrix in Table 11) translates to $\lesssim 10^4$ false positives in our light curve set, though the frequency of the major contaminant (cataclysmic variables) is likely overrepresented in the training and validation sets as compared to the actual VVV data, so this calculation is unduly pessimistic.

None the less, a visual inspection was then required to remove the remaining false positives. First, a preliminary classification of all 6677 light curves was made by myself to produce Fig. 48. The top panel shows the ratio of microlensing to not microlensing as a function of the decision tree microlensing class probability p_m . The bottom panel shows the purity as determined from this first visual inspection. Using the flattening of the purity curve as a guide, we chose $P_{\text{thresh}} = 0.825$ to produce a set of 4132 light curves for further scrutiny with the aim of refining to produce a very pure sample.

It is noted that many of the misclassified microlensing events were consistent with being low-amplitude long-period variable stars. Long-

period variable stars are common in the VVV (e.g. [Angeloni et al., 2014](#)), which suggests that the inclusion of template NIR long-period variable light curves in the training set could help to eliminate them in the machine produced sample.

Each of the 4132 microlensing candidates was now visually inspected at minimum, by a further two people (meaning at least three people saw each candidate) who adopt the scoring of -1 for not microlensing, 0 for unsure and 1 for microlensing. We decided that light curves with a mean score ≥ 0.5 were almost certainly genuine microlensing events, yielding a final list of 1959. This catalogue is presented in electronic form in the supplementary material associated with this thesis as Table 13. Some sample light curves that are classified as microlensing are displayed in Fig. 56 together with results from our later maximum likelihood and Bayesian modeling (see Section 7.5.2). In Sections 7.5 and 7.6 we analyze the final sample of 1959 human-vetted events.

7.4.2 Visual inspection

Spread randomly among the 4132 real light curves were 150 simulated events that had passed the classifier. We included these to gauge completeness with each participant seeing the same set of simulations. In addition, since the visual inspection was performed in two rounds, 100 event candidates that each participant had seen in the first round were also included in the second round to test consistency.

The completeness of the visual inspection, as determined by the recovery of the simulated microlensing, was particularly surprising. The most complete participant still missed 23% of the simulated events, the least complete missed 70% of them; the remaining two missed 27% and 51%. The large variation in completeness between participants was astonishing to us. Completeness was most highly correlated with baseline magnitude and u_0 , with higher values yielding lower completeness. This is not surprising since bright, high amplitude events have the clearest signals. Considering simulated events having a mean score ≥ 0.5 as above, our visual inspection recovers 73 of 150 events, or an overall completeness of just 50%!

The consistency of three of the participants was such that only $\sim 6\%$ of classifications were significantly different between rounds, where ‘significantly different’ is deemed to be a switch from microlensing to not microlensing, or vice-versa. One participant was inconsistent

at the 20% level. We note that the direction of change in disposition of the participants also varied. Two participants clearly became more strict in the second round of inspections (i.e. where a change in classification occurred it was generally towards not microlensing), one was only marginally more strict, and the last was more relaxed relative to the first round.

In conclusion, visual inspection is likely to yield time-varying results with poor completeness. We partially mitigated the effects by having four inspectors. For future surveys, we advocate having far more inspectors view each candidate and taking an aggregate score. This could be achieved with a citizen science project which allows tens of thousands of volunteers to classify astrophysical objects (e.g., [Lintott et al., 2011](#)), or photometric time series data (e.g. [Thiemann et al., 2021](#)). This should provide robust and reliable classifications. Unfortunately, the results of the completeness test indicate that visual classification is no guarantee of a high level of completeness. Even with a large number of inspectors, there is likely to be a point beyond which higher completeness cannot be reached. It is better to design a classification method which requires no human intervention at all to produce a final (incomplete) catalogue of events, though even this is difficult to do for such an intrinsically rare phenomenon as microlensing. However, it is then much easier to quantify completeness and contamination of an exclusively machine classification.

7.4.3 *Spatial variation*

Spatial variation in transient detection efficiency is to be expected when observations are sparse and irregularly sampled. The VVV survey observation strategy is split between four different zones (spatial ranges approximated): The Galactic disc fields at $|b| < 2^\circ$ and $295 < l^\circ < 350$ have ~ 65 observations per pointing, spread relatively evenly across the 10 year survey; the mid-plane bulge fields at $|l - 0^\circ| < 10^\circ$ and $|b| < 2.5^\circ$ have ~ 95 observations per pointing with the majority in the calendar years 2012 to 2015 (inclusive); the high-latitude Galactic bulge fields at $|l - 0^\circ| < 10^\circ$ and $|b| > 2.5^\circ$ have ~ 80 observations per pointing with a similar spread to the disc fields; The high cadence region in the inner bulge at $1.5 < l^\circ < 7.5$ and $-3.5 < b^\circ < -1.5$ has ~ 335 observations per pointing with the majority being in the 2012, 2014, and 2015 calendar years. It should be noted that these counts are per pointing, but the VIRCAM tiling strat-

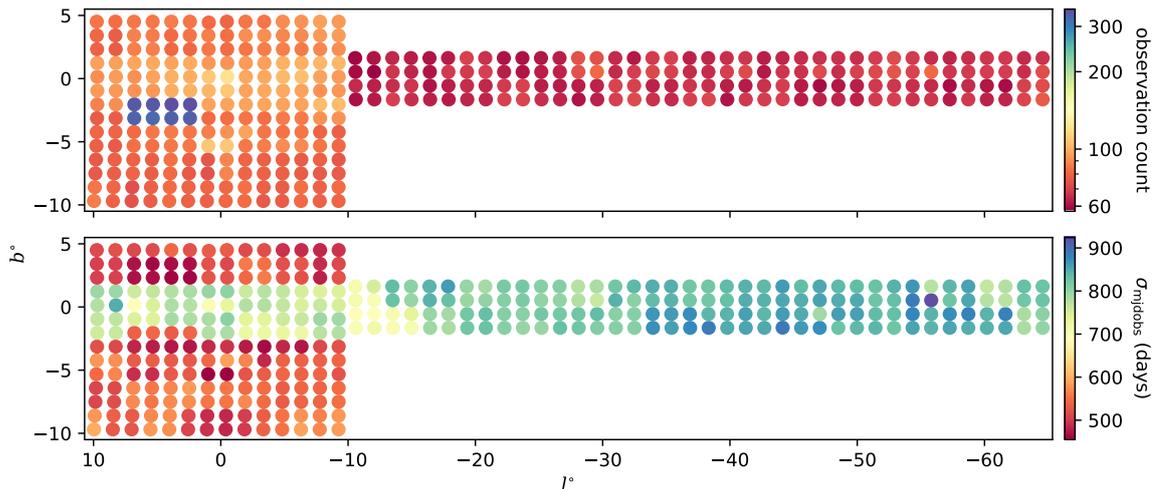


Figure 49: The spatial and temporal density and distribution of the VVV survey. This includes VVVX observations that cover the original VVV survey area. The upper panel colour axis shows the number of observations per pointing position, for a given tile (note the logarithmic scaling for this panel), and the lower panel shows the standard deviation in observation date in Julian days.

egy means sources are covered between 1 and 6 pointings (2 being the modal value). This means a randomly selected position within the survey may be observed anywhere between ~ 65 and ~ 2000 times. The observation count of positions separated by an arcminute can differ by an order of magnitude. The spatial distribution of VVV tiles, coloured by both a per pointing observation count and the standard deviation of the date of their observation (in days), are shown in Fig. 49.

To evaluate the impact of the spatial variation in observation distribution on the microlensing event recovery efficiency, we measured the fraction of simulated microlensing light curves recovered by our classifier (events with $p_m > 0.825$) as a function of sky position and Einstein crossing time. The VVV survey area was binned into $0.5^\circ \times 0.5^\circ$ sections with simulation seed sources selected randomly from within each section. Microlensing event parameters used a discrete value of t_E , a u_0 value randomly selected from $\mathcal{U}(10^{-3}, 1.5)$, and the remaining parameter distributions as in Sec. 7.2.1.1. Simulated events were run through our classifier until the recovery efficiency, the fraction of events classified as microlensing relative to the simulation count, reached a signal-to-noise ratio of 10. This experiment was repeated for a total of five values of t_E . The resulting recovery effi-

ciency maps are shown in Fig. 50 with the positions of real recovered microlensing events included in the map with the closest t_E value.

We note that the regions covered by $N \neq 2$ pointings are not clearly visible in Fig. 50, as they are much smaller in size than our spatial bins. Observations from the same VIRCAM tile visit, are generally all within a few minutes of each other, hence they do not significantly alter the epoch distribution. Consequently, their influence on the light curves for events of a much longer timescale than this is simply to more tightly constrain the magnitude at that time point, typically by a factor \sqrt{N} . Their contribution to the completeness efficiency of each is still included in these maps, as the value for each bin is an average efficiency across the whole bin.

A cursory inspection of Figures 49 and 50 indicate that high recovery efficiency for low t_E events depends more on observation density than it does for high t_E events, for which large observation spread is more important. The high cadence region, which has high observation density but small spread, is very evident as a region of (relative) high efficiency in the $t_E \leq 10$ day panels but largely blends into the background in the high t_E plots. Conversely, the entire mid-plane ($|b| \lesssim 2.5^\circ$) has higher efficiency for long timescale events than the high latitude fields due to its larger relative temporal spread. A number of smaller features are evident in the various recovery efficiency maps in Fig. 50, all of which appear to be mirrored to some degree in at least one of the maps of observation density and spread shown in Fig. 49. The relationship between event recovery efficiency and observation density and distribution for long and short timescale events is clearly visible in Fig. 51 which shows the mean epoch count and mean standard deviation in observation date (in days), for all sources in each spatial bin shown.

The absolute values of recovery efficiency vary hugely, ranging from $< 0.2\%$ for $t_E = 3$ days in the fields with the lowest observation counts, to $> 14\%$, for $t_E = 300$ days in the fields with the highest temporal spread in their observations. Even within single values of t_E , the efficiency can vary by up to an order of magnitude depending on position. This reinforces our view that there is a need to take great care when estimating microlensing survey completeness where inferences on the underlying event rate are to be made.

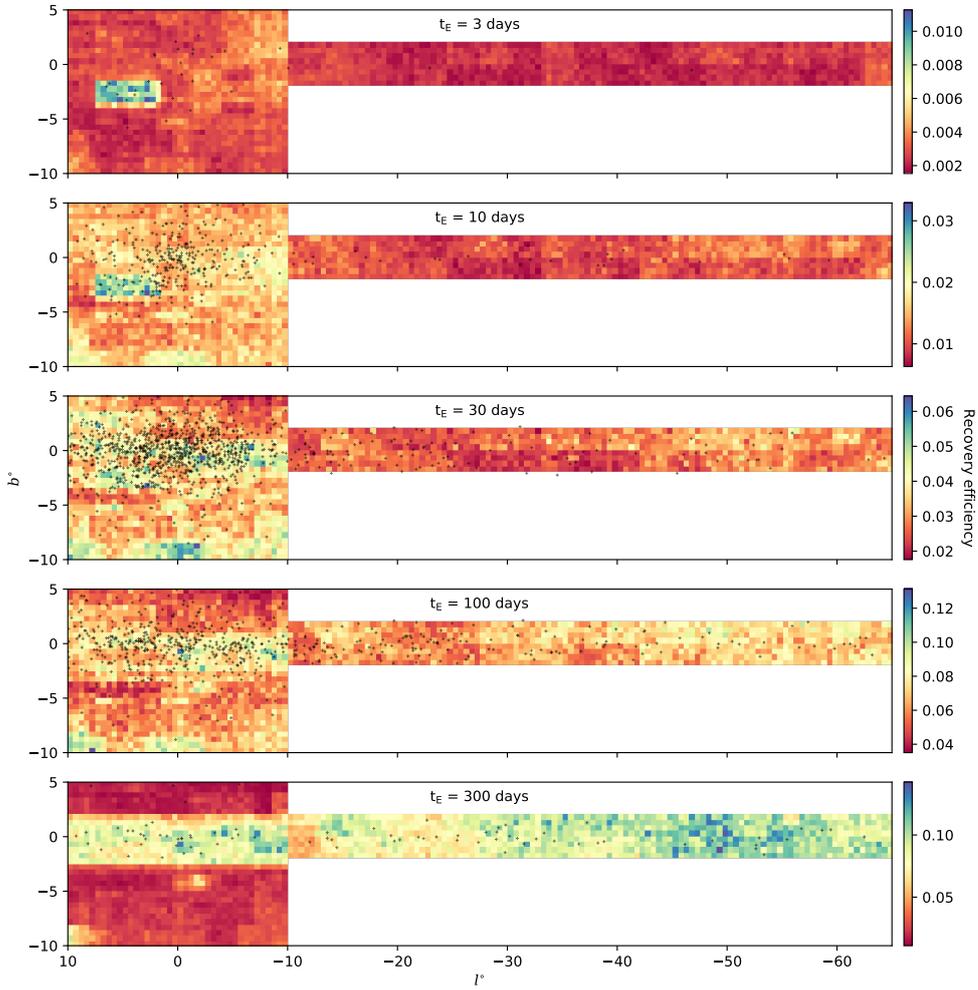


Figure 50: Simulated microlensing event recovery efficiency for five different values of Einstein time and for $10^{-3} < u_0 < 1.5$. For each spatial bin, events are simulated and classified, and the recovery efficiency is the fraction of events classified as microlensing. There is a 10% uncertainty on these recovery efficiency values. The black points indicate the position of recovered microlensing events. Their median t_E dictates the map on which they are placed. These maps highlight the dependence of event recovery efficiency on observation cadence and distribution. The high-cadence VVV fields in the bulge at $l, b \approx 5, -2^\circ$, are clearly visible with a higher recovery efficiency for small t_E , but are less powerful for longer timescale events. The band of high recovery efficiency at $|b| < 2^\circ$ in the $t_E = 300$ day map corresponds to a broader spread in observation dates for these fields.

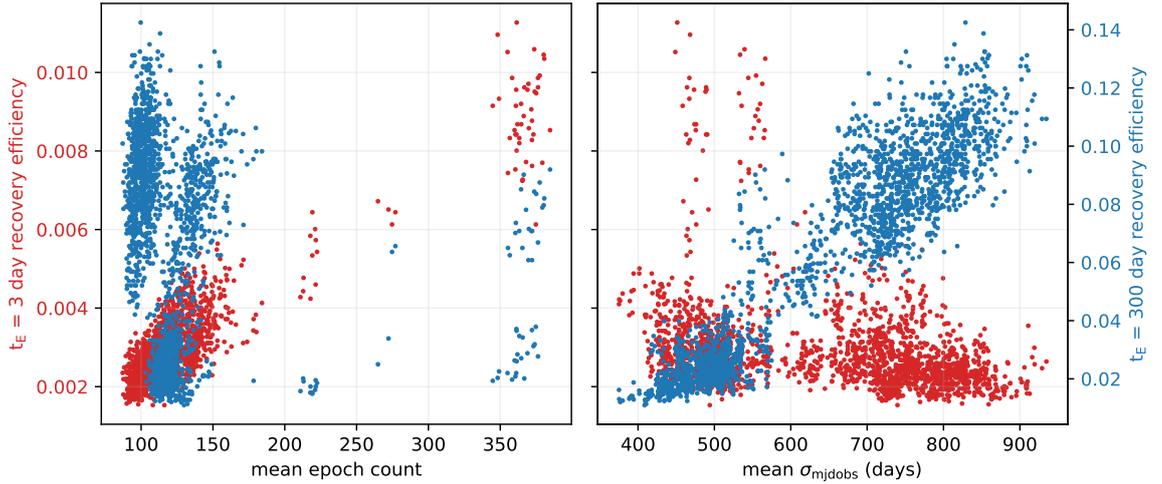


Figure 51: Simulated microlensing event recovery efficiency for $t_E = 3$ day and $t_E = 300$ day events, as a function of mean epoch count and mean observation temporal spread of sources in each spatial bin, as described in Section 7.4.3.

7.5 COMPARISON WITH PREVIOUS VVV MICROLENSING SEARCHES

Here we compare our recovered sample of microlensing events to the three catalogues of previously found microlensing events in the VVV (Navarro et al., 2020a,b,c, ;hereafter the Navarro sample). First, we report on the number of previously known microlensing events our decision tree classifier recovers. We then introduce our Bayesian methodology for inferring the PSPL parameters for each microlensing event. We compare our Bayesian approach with the previously used fitting methods for inference of PSPL microlensing parameters in the VVV.

7.5.1 Known event recovery

The Navarro sample reports 855 unique microlensing events, of which 526 survive the quality cuts outlined in Section 7.3.3.1. Finally, 256 of those make it into the catalogue presented in this Chapter. In the end, our catalogue contains 30% of the reported microlensing events in the Navarro sample.

This relatively low recovery fraction highlights the trade-off between careful visual inspection of all light curves over a small area of the survey versus a more automated scalable search over the full survey area. Both methods have advantages and drawbacks. In the case of a visual inspection, almost all the high amplitude microlens-

ing events will be found. For the low-amplitude, noisy events, our testing and experimentation detailed in Section 7.4 suggests that performance can be patchy and subjective. A careful visual search could provide a somewhat complete sample, at the cost of lack of purity, but this method is not scalable to the $\sim 10^9$ light curves in the survey. This limits analysis to relatively small numbers of events. Moreover, relying entirely on visual inspection introduces biases into the selection of events that are hard to quantify. This causes difficulties when subsequent population analyses of events are performed, for example, optical depth, or timescale distribution calculations.

In contrast, an automated search scales to all $\sim 10^9$ light curves in the survey, allowing many more events uniformly over the survey to be found. As demonstrated by the low recovery fraction, this is at the cost of completeness. The real strength of the method is the ability to better quantify the selection, function as shown in Section 7.4.3 which is vital for subsequent event population analyses. Though the catalogue presented in this work is likely less complete than the Navarro sample, it has the advantage of a comparatively better quantified selection function. This is because we have minimized as much as possible the human component of selecting events, and precisely characterized the selection function of the decision tree classifier.

We now focus on the 526 events from the Navarro sample that passed our initial quality cuts. Some of these events did not make our final catalogue. The reason can best be seen by inspecting the decision surface of the classifier in $\sigma_{\text{base}} - \sigma_{\text{amp}}$ space. Fig. 52 (top panel) shows the 526 events in this space with the decision tree classifier microlensing class probability overlaid. The decision surface shows that events with both high σ_{base} and σ_{amp} are confidently classified as not microlensing (dark blue) which is as expected. In these cases, the light curves have many data points far away from the PSPL maximum likelihood model both on the amplification and in the baseline (taking into account their respective errorbars). For events with σ_{amp} and $\sigma_{\text{base}} \approx 1$, the classifier confidently labels them as microlensing with high probability (dark red). This reflects the simulated microlensing event data used to train the decision tree classifier. We note that for events with $\sigma_{\text{base}} \approx 1$, but with $\sigma_{\text{amp}} < 1.0$, the classifier labels them as microlensing but with less confidence (light red). Promisingly, this indicates that the classifier is using the quality of the maximum likelihood PSPL fit on the microlensing amplification, to differentiate be-

tween very high probability events (dark red) and high probability events (light red).

However, for events with both low σ_{base} and σ_{amp} , the classifier labels them as not microlensing with intermediate probabilities ≈ 0.4 (light blue). Fig. 52 (bottom panel) shows the light curve for one of these events (Navarro ID 62565 in tile b347). Visual inspection of the light curve clearly identifies 62565 in tile b347 as microlensing, indicating room for improvement in the classification algorithm. For such events, the PSPL maximum likelihood solution appears to over-fit the light curve data. This is because, both in the baseline and amplification, the data are too tightly distributed around the model given the size of their errorbars. If 62565 in tile b347 is a genuine PSPL microlensing event, this suggests an assumption made during the modeling has been violated. In finding the PSPL maximum likelihood solution (Section 7.3.1), simulating the microlensing training data (Section 7.2), and interpreting the σ_{amp} and σ_{base} features (Section 7.3.3.2), we assumed independent Gaussian noise with standard deviation of the reported or expected light curve errorbars. This is a strong assumption and is unlikely to hold true for all sources in the survey.

For example, the reported errorbars could be either over or underestimated which could cause σ_{amp} and σ_{base} to be smaller. Moreover, noise in the light curve could be correlated over small timescales (e.g. Golovich et al., 2020), which would change the values of σ_{amp} and σ_{base} . These issues could be partly alleviated by relaxing the independent Gaussian noise assumption (by including a free parameter that inflates or shrinks the reported errorbars, or inclusion of correlated noise). Of course, these additions would come at extra computational cost which would make the classification algorithm less scalable.

Finally, for σ_{amp} and σ_{base} , we are assuming the maximum likelihood solution is the true model. If this is not representative of the possible values (posterior distribution) of the PSPL parameters then this assumption will likely cause issues. In cases of microlensing events with low signal-to-noise, this can cause the posterior in the PSPL to have a large dispersion. In cases with sparse coverage, this can give rise to many plausible solutions and a multimodal posterior. As we will see in Section 7.5.2, this does turn out to be the case for some events. These issues could be partially rectified by, instead of using the PSPL maximum likelihood solution residuals, computing residuals with respect to an approximation to the posterior (e.g. with a

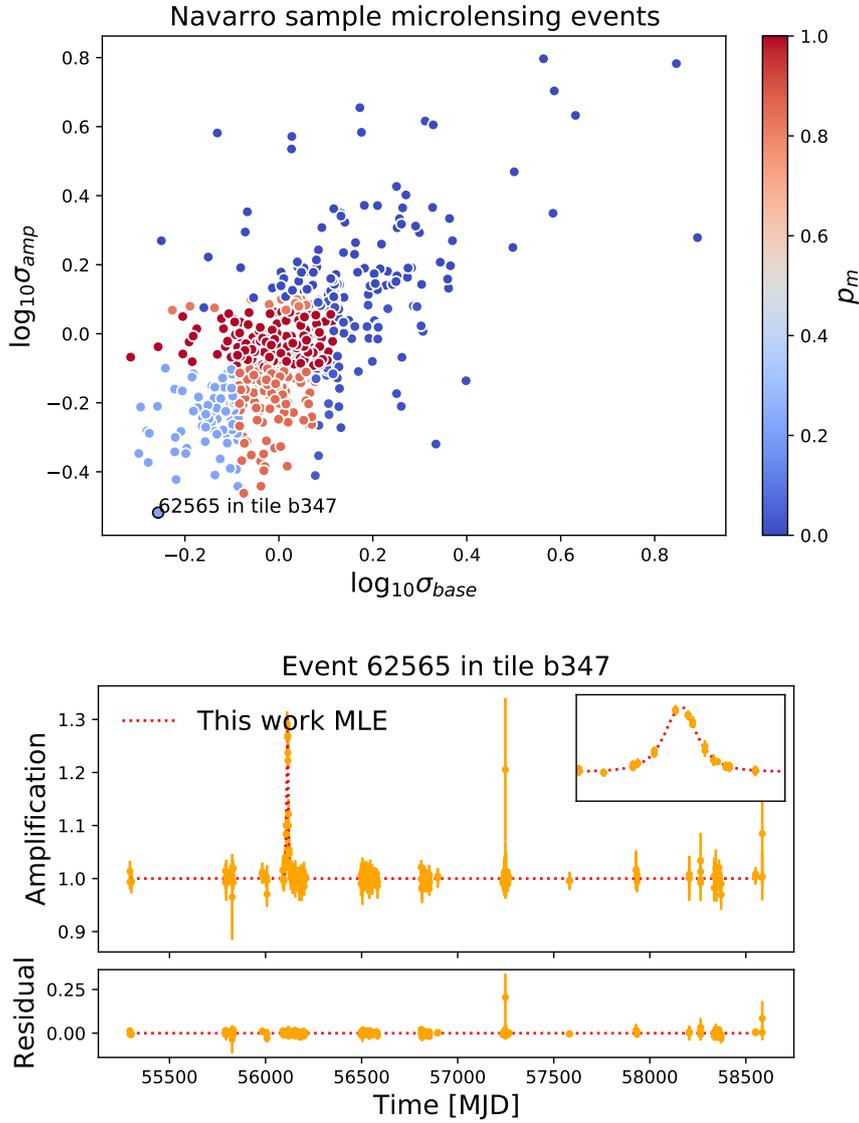


Figure 52: **Top:** Projection of the feature space in $\sigma_{amp} - \sigma_{base}$, for the 526 previously found Navarro microlensing events that meet our initial quality cuts outlined in Table 10. Events are coloured by their decision tree microlensing class probability (p_m). **Bottom:** Light curve of Navarro ID 62565 in tile b347. The position of this event in $\sigma_{amp} - \sigma_{base}$ space is indicated in the top panel. Inset: in the top right is a zoomed-in region of the microlensing amplification. The light curve is plotted in amplification space which is calculated with respect to the maximum likelihood baseline flux. The maximum likelihood solution from this work is shown over the data.

Laplace approximation). However, this too would come at the cost of the scalability of the classification algorithm.

7.5.2 Bayesian inference of PSPL parameters

In the Navarro sample, PSPL parameters for each event were obtained using a non-linear, least-squares, fitting routine implemented in the SciPy Python package (`optimize.curve_fit`; Virtanen et al., 2020). This routine first finds the MLE of the PSPL model under the same Gaussian noise assumptions as in Section 7.3.1 using the gradient based trust region reflective optimization algorithm (Branch et al., 1999). In this algorithm, the gradients are computed numerically with a finite difference scheme and the algorithm converges onto a local optimum of the log likelihood function to produce an MLE. The curvature of the likelihood function around the MLE is then approximated to first order, permitting a covariance matrix for the maximum likelihood parameters to be computed. The diagonal elements of the covariance matrix contain the standard errors on the PSPL parameters, and are reported as one standard deviation errors on the MLE PSPL parameters in the Navarro sample.

In this work, we instead opt for a Bayesian approach to determine the PSPL parameters. We wish to sample the posterior distribution of the PSPL parameters, $\vec{\theta} \equiv [t_0, t_E, u_0, m_0, f_{bl}]$, for each event. By Bayes theorem, the posterior distribution is given by

$$p(\vec{\theta}|\mathcal{D}, \text{PSPL}) = \frac{p(\mathcal{D}|\vec{\theta}, \text{PSPL})p(\vec{\theta}|\text{PSPL})}{p(\mathcal{D}|\text{PSPL})} \quad (85)$$

where

$$p(\mathcal{D}|\text{PSPL}) = \int_{\Omega_{\vec{\theta}}} p(\mathcal{D}|\vec{\theta}, \text{PSPL})p(\vec{\theta}|\text{PSPL})d\vec{\theta}. \quad (86)$$

Here, $p(\vec{\theta}|\text{PSPL})$ is the prior over the PSPL parameters, $p(\mathcal{D}|\vec{\theta}, \text{PSPL})$ is the likelihood as defined in Eq. (72), and $p(\mathcal{D}|\text{PSPL})$ is the model evidence which is an integral overall possible values of the PSPL parameters ($\Omega_{\vec{\theta}}$). For the fitting, we transform the data into flux space assuming a zero point magnitude of 22.0, meaning we fit a baseline flux f_0 instead of a baseline magnitude m_0 . We also fit t_E and f_0 in natural logarithm space as these parameters are then restricted to be positive.

Table 12: Description of priors used in the modeling of the PSPL events. Here, t_{\min} and t_{\max} are the minimum and maximum times of each light curve. F_{med} and F_{σ} are the median and standard deviations of the fluxes in the light curve.

Parameter	Prior distribution
$\ln t_E$	normal(mean=3, standard deviation=6)
u_0	exponential(rate=0.5)
t_0	uniform(lower= t_{\min} , upper= t_{\max})
$\ln F_{\text{base}}$	normal(mean= $\ln(F_{\text{med}})$, standard deviation= $\ln(3F_{\sigma})$)
f_{bl}	uniform(lower=0, upper=1.2)

Our strategy is to use weakly informative priors to constrain the model parameters to reasonable and physical regions of the parameter space. The prior used for t_0 is set dynamically for each event and is uniform between the minimum and maximum times in the light curve. The prior for $\ln f_0$ is also set dynamically for each event to be Gaussian with mean of the median flux of the light curve and standard deviation to be three times the standard deviation of the flux in the light curve. The prior for $\ln t_E$ is set to be Gaussian with mean 3 ($t_E \approx 20$ days) and a standard deviation of 6, allowing t_E to range from $\approx 10^{-2} - 10^4$ days. For u_0 , the prior is an exponential with rate = 0.5. This constrains u_0 to be positive and keeps its value in a reasonable physical range as we do not expect to have detectable microlensing events with $u_0 \gg 1$. Finally, the prior used for f_{bl} , is uniform with lower and upper bounds being 0.0 and 1.2 respectively. The upper bound is above the physical limit of 1.0 to allow for a small amount of negative blending. Negative blending, or values of f_{bl} just above one, are not physical (in such a case the source appears to contribute more than 100% of the flux to the blend), but are possible due to systematic effects in the DoPHOT photometric reduction pipeline (Park et al., 2004; Smith et al., 2007). The prior factorises over the PSPL parameters and the details are summarised in Table 12.

To obtain samples from the posterior distribution of the PSPL parameters, we use the dynamic nested sampling algorithm (Higson et al., 2019) implemented by the dynesty Python package (Speagle, 2020). First proposed by Skilling (2004), nested sampling aims to numerically estimate the model evidence. Nested sampling works by drawing a number of live points from the prior and iteratively removing the point with the lowest likelihood and replacing it with a new point drawn from the prior with a higher likelihood. The pro-

cedure allows the prior to be integrated in nested shells of constant likelihood, in turn, allowing the model evidence of Eq. (86) to be estimated. The procedure is repeated until the estimated remaining fractional model evidence drops below some threshold level. As a by-product of this procedure, posterior samples for the model parameters are generated. In the case of dynamic nested sampling, the number of live points are dynamically allocated which allows a better estimate of the posterior density at a cost of larger error in the model evidence estimate⁴. In this study, we fit events with 1000 initial live points, random walk sampling (Skilling et al., 2006), and multiple bounding ellipsoids. We allocated samples 100% of weight on the posterior and used a stopping criterion in the remaining fractional evidence of 0.01.

The main reason for using nested sampling to characterize the posterior distributions as opposed to MCMC based samplers (e.g. ensemble, Foreman-Mackey et al. (2013), or Hamiltonian Monte Carlo, Golovich et al. (2020)), is its ability to sample from complex multimodal distributions. Due to the sparse and relatively low signal-to-noise nature of microlensing light curves in the VVV, the posterior distributions for PSPL parameters for some events, do turn out to be complex and or multimodal. As an illustration, Fig. 56 shows a selection of 24 light curves from our sample of 1 959 microlensing events. We have chosen them to give an idea of the range of light curves in our catalogue, including examples of high and low amplification events with a variety of noise properties. Overplotted on the K_s -band data, are one hundred posterior realizations together with the MLE. Notice that, especially for noisy and low amplification events, the posterior realizations yield a much broader range of light curve behaviour than indicated by the maximum likelihood curve.

We now turn to comparing our posterior inferences on t_E for each event to the methods used on the Navarro sample. We focus our attention on the 53 events in the Navarro sample with a reported one standard deviation error $< 10\%$ of the reported value of t_E . Fig. 53 shows the comparison between our posterior inference and those from Navarro. In general, our posterior median for t_E is in good agreement with the reported value from the Navarro sample (left panel).

⁴ See the dynesty documentation (<https://dynesty.readthedocs.io/>) and references therein for a detailed explanation of nested sampling.

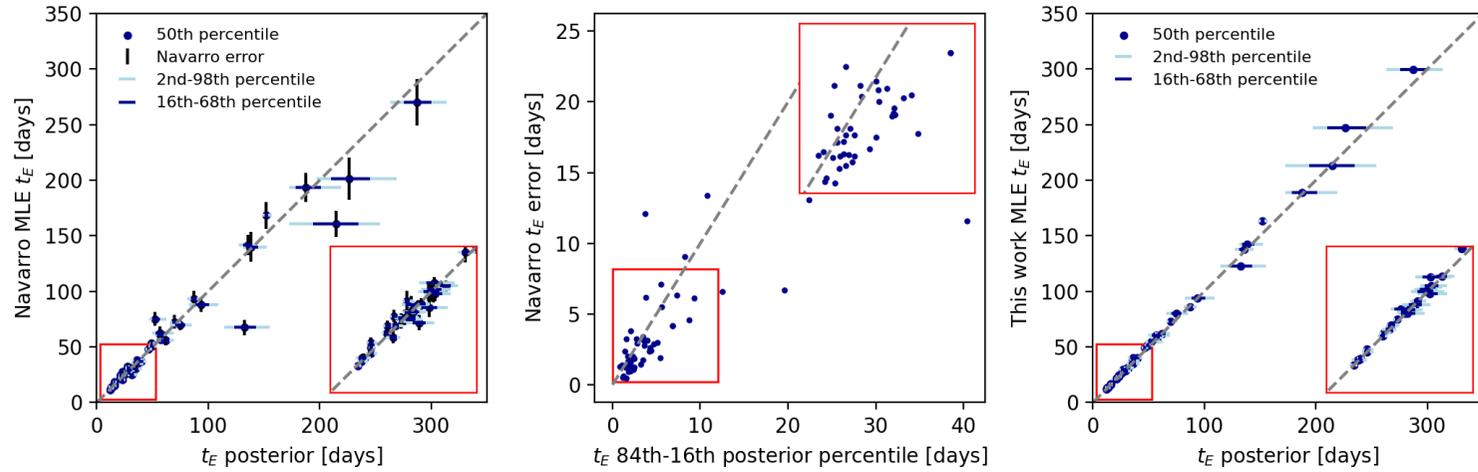


Figure 53: Comparison between the inference on t_E between the analysis of the Navarro sample and a Bayesian analysis from this work. All 53 events in the Navarro sample, where a $\leq 10\%$ error on the inferred value of t_E is reported by the Navarro analysis, are shown. **Left:** Reported Navarro t_E value against Bayesian analysis posterior median from this work. Good agreement for most events is seen. **Middle:** One standard deviation error from the Navarro analysis against 84th-16th percentile of the posterior from this work. The Navarro analysis tends to underestimate the error for the best constrained events. **Right:** MLE computed in this work (section 7.3.1) compared with the posterior median. Better agreement is observed compared with the Navarro analysis in the right panel. This is due to the analysis in this work using the global differential evolution optimization algorithm.

However, we find that the fitting routine used in the Navarro sample tends to underestimate the uncertainty in t_E when comparing the one standard deviation error with the 84-16th percentile of the posterior for the events with a well constrained t_E (middle panel).

Fig. 53 (right panel) shows the posterior distribution compared to the MLE parameters obtained in this work and described in Section 7.3.1. We see better agreement between the posterior median and our MLE parameter values compared with the Navarro sample maximum likelihood solution. This difference is likely due to the fact that the differential evolution optimization algorithm is better able to find the global optimum of the likelihood function compared with the gradient based trust region reflective optimization algorithm.

Fig. 54 shows an example of two complex posterior distributions for t_E which are poorly characterized by the maximum likelihood solution. The top row shows the event with Navarro ID 79184 in tile b305. In this case, both the maximum likelihood methods from this work and Navarro’s analysis find the mode of the posterior. However, the Navarro error estimate overestimates the spread of the posterior. Moreover, the Navarro error estimate fails to capture the asymmetric nature of the posterior distribution which has a heavy right tail. The bottom row of Fig. 54 shows the event with Navarro ID 83519 in tile b337. In this event, the posterior is multi-modal in t_E , and both MLEs fail to precisely find the major mode. Although the Navarro error estimate covers most of the posterior distribution, it does not capture its complex shape.

We note that in many cases, offsets in the Navarro baseline magnitude and the baseline from our reduction of the VVV data are present. This is likely due to differences in the photometric processing of the VVV images, specifically their use of alternative profile fit photometry software and photometric calibration methods. We conclude that the sparse and low signal-to-noise ratio nature of many microlensing events in the VVV necessitate a full Bayesian analysis to fully characterize the inferences on t_E .

7.6 INFERENCES ON THE EINSTEIN TIMESCALE

For the 1959 microlensing events in our catalogue, we fit the PSPL model by the Bayesian method described in Section 7.5.2, and obtain posterior distributions for the Einstein timescale, t_E . In Table 13 con-

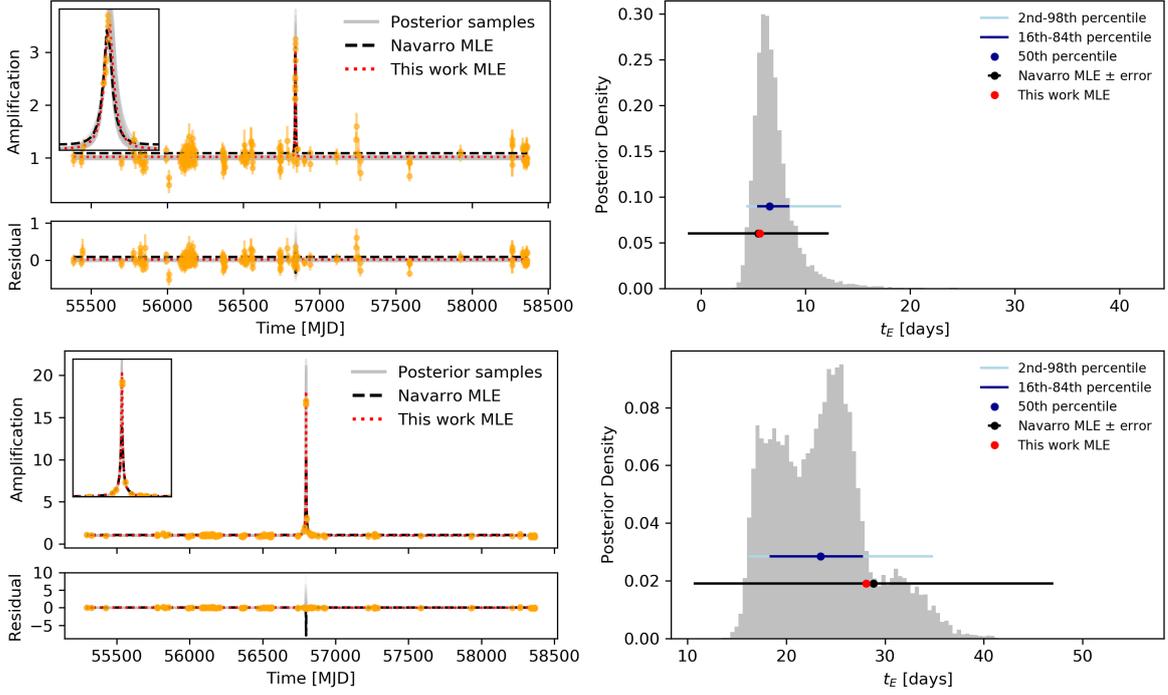


Figure 54: **Top:** (Left) Light curve of the event with Navarro ID 79184 in tile b305 with the posterior inference and MLE from the Navarro analysis and this work overlaid. (Right) marginal posterior distribution with percentiles and maximum likelihood solution overlaid. Both MLEs find the mode of the posterior. In this case, Navarro’s error estimate fails to capture the asymmetric shape of the posterior. **Bottom:** Same as the top row, but for the event with Navarro ID 83519 in tile b337. In this case, both MLEs fail to capture the multimodal or complex nature of the posterior.

tains the 16th, 50th, and 84th posterior percentiles of t_E , for each event along with the same percentiles for all of the other PSPL parameters.

The top panel of Fig. 55 shows the distribution of constraints on t_E obtained by the Bayesian analysis of the events. We can see that typically t_E is only constrained within $\sim 40\%$. These constraints are somewhat poorer than those typically achieved by dedicated microlensing surveys such as OGLE with better coverage of events and higher observing cadences (e.g. see Figure 2 in Mróz et al., 2019).

The bottom panel of Fig. 55 shows the distribution of the median of the posterior of t_E . It shows median values of t_E , ranging from $10^0 - 10^3$ days, and is peaked at ≈ 30 days. One may be tempted to convolve this histogram in its raw form with detection efficiency (as is commonly done in microlensing analyses e.g. Mróz et al., 2020a) in order to make inferences about the t_E distribution over all events. This method, however, neglects the uncertainty on the t_E value for

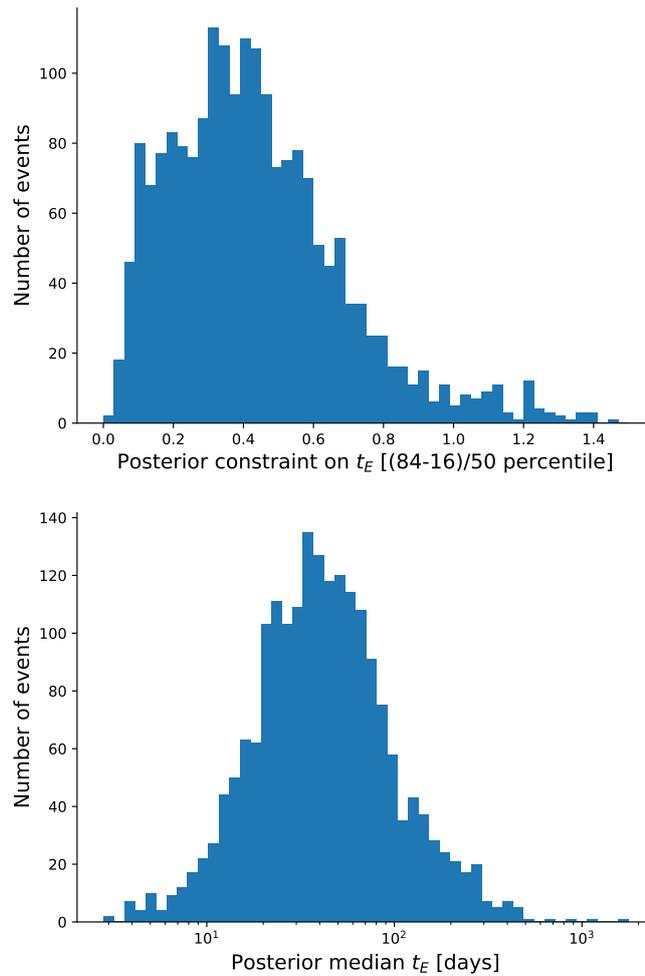


Figure 55: **Top:** Distribution of the posterior constraint provided on t_E , for all 1959 microlensing events found in this work. Typically events have t_E constrained within $\sim 40\%$. **Bottom:** Distribution of the median values of the posterior on t_E , for all the microlensing events.

each event which is not negligible for the sample of microlensing events found in this work.

A method that allows the population level distributions to be inferred from a set of noisy objects, is presented in [Hogg et al. \(2010\)](#) and [Foreman-Mackey et al. \(2014\)](#). It was applied while inferring the population t_E distribution of OGLE microlensing events by [Golovich et al. \(2020\)](#). By modeling in this way, and accounting for correlated noise and parallax, [Golovich et al. \(2020\)](#) showed this can have a large impact on the high end of the inferred t_E population distribution when compared to the simple histogram approach. Given the large uncertainty in each event t_E in the sample presented here, robust inference of population level distributions will likely be best achieved

by careful application of the methods described by [Hogg et al. \(2010\)](#) and [Foreman-Mackey et al. \(2014\)](#). We therefore defer this analysis for future work.

7.7 CONCLUSIONS

We have performed the first search over the full VVV survey area of the inner Galaxy and provided a sample of 1959 microlensing events. We achieved this using a comparatively simple and interpretable machine learning algorithm with very few features. Although applications of machine learning to microlensing have been studied before ([Wyrzykowski et al., 2015](#); [Godines et al., 2019](#)), this is the first successful attempt to implement an algorithm on the full VVV data set of ~ 700 million light curves.

We used a decision tree classifier which is a non-parametric, supervised learning algorithm. Given a set of features, the classifier splits the data into two partitions, microlensing and not microlensing. This happens iteratively, as the data are submitted to more and more tests and split further and further on each of the branches of the decision tree, producing a fast and scalable classifier. Unlike earlier microlensing classifiers ([Wyrzykowski et al., 2015](#); [Godines et al., 2019](#)), we used just four features; specifically, standardized residuals between the data and the MLE of PSPL model both (i) inside, and (ii) outside, two Einstein times of maximum amplitude, (iii) the change in the AIC between PSPL and constant light curve models, and (iv) the von Neumann ratio. The classifier achieves 97% accuracy in identifying microlensing events in training data comprised of random VVV light curves and simulated microlensing and cataclysmic variable events.

Even though our training set contains $\sim 10^4$ examples, there are still some anomalous, very noisy or unusual VVV light curves that are not recognised. These can end up being misclassified and have to be removed by hand. So, the very final stage requires human intervention. This, though, is at a manageable level of inspecting $\sim 10^4$ light curves from the original starting sample of $\sim 10^9$. So the classifier has made the task of analysing the entire VVV light curve database manageable. The final catalogue of 1959 microlensing events, together with their positions and inferences on their timescales, impact parameters, and blending fractions, are available in [Table 13](#). Some sample light curves are displayed in [Fig. 56](#), together with the MLE and Bayesian modeling.

Parts of the VVV footprint have previously been scrutinised for microlensing by Navarro et al. (2018, 2020a,b,c). They used a conventional mixture of parameter cuts to remove variable stars and other contaminants, combined with visual assessment. Our sample is less complete compared to the previous Navarro et al. searches in the common regions of sky, but it extends over the whole VVV survey area and possesses a better characterized selection function. However, we do acknowledge that in the final visual classification step, we do potentially introduce hard to quantify error into our selection function. This is partially mitigated by averaging visual classification of different people for each event, and by precise characterization of the decision tree selected sample, of which the visually confirmed events are a subset.

For our events, we provide spatially resolved classifier efficiency maps as a function of the Einstein time, t_E . This is vital for population studies. For example, these efficiency maps show large features in t_E due to variable survey observing strategies, and are important to take into account when doing spatially resolved population studies. The latitude or longitude dependence of the optical depth or timescale distribution contains information on the lens and source population (Evans & Belokurov, 2002), but it is not possible to extricate without understanding the variation of efficiency on the sky (Navarro et al., 2020a).

Due to the relatively sparse and noisy VVV light curve data, as compared with dedicated microlensing surveys like OGLE, we implemented a Bayesian scheme to determine the parameters of our 1959 events. We find that the posterior distributions for t_E are often complex. We show that maximum likelihood approaches used in previous analyses do not capture this complexity, and therefore fail to entirely characterize the uncertainty in t_E . Only for the well-constrained events, ($< 10\%$ error on t_E) do we find good agreement between our Bayesian approach and MLEs. This has implications for forthcoming surveys that are not dedicated to microlensing and so have sporadic sampling, such as the Legacy Survey of Space and Time at the Vera Rubin Observatory (Sajadian & Poleski, 2019) or the Zwicky Transient Facility at Palomar Observatory (Medford et al., 2020). Also, for many events in our sample, there is typically a $\approx 40\%$ constraint on t_E . This necessitates a Bayesian approach to population inference to handle uncertainty (e.g. Golovich et al., 2020), rather than previous approaches of histograms which ignore the uncertainty in each t_E .

For the VVV survey, t_E is typically not well enough constrained to study distributions of timescales without allowing for this.

In our final step involving visual inspection, every light curve was examined by at least three people. The consistency of any individual is often good, in the sense that when provided with the same light curve again, the same decision as to microlensing or not microlensing is made. However, from individual to individual, there is considerable variation as to the classification, especially if the light curve is noisy or sparsely sampled. We dealt with this by averaging over the scores of all the individuals participating in our visual assessment. However, visual inspection is not completely reliable as a final step. In ideal circumstances, we would prefer a fully automated classification algorithm. We suspect that the way to improve ours is to model and simulate the noise properties of the data better.

There are a number of future directions for the work in this Chapter. First, we plan to carry out statistical analyses of the longitude and latitude dependence of properties of our sample of VVV microlensing events. This has already been examined over a smaller portion of the VVV footprint by [Navarro et al. \(2018, 2020a\)](#). The advantage of revisiting the problem with the new sample is that our efficiency maps enable us to remove spatially dependent artefacts associated with properties of the survey or extinction. With these effects untangled, we can reconstruct the distribution of timescales t_E , as a function of position. In our sample, the timescale distribution between $t_E = 30$ days and 100 days drops by $\approx 50\%$ in the bulge fields, but appears to remain fairly constant in the disc. Is the change in recovery efficiency large enough to explain this or is this a real feature? An answer to this question will give us clues as to the properties of the lens and source populations.

Secondly, the number and properties of microlensing events depend on the distribution and kinematics of stars and compact objects along the line of sight. Precise measurements of the microlensing optical depth and the event rate toward the Galactic bulge enable tests of competing models of the inner Milky Way, including the orientation and mass of its bar. Such calculations are only possible because of our accurate characterization of the efficiency of the classification algorithm for events. The most recent calculations by OGLE are based on ~ 8000 events with optical light curves ([Mróz et al., 2019](#)). However, the infrared bands of the VVV survey are better able to penetrate

the extinction close to the Galactic plane. Our microlensing sample is a complementary probe to OGLE in the inner Galaxy.

Thirdly, anomalous microlensing events can show deviations from the PSPL model. Of course, it is the important category of binary lenses that shows the most substantial morphological variety. The light curves can exhibit multiple peaks and caustic crossings and can be very different to the familiar time-symmetric PSPL light curve (e.g., [Mao & Paczynski, 1991](#); [Di Stefano & Perna, 1997](#)). There are also less drastic changes caused by parallax or finite source effects for single lenses. Our classifier is not trained to identify such light curves and we expect most in the VVV survey will have been missed. The problem of extending machine learning methods to identify anomalous events is both interesting and challenging ([Mróz, 2020](#); [Khakpash et al., 2021](#)). In the wider context, the VVV survey is already proving a useful testing ground for automatic classification of all manner of variable phenomena such as RR Lyrae ([Cabral et al., 2020](#)) and other types of variable star ([Medina et al., 2018](#)). We expect the importance of this activity to increase as we prepare for larger surveys and larger telescopes to map out the terrain of time-domain astronomy.

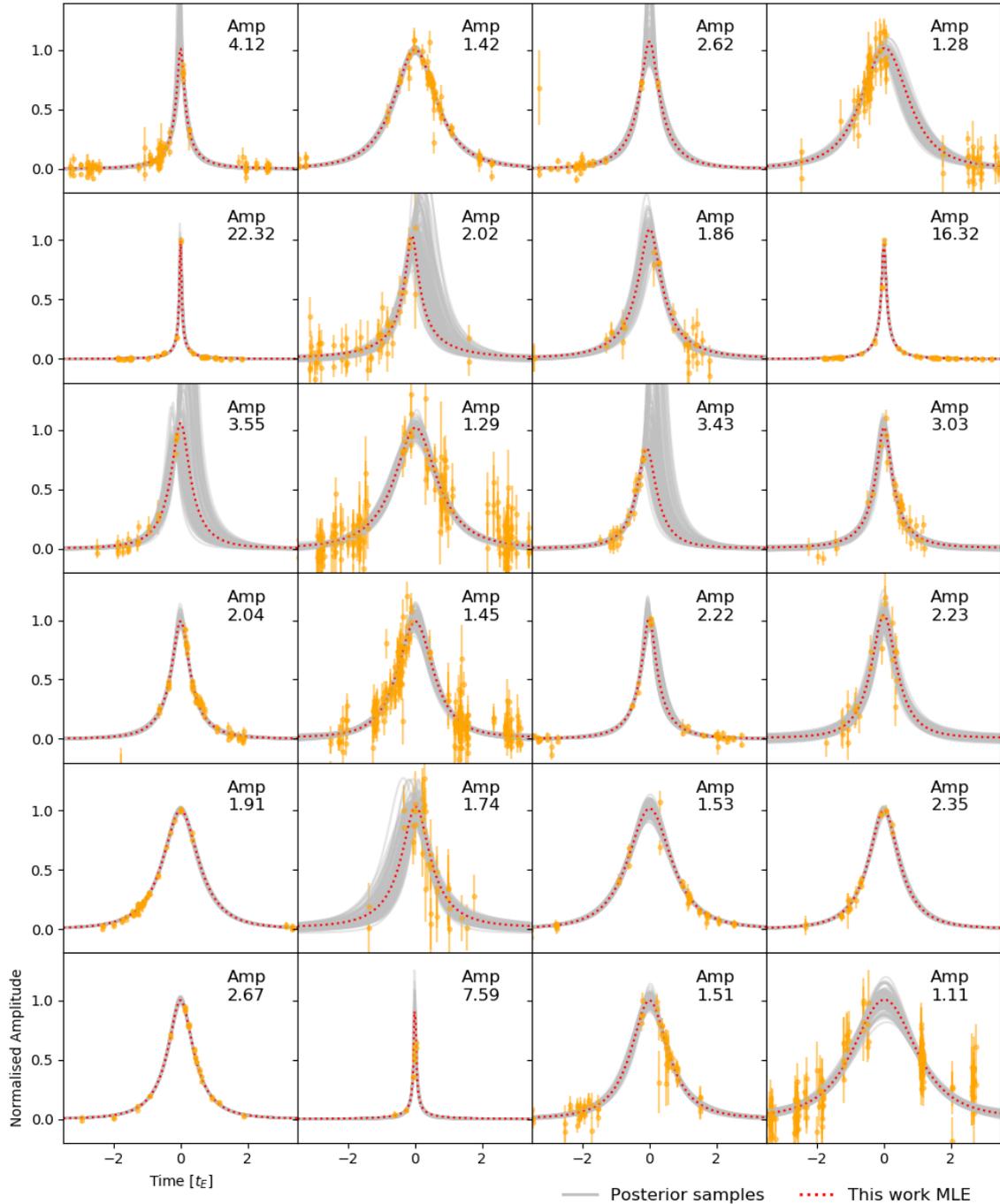


Figure 56: Selection of found microlensing events with one hundred posterior realizations and the MLE overlaid on the K_s -band light curve data. Each event is plotted in normalized amplitude and t_E space, to allow them all to be plotted on the same scale. For each event, the flux is offset by the posterior median baseline flux and normalized by the posterior median of the flux at maximum amplification. An estimate of the maximum amplification is shown as text in each plot. This maximum amplification is calculated by dividing the posterior median maximum flux by the posterior median baseline flux.

CONCLUSIONS

I will first outline my original research contributions presented in this thesis, chapter by chapter. I will then summarize general conclusions tying the results of different chapters together, and present future directions of research.

8.1 CONTRIBUTIONS OF THIS THESIS

Chapter 3:

- I presented the first search for specific predicted microlensing events using high-precision astrometric data from GDR1, resulting in one high-confidence event prediction.
- This event was followed up by HST and resulted in a direct gravitational mass determination of the DQ type white dwarf LAWD 37. The mass provided the first semi-empirical test of the MRR for an isolated white dwarf, and marked only the third detection of the astrometric microlensing effect for a predicted microlensing event outside the solar system.
- I was the first to apply a Bayesian analysis framework to the analysis of a predicted astrometric microlensing event. This allowed the combination of prior information on the lens and source from GEDR3 and HST astrometric follow-up to extract a mass for LAWD 37, capturing all uncertainties in a consistent manner. This is in contrast to the analysis of the two previous predicted astrometric events in [Sahu et al. \(2017\)](#) and [Zurlo et al. \(2018\)](#) where prior information on the source and lens was fixed and a χ^2 minimization was used to extract the mass of the lens.
- I was the first to apply the Bayesian LOO cross-validation score to the analysis of a predicted microlensing event, and to my knowledge, this was first application of LOO in the astrophysics literature. The LOO cross-validation score proved critical to interpreting, comparing, and reasoning with the different model fits. Although the computation of the LOO cross-validation

score is common practice in applied statistics (the LOO method paper, [Vehtari et al. \(2017\)](#), has ≈ 2000 citations at the present time), its potential use in the analysis of astrophysics problems has yet to be realized.

Chapter 4:

- I searched GDR2 for predicted microlensing events with detectable photometric signals. I found two events which had maximums in J2019 which were independently found by both [Bramich \(2018\)](#) and [Klüter et al. \(2018b\)](#).
- Using a new Bayesian framework, I demonstrated that GDR2 astrometry combined with just the photometry of the two events is sufficient to extract a lens mass. In other words, comparatively expensive and likely challenging astrometric follow-up is not required to obtain lens masses for these events.
- I showed that fine-grain optimization of the photometric follow-up (e.g. number of data points, how they are distributed on the microlensing signal, and their precision) can affect the extracted lens mass precision at 5% – 10% level (Figs. [35](#) and [36](#)).

Chapter 5:

- I was the first to extend predictive microlensing searches towards the Galactic bulge and plane in the NIR using GDR2 combined with background source astrometry from VIRAC. I found two predicted microlensing events missed by searches solely using GDR2.
- While the events were not followed up, they demonstrated GDR2 is incomplete for predicted events and astrometry from VIRAC is of sufficient astrometric precision to predict event signals with reasonable confidence.

Chapter 6:

- I uncovered a previously unknown and significant type of contaminant that needs to be considered and mitigated for when predicting microlensing events with the Gaia data. Contrary to previous searches which focused on removing spurious high-proper motion lenses (e.g. [Salim & Gould, 2000](#); [Proft et al.,](#)

2011), I found that a significant contaminant was spurious background sources caused by, either a marginally resolved binary component of the candidate lens or failure modes in the Gaia processing pipeline.

- I showed that removing this contaminant eliminates 61% of predicted events that would have been detected by Gaia over its extended mission, including some of the most promising candidates. This significantly changes the current outlook for detecting predicted microlensing events with Gaia as presented in Klüter et al. (2020).
- Removing events with spurious sources also eliminated events with predicted maximums in the future, which would likely be the subjects of follow-up campaigns. By identifying these spurious events, I provided a service to the community by stopping pointless and expensive astrometric follow-up campaigns.

Chapter 7:

- I presented the first search for photometric microlensing events across the whole VVV survey footprint, and found 1952 events in the NIR and located across the inner regions of the Galactic bulge and plane.
- To extract the events, I used an interpretable and scalable decision tree classifier, which was trained only using four features based on model fitting, checking, and smoothness of the light curve. The trained classifier performed similarly (97% accuracy on the simulated training set) to other classifiers designed to find microlensing events but which used tens of features (Wyrzykowski et al., 2015; Godines et al., 2019).
- The scalability of the decision tree classifier permitted spatially resolved efficiency maps to be computed for t_E which are required for population studies. This is critical for VVV microlensing events as the observing cadence varies significantly over the VVV footprint, causing orders of magnitude change in the detection of efficiency for a given t_E and event location (Fig. 50).
- I showed that the use of a Bayesian analysis framework in combination with nested sampling can characterize sparsely sampled events with complex posterior distributions. This method solves the so called 'forsaken events' problem presented in

Navarro et al. (2020b). An event was dubbed ‘forsaken’ when, due to sparse light curve sampling, it could be explained by two distinct sets of PSPL parameters (two modes in the posterior). This resulted in Navarro et al. (2020b) discarding these events in subsequent analyses. Whereas using my method, nested sampling allows the full characterization PSPL parameter posterior distributions, even in the case of multimodality. Ultimately this means that these previously “forsaken” events do not have to be discarded in subsequent analyses.

8.2 CONCLUSIONS AND FUTURE WORK

8.2.1 *Follow-up of predicted microlensing events can be optimized*

This conclusion is supported by both Chapters 3 and 4. In general, if a microlensing event can be predicted, most of the event’s parameters are known ahead of time. Most directly, this allows follow-up parameters like the required photometric or astrometric precision, number of data points, and which filter to observe in, to be optimized ahead of time. In Chapter 4, and under the assumption of white noise corruption, it was seen that clustering follow-up around the event maximum signal improved the lens mass constraints.

However, the analysis of the astrometric microlensing event by LAWD 37 in Chapter 3 suggests more subtle follow-up optimization opportunities. These opportunities are due to the non-trivial correlated structure of the noise present in the LAWD 37 HST follow-up data and are likely common to the follow-up of the majority of predicted astrometric events. Naively, the HST follow-up for the LAWD 37 event clustered most of the data around the predicted event maximum. This seems like a sensible approach in the first instance. However, typical predictable microlensing events are caused by nearby bright lenses (Dominik & Sahu, 2000; Belokurov & Evans, 2002) which means typically, we will be in a similar scenario to the analysis in Chapter 3, where the lens PSF subtraction introduces significant correlated noise into the data. This correlated noise is usually larger at closer lens-source separations (see e.g. Table 2), or at the largest deflection signals. Moreover, in Chapter 3, it was shown that the correlated noise can mimic the astrometric deflection signal. Ultimately, this means that for this analysis, epoch 2, in the tails of the event, was critical for measuring the deflection signal, and not the

data at the predicted maximum. The utility of epoch 2 was due to a high signal-to-noise offset in a direction that could not be explained by alteration of the source’s astrometric parameters.

With the power of hindsight, the importance of taking observations at epoch 2 could have been worked out ahead of planning the HST observations. This is because all of the lensing system’s parameters, (lens and source astrometry and lens mass estimate) are already known, and the size of the correlated noise introduced by the lens PSF subtraction could be roughly estimated ahead of time. All of this information could be used in a simulation to optimize the HST observation times to get the optimal constraint on the lens mass and or ruling out purely correlated noise explanations of the data. A future line of research would be to work out the formal optimization problem to find the best follow-up observation times and this could be tested on future predicted microlensing events (e.g. [Mustill et al., 2018](#); [Bramich & Nielsen, 2018](#); [Klüter et al., 2018b](#)).

8.2.2 *Gaia is incomplete for predictable microlensing events*

This conclusion is supported by both Chapters 5 and 6. While astrometric data from Gaia has revolutionized the prediction of microlensing and has opened the prediction channel in microlensing (there is now a steady stream of ≈ 10 predicted events/year over the next century [Bramich, 2018](#); [Klüter et al., 2018b](#); [Bramich & Nielsen, 2018](#)), searches are far from complete. One direction of future research would be to solve the three outstanding issues with current searches, and complete the predicted microlensing event almanac. The three outstanding issues are as follows.

Current searches are incomplete for faint low-mass and high proper motion objects. [Gaia Collaboration et al. \(2018\)](#) estimates $\approx 17\%$ of higher proper motion objects are missing from GDR2. Moreover, faint lenses such as brown dwarfs are not selected as lenses in current searches because of the strict astrometric qualities which are applied (e.g. [Bramich, 2018](#)). This could be rectified by supplementing GDR2 with external proper motion catalogues containing the missing faint high proper motion objects. While efforts in this direction have been started ([Nielsen & Bramich, 2018](#)), they are far from complete, especially for objects like brown dwarfs (e.g. [Best et al., 2020](#)). Moreover, objects such as brown dwarfs which are

too faint in the optical to be detected by Gaia, will not be present in future Gaia data releases.

Current searches are incomplete for background sources which caused events to be missed as demonstrated in Chapter 5. GDR2 only has ≈ 0.01 sources/arcsecond² in the Galactic bulge region. This is mainly due to heavy extinction and source crowding issues yet to be resolved in Gaia's data processing. While this is likely to improve with future data releases, Gaia will not produce a complete background source list in the Galactic bulge or plane. Regions of the Galactic bulge and plane where we expect most of the lensing events are much better covered by dedicated surveys such as VIRAC (Smith et al., 2018) and DECAPS (Schlafly et al., 2018) which have source densities approximately one hundred times greater than that of GDR2.

Current searches are blind to unresolved events close to the GDR2 reference epoch (J2015.5), as we saw in Chapter 6 Fig. 42 which could be analyzed when Gaia eventually releases its time-series astrometry. This severe dip in the number of events around the Gaia reference epoch shown in Fig. 42 is due to the lens and source in these cases being too close for Gaia to resolve and hence the event could not be predicted. In Chapter 6 I conclude that at least half of the microlensing events expected to occur during the Gaia mission have yet to be identified because of this issue. This could be resolved by using GDR2 in combination with legacy source catalogues like 2MASS (Skrutskie et al., 2006) and DSS. These legacy source catalogues have sufficiently separated reference epochs for events happening over the Gaia mission time to be resolved, and hence would allow the events to be predicted.

8.2.3 *A Bayesian analysis is the natural choice for predicted microlensing events and sparsely sampled photometric events*

The work in Chapters 3 and 4 involved combining prior astrometric information from Gaia on the lens and source with follow-up data to extract the lens mass. This analysis scenario is general to all predictable microlensing events. For all predictable microlensing events, because the event was able to be predicted in the first place, naturally, there is some level of prior information available on the lens and source. I chose a Bayesian framework for these analyses because it readily allows the inclusion of the prior information coherently with the information contained in the follow-up data.

In the case of modeling the photometric events found in the VVV and presented in Chapter 7, a Bayesian analysis proved critical but for a different reason. Because these events were sparsely sampled, a complex and large range of possible parameter combinations could explain the data. This meant that traditional point-estimation analysis techniques (χ^2 or a maximum likelihood analysis) could not fully characterize the range of possible parameter values (Navarro et al., 2020b). In contrast, the Bayesian analysis framework, combined with nested sampling, handled the resulting complex posterior distributions without issue. Looking forward, this will become more important for modeling and extracting information from microlensing events found in large-scale surveys (e.g. LSST) which will not have observing strategies optimized for microlensing (e.g. Sajadian & Poleski, 2019).

8.2.4 *Historical imaging data is a vital diagnostic for predicted microlensing events*

In Chapter 6, I discovered that spurious background sources are a significant contaminant when predicting microlensing events with GDR2. The key insight enabling the work in Chapter 6 was that because most predicted microlensing events are caused by high proper motion objects, the lens and source should be well separated in historical imaging data. This means in most cases, the lens and source should be resolved and clearly identifiable in the images. This visual inspection check was performed for all candidate events in this thesis (Chapters 3, 4, and 5), and proved robust, at least in the case of the LAWD 37 event in Chapter 3. While not scaleable or automated, visual inspection of historical imaging of predicted microlensing event candidates to identify the source and lens has proved a robust, widely applicable, and powerful diagnostic tool.

8.2.5 *Future work for the photometric VVV events*

The most immediate application for the sample of photometric VVV events found in Chapter 7 would be comparison of the average spatially resolved t_E distribution with Galactic models (e.g. MaBulS; Specht et al., 2020). Due to the high uncertainty in the t_E posteriors for many events, comparison between the samples of events and the MaBulS Galactic models will likely be best achieved by convolu-

tion of the spatial efficiency maps with Bayesian population analysis techniques (e.g. [Hogg et al., 2010](#); [Golovich et al., 2020](#)). If the number of stars monitored in the VVV survey could be estimated, then spatially resolved optical depth could also be compared with the MaBulS Galactic models. Additionally, identifying regions of high optical depth, or lines of sight with high event rates and therefore high planet yields, in the inner bulge, will be useful for the future exoplanet microlensing survey by WFIRST ([Penny et al., 2019](#)).

The analysis of single events may also provide a fruitful line of future research. In particular, the sample of VVV events contains a large fraction of long timescale events (there are \approx hundreds of events with $t_E > 100$ days in Fig. 55). For these events, detection of annual parallax anomalies ([Wyrzykowski & Mandel, 2020](#)), or even the absence of any such anomalies (e.g. [Karolinski & Zhu, 2020](#); [Lam et al., 2020](#)), could be used to identify whether the event was caused by an isolated stellar mass black hole. Detection of such black holes may provide insight into the apparent lack of black holes in the so called ‘mass-gap’ between the lightest known black holes, $\approx 5M_\odot$, and the heaviest neutron stars, $\approx 2M_\odot$ ([Wyrzykowski & Mandel, 2020](#)).

8.3 FINAL REMARKS

Thanks to the astrometric data from Gaia and the NIR photometric data from the VVV, two new windows in microlensing are now open. First, Gaia has delivered the era of high-confidence microlensing event prediction, permitting target and optimized follow-up of events. Follow-up of the steady stream of predicted ≈ 10 microlensing events per year occurring over the next century and found with Gaia will permit direct and precise mass measurements for stars across the Hertzsprung–Russell diagram, and in the solar neighbourhood (e.g. [Bramich & Nielsen, 2018](#)). Second, NIR events found in the VVV will provide the first microlensing probes of inner regions of the Galactic bulge (e.g. [Gould, 1995](#); [Navarro et al., 2020a](#)) giving clues to its structure. While microlensing probes of the solar neighbourhood and the Galactic bulge are open, the work of fully utilizing them has only just begun.

This appendix contains additional material for Chapter 3. This includes details on the Leave-One-Out (LOO) cross-validation approximation and alternative fitting procedures for inference of LAWD 37's mass.

A.1 LEAVE-ONE-OUT CROSS-VALIDATION APPROXIMATION

In this Section we explain how the LOO PSIS approximation is applied to the models considered in Chapter 3. This section closely follows [Vehtari et al. \(2017\)](#) and [Bürkner et al. \(2020\)](#). The goal of the approximation, given that we have obtained S posterior samples from fitting \mathcal{M}_{TN} to the full data set \mathcal{D} , $\{\vec{\theta}^s\}_{s=1}^S$, we wish to approximate,

$$p(\mathcal{D}_i | \mathcal{D}_{-i}, \mathcal{M}_{\text{TN}}) = \int p(\mathcal{D}_i | \vec{\theta}, \mathcal{D}_{-i}, \mathcal{M}_{\text{TN}}) p(\vec{\theta} | \mathcal{D}_{-i}, \mathcal{M}_{\text{TN}}) d\vec{\theta}. \quad (87)$$

Specifically, we want to use $\{\vec{\theta}^s\}_{s=1}^S$ to approximately evaluate Eq. (87) instead of having to refit the model (which is expensive) with the i th data point left-out to obtain $\{\vec{\theta}_{-i}^s\}_{s=1}^S$. This is achieved by using an importance sampling approximation and re-weighting $\{\vec{\theta}^s\}_{s=1}^S$ accordingly.

Importance sampling is a Monte Carlo method to used to compute expectations. We would like to evaluate the expectation,

$$\mathbb{E}_f[h(\theta)] = \int h(\theta) f(\theta) d\theta, \quad (88)$$

where f is some hard-to-sample-from distribution. Instead of using samples from f we can use samples from an easier-to-sample-from proposal distribution g , $\{\theta_g^s\}_{s=1}^S$, provided we know the the ratio between f and g which is $r(\theta_g^s) = f(\theta_g^s)/g(\theta_g^s)$. The importance sampling approximation is then,

$$\mathbb{E}_f[h(\theta)] \approx \frac{\sum_{s=1}^S r(\theta_g^s) h(\theta_g^s)}{\sum_{s=1}^S r(\theta_g^s)}. \quad (89)$$

The quality of this approximation is sensitive to the distribution of the importance weights, $r(\theta_g^s)$. If the proposal distribution, g , is not representative of the target distribution f , then the importance weights become unstable. This instability comes from the importance weights being dominated by a few extreme values which leads to them having a large or infinite variance. Ultimately this leads to a poor importance sampling approximation (Vehtari et al., 2015). To mitigate this problem, $r(\theta_g^s)$ are fitted to a Pareto distribution and the extreme values are removed and are replaced with draws from the fitted Pareto distribution, $\tilde{r}(\theta_g^s)$. These smoothed weights, $\tilde{r}(\theta_g^s)$, are dropped in as a replacement for $r(\theta_g^s)$ in Eq. (89). This is called Pareto Smoothed Importance Sampling (PSIS; Vehtari et al., 2015). In addition to stabilising the importance sample approximation, PSIS also provides a diagnostic on the quality of the importance sampling approximation. This diagnostic is the fitted shape parameter, k , of the Pareto distribution. k traces the number of finite moments of the importance weights distribution and therefore the quality of the PSIS approximation. Vehtari et al. (2015) find that $k < 0.7$ indicates PSIS will work well whereas a value of $k > 0.7$ indicates the PSIS approximation is likely to be poor and should not be used.

For approximating the LOO score, our hard-to-sample-from target distribution is $p(\vec{\theta}|\mathcal{D}_{-i}, \mathcal{M}_{\text{TN}})$. This is the posterior distribution obtained when fitting the model to the data set with the i th data point left-out. This is hard to sample from because it is expensive and we would have to refit the model, which we want to avoid. h is $p(D_i|\vec{\theta}, \mathcal{D}_{-i}, \mathcal{M}_{\text{TN}})$. Our easy to sample from proposal distribution is $p(\vec{\theta}|\mathcal{D}, \mathcal{M}_{\text{TN}})$ which we readily have samples from - $\{\vec{\theta}^s\}_{s=1}^S$. The only remaining task is to compute the ratio between the target and proposal distributions.

$$r(\vec{\theta}) = \frac{p(\vec{\theta}|\mathcal{D}_{-i}, \mathcal{M}_{\text{TN}})}{p(\vec{\theta}|\mathcal{D}, \mathcal{M}_{\text{TN}})} \propto \frac{p(\vec{\theta}|\mathcal{M}_{\text{TN}})p(\mathcal{D}_{-i}|\vec{\theta}, \mathcal{M}_{\text{TN}})}{p(\vec{\theta}|\mathcal{M}_{\text{TN}})p(\mathcal{D}|\vec{\theta}, \mathcal{M}_{\text{TN}})} = \frac{p(\mathcal{D}_{-i}|\vec{\theta}, \mathcal{M}_{\text{TN}})}{p(\mathcal{D}|\vec{\theta}, \mathcal{M}_{\text{TN}})}. \quad (90)$$

Here we have used Bayes rule which states that the ratio of the posteriors is proportional to the ratio of the prior \times likelihood, and the priors terms have cancelled. For all models the likelihood factorises over the different epochs (Eq. 52). If the left-out data point, D_i , is the

j th data point in epoch \bar{e} , all other epochs cancel from the likelihood ratio in Eq. (90) and,

$$r(\vec{\theta}) \propto \frac{p(\mathcal{D}_{-i}|\vec{\theta}, \mathcal{M}_{\text{TN}})}{p(\mathcal{D}|\vec{\theta}, \mathcal{M}_{\text{TN}})} = \frac{1}{p(\mathcal{D}_{\bar{e},j}|\mathcal{D}_{\bar{e},-j}, \vec{\theta}, \mathcal{M}_{\text{TN}})}. \quad (91)$$

The ratio is proportional to the inverse of the likelihood of the left-out data point (Bürkner et al., 2020). In all models considered in this work, $p(\mathcal{D}_{\bar{e},j}|\mathcal{D}_{\bar{e},-j}, \vec{\theta}, \mathcal{M})$ is the product of two multivariate Gaussian distributions (in the X and Y directions) as described in Eq. (51). Using the results from Sundararajan & Keerthi (2001), Bürkner et al. (2020) and Eq. (51), this likelihood can be computed efficiently as,

$$\begin{aligned} \log p(\mathcal{D}_{\bar{e},j}|\mathcal{D}_{\bar{e},-j}, \vec{\theta}, \mathcal{M}_{\text{TN}}) &= -\log 2\pi\tilde{\sigma}_{\bar{e},j} \\ &\quad - \frac{1}{2\tilde{\sigma}_{\bar{e},j}} \left(\left(\left[\vec{X}_{\bar{e}} \right]_j - \tilde{X}_{\bar{e},j} \right)^2 \right. \\ &\quad \left. + \left(\left[\vec{Y}_{\bar{e}} \right]_j - \tilde{Y}_{\bar{e},j} \right)^2 \right). \end{aligned} \quad (92)$$

Here, $[\bullet]_j$ is the j th component of vector \bullet . The conditional means, $\tilde{X}_{\bar{e},j}$ and $\tilde{Y}_{\bar{e},j}$, and conditional standard deviation $\tilde{\sigma}_{\bar{e},j}$ are given by Bürkner et al. (2020) as,

$$\tilde{X}_{\bar{e},j} = \left[\vec{X}_{\bar{e}} \right]_j - \frac{\left[(\boldsymbol{\Sigma}_{\bar{e}}^{\text{N}})^{-1} (\vec{X}_{\bar{e}} - \vec{X}_{\bar{e}}^{\text{T}}) \right]_j}{\left[(\boldsymbol{\Sigma}_{\bar{e}}^{\text{N}})^{-1} \right]_{jj}}, \quad (93)$$

$$\tilde{Y}_{\bar{e},j} = \left[\vec{Y}_{\bar{e}} \right]_j - \frac{\left[(\boldsymbol{\Sigma}_{\bar{e}}^{\text{N}})^{-1} (\vec{Y}_{\bar{e}} - \vec{Y}_{\bar{e}}^{\text{T}}) \right]_j}{\left[(\boldsymbol{\Sigma}_{\bar{e}}^{\text{N}})^{-1} \right]_{jj}}, \quad (94)$$

and,

$$\tilde{\sigma}_{\bar{e},j} = \frac{1}{\left[(\boldsymbol{\Sigma}_{\bar{e}}^{\text{N}})^{-1} \right]_{jj}}. \quad (95)$$

Here, we have dropped the dependence of $\boldsymbol{\Sigma}_{\bar{e}}^{\text{N}}$, $\vec{X}_{\bar{e}}^{\text{T}}$ and $\vec{Y}_{\bar{e}}^{\text{T}}$ on $\vec{\theta}$ for brevity. $[\bullet]_{jj}$ denotes the j diagonal element of the matrix \bullet .

Using these weights with PSIS, we can then approximate Eq. (87) for each left-out data point using the posterior samples from the full dataset, $\{\vec{\theta}^s\}_{s=1}^S$. For each left-out data point, we obtain the Pareto k

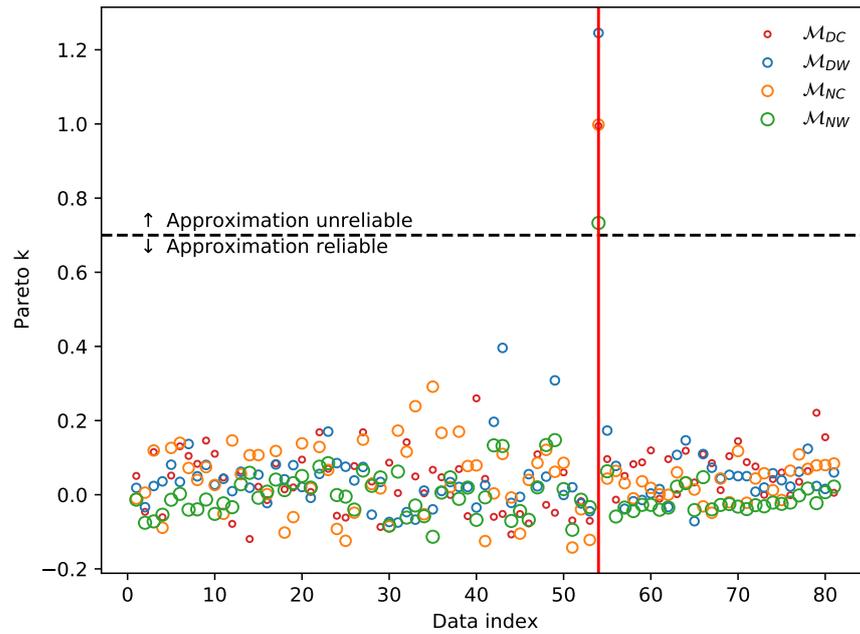


Figure 57: Pareto k values in the PSIS-LOO approximation used for each model. Different colours indicate different models. Different sized markers are used so points on top of each other can be differentiated. Dashed horizontal line indicates the threshold value of Pareto k (0.7) used to judge if the importance sampling approximation is reliable. For all models, the importance sampling approximation fails for the data point with index 54 indicated as a vertical red line. To obtain the LOO score for data point 54 the model was refit.

diagnostic. If for any left-out data point $k > 0.7$, we do not use the PSIS approximation, and instead perform a full refit to evaluate Eq. (87) exactly. Fig. 57 shows the Pareto K diagnostic for each left-out data point for all models. For every model the approximation was reliable for all left-out data points apart from D_{53} . For all models we therefore computed this LOO term exactly by refitting the model with D_{53} left-out. Overall, with PSIS, we obtained an approximate LOO score for each model with only one additional refit of the model compared to the 81 refits required to compute LOO exactly.

To further check the safety of the PSIS approximation, I computed all LOO terms for the model, \mathcal{M}_{DC} , exactly and compared them to the PSIS approximation. Fig. 58 shows the comparison of the PSIS approximation and the exactly computed LOO terms from \mathcal{M}_{DC} . The PSIS approximation matches the exact LOO computation well for all left-out data points apart from D_{54} . Encouragingly, D_{54} is the same data point that was flagged by PSIS as the approximation being unreliable (Fig. 57). Consequently, the model was refitted and the exact

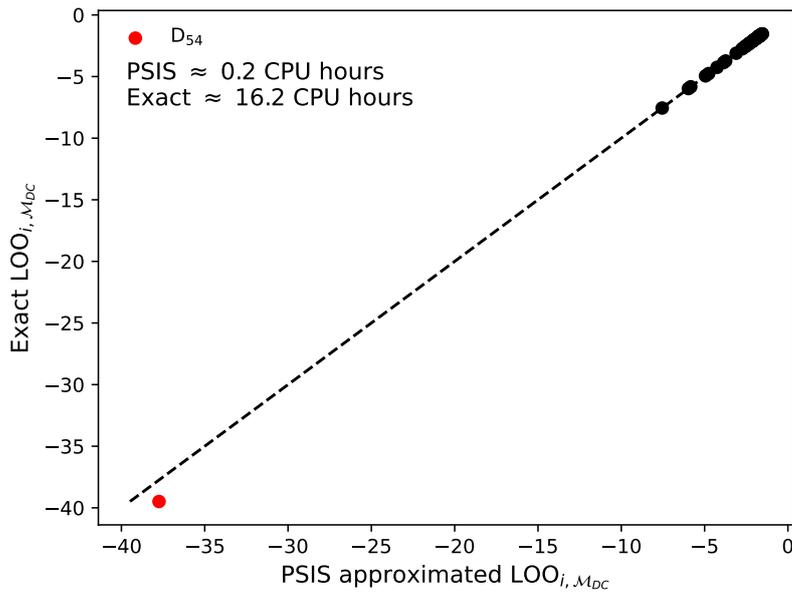


Figure 58: PSIS approximation versus exact computation of the LOO terms for \mathcal{M}_{DC} . The PSIS is in good agreement with the exact values. The dashed line is the one-to-one relationship. The data point that failed the PSIS approximation D_{53} , is marked in red. The approximate CPU computation times for the exact and PSIS methods are shown as text.

LOO value was used for D_{54} . The large negative value of $\text{LOO}_{54, \mathcal{M}_{DC}}$ indicates that D_{54} is an outlying data point which is not well predicted by \mathcal{M}_{DW} . The failure of PSIS approximation for D_{54} suggests that the removal of this data point changes the posterior by a large amount compared with the removal of all other data points. Fig 58 also shows that the PSIS approximation was ≈ 80 times faster than the exact LOO computation.

A.2 POSTERIOR APPROXIMATIONS

In this Section we briefly compare posteriors obtained via MCMC in this work compared with other approximate methods to fit for Θ_E used in the literature (e.g. [Sahu et al., 2017](#); [Zurlo et al., 2018](#)).

A.2.1 The Laplace approximation

The first alternative fitting method we consider is the Laplace approximation to the posterior distribution. The Laplace approximation

works by matching a Gaussian distribution at the mode of the posterior (maximum a posterior estimate, $\vec{\theta}_{\text{MAP}}$) by numerical minimization of $-\ln p(\vec{\theta}|\mathcal{D}, \mathcal{M}) \propto -\ln p(\vec{\theta}, \mathcal{D}|\mathcal{M})$. The curvature around this posterior mode in each of the parameters is calculated using the Hessian matrix \mathbf{H} . Overall, a mean and covariance matrix are calculated as,

$$\vec{\theta}_{\text{MAP}} = \underset{\vec{\theta}}{\text{argmin}} \left[-\ln p(\vec{\theta}, \mathcal{D}|\mathcal{M}) \right], \boldsymbol{\Sigma} = \mathbf{H}^{-1} \quad (96)$$

where,

$$\mathbf{H}_{ij} = - \left. \frac{\partial^2 \ln p(\vec{\theta}, \mathcal{D}|\mathcal{M})}{\partial \vec{\theta}_i \partial \vec{\theta}_j} \right|_{\vec{\theta}=\vec{\theta}_{\text{MAP}}}, \quad (97)$$

respectively. The Laplace approximated posterior is then,

$$p(\vec{\theta}, \mathcal{D}, \mathcal{M}) \approx \mathcal{N}(\vec{\theta}|\vec{\theta}_{\text{MAP}}, \boldsymbol{\Sigma}). \quad (98)$$

Fig. 59 shows the Laplace approximation of the posterior on Θ_E compared with the posterior distribution obtained via MCMC for the \mathcal{M}_{DC} and \mathcal{M}_{DW} models. For \mathcal{M}_{DW} , Fig. 59 shows that the Laplace approximation matches the full MCMC posterior well. In contrast, for \mathcal{M}_{DC} , the Laplace approximation is poor and overconfident (lower variance) compared with the full MCMC posterior. The Laplace approximation is poor here because it fits the curvature of the peak at the \mathcal{M}_{DC} posterior distribution which is not representative of its heavy tails. (Fig. 27 shows this most clearly). It is noted that the posterior mode in both models is well captured by the Laplace approximation.

A.2.2 Monte Carlo MAP estimate for correlated noise

Another commonly used approach, especially in physics, is a non-linear χ^2 optimization (see e.g. [Bevington & Robinson, 2003](#)) combined with a Monte Carlo simulation to estimate the range of possible values of the parameter of interest in a model. The χ^2 of a particular model is in this case,

$$\chi^2(\vec{\theta}) = \sum_{i=1}^N \sum_{d \in \{X, Y\}} \frac{(\text{data}_{i,d} - \text{model}_{i,d}(\vec{\theta}))^2}{\text{error}_{i,d}^2}. \quad (99)$$

i is indexed over the data points and d is indexed over the directions. $\text{model}_{i,d}(\vec{\theta})$ indicates the model prediction for parameter values $\vec{\theta}$

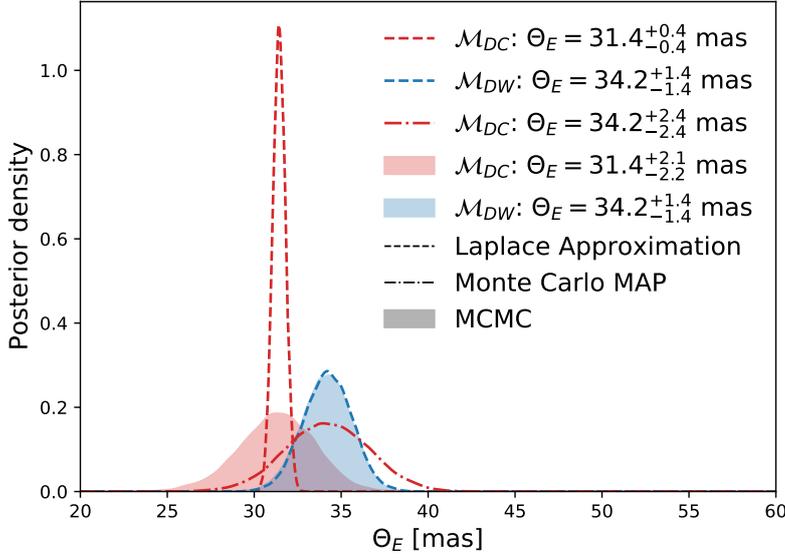


Figure 59: Comparison of the Laplace approximation to the marginal posterior distribution of Θ_E (described in Section A.2.1), the Monte Carlo MAP approximation (described in Section A.2.2), and with the posterior distribution obtained via MCMC. The comparison is shown for both deflection models.

for data point i in direction d . $\text{error}_{i,d}$ is the error at data point i in direction d . This is proportional to the likelihood of models with white Gaussian noise. Further penalty terms representing priors can be added to Eq. (99).

The χ^2 and Monte Carlo procedure essentially splits the modeling into two parts, a part considering the purely random statistical uncertainty (σ_{stat}), and the other dealing with the systematic uncertainty of the model (σ_{sys}). Once these two uncertainties are determined, a Gaussian distribution for the inferred value of the parameter of interest is assumed with total standard deviation $\sqrt{\sigma_{\text{sys}}^2 + \sigma_{\text{stat}}^2}$.

For the purposes of this analysis, this method can be used to approximate the posterior on Θ_E , for the correlated noise deflection model \mathcal{M}_{DC} . The model is split into the statistical and systematic parts in the following way. The statistical part of \mathcal{M}_{DC} is chosen to be the white noise part of the model only, ignoring the correlated terms. Essentially this is equivalent to first performing a χ^2 minimization on \mathcal{M}_{DW} , which is equivalent determining $\vec{\theta}_{\text{MAP}}$ with the inclusion of prior information. The reported value for Θ_E is then the MAP estimate of \mathcal{M}_{DW} or $\Theta_E^{\text{MAP}, \mathcal{M}_{DW}}$. σ_{stat} is determined by the Laplace approximation (as described in Section A.2.1) of \mathcal{M}_{DW}

around $\Theta_E^{\text{MAP}, \mathcal{M}_{\text{DW}}}$, which we denote $\sigma_{\text{stat}}^2 = \Sigma_{\Theta_E}^{\mathcal{M}_{\text{DW}}}$ or the diagonal Θ_E component of the covariance matrix Σ described in Eq. (96) for \mathcal{M}_{DW} .

The uncertainty that the correlated noise part of the model introduces is then dealt with separately and is summarised as a systematic error by σ_{sys} . σ_{sys} is calculated by corrupting the data (creating a new data set $\tilde{\mathcal{D}}^r$) with a realization of the correlated noise and recalculating the MAP estimate for \mathcal{M}_{DW} , which we denote $\Theta_E^{\text{MAP}, \mathcal{M}_{\text{DW}}, \tilde{\mathcal{D}}^r}$, where r is one realization. This is then repeated R times in a Monte Carlo simulation, building a distribution of values of Θ_E . The variance of this distribution is taken to be σ_{sys}^2 or,

$$\sigma_{\text{sys}}^2 = \text{Var} \left(\left\{ \Theta_E^{\text{MAP}, \mathcal{M}_{\text{DW}}, \tilde{\mathcal{D}}^r} \right\}_{r=1}^R \right). \quad (100)$$

The posterior distribution of Θ_E , of the correlated noise model \mathcal{M}_{DC} , is then approximated as,

$$p(\Theta_E | \mathcal{D}, \mathcal{M}_{\text{DC}}) \approx \mathcal{N} \left(\Theta_E | \Theta_E^{\text{MAP}, \mathcal{M}_{\text{DW}}}, \Sigma_{\Theta_E}^{\mathcal{M}_{\text{DW}}} + \sigma_{\text{sys}}^2 \right). \quad (101)$$

Here σ_{sys}^2 is defined in Eq. (100) and is not expressed in its full form for brevity.

Fig. 59 shows the Monte Carlo MAP Θ_E posterior approximation (using $R=100$) compared to the MCMC estimate and the Laplace approximation. We determine $\sigma_{\text{stat}} = 1.4$ mas and $\sigma_{\text{sys}} = 1.9$ mas. The Monte Carlo MAP estimate provides a reasonable estimate for the posterior spread (± 2.4 mas) compared with the MCMC posterior (${}^{+2.1}_{-2.2}$ mas), if not slightly over-estimated. This difference is due to the Monte Carlo MAP approximation adding additional scatter to the data and violating the implicitly assumed data generation process for \mathcal{M}_{DC} (Eq. 49).

The concerning issue, however, is that the Monte Carlo MAP approximation is biased towards high values (median: 34.3 mas) of Θ_E compared to the MCMC posterior (median : 31.4 mas) for \mathcal{M}_{DC} . The higher median value for the Monte Carlo MAP approximation for \mathcal{M}_{DC} matches the MCMC posterior median for the white noise model, \mathcal{M}_{DW} . This is because the mean of the Monte Carlo MAP Gaussian approximation for \mathcal{M}_{DC} is calculated using the likelihood of the \mathcal{M}_{DW} model. Similar to the spread of the posterior, the Monte Carlo MAP approximation fails to approximate the posterior because it violates the implicit data generation assumptions in \mathcal{M}_{DC} by using a MAP estimate from \mathcal{M}_{DW} . Specifically, the Monte Carlo MAP ap-

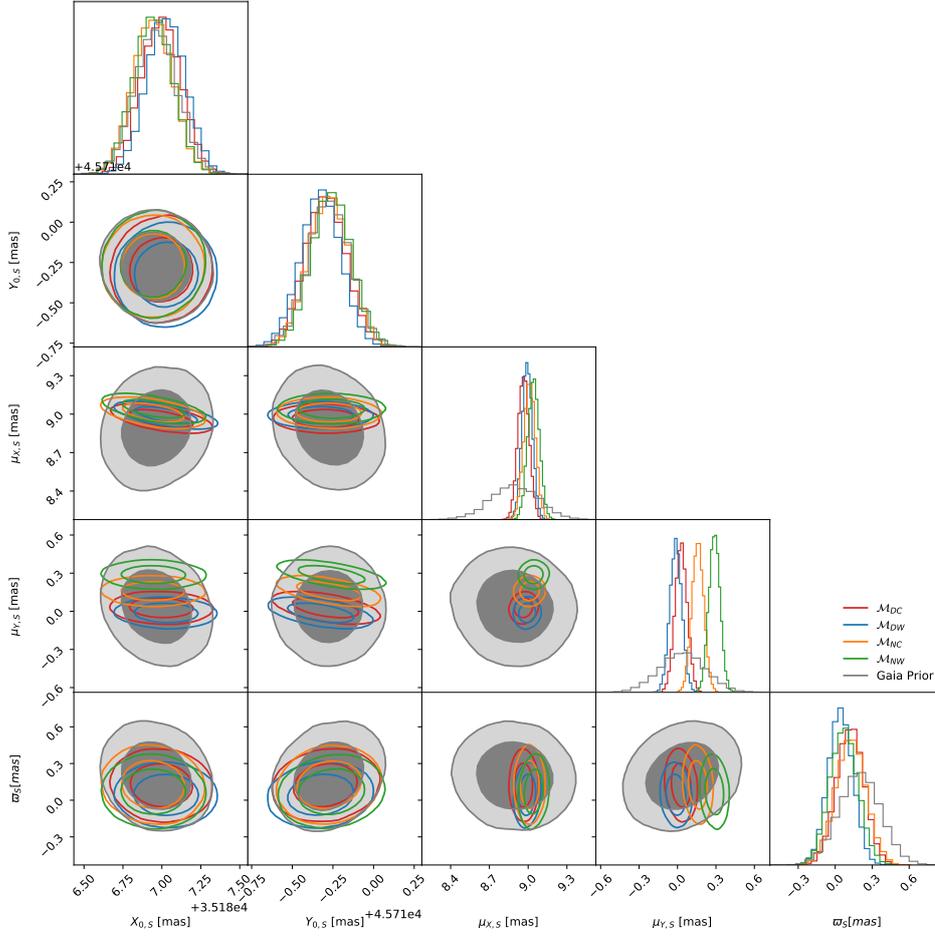


Figure 60: GEDR3 priors versus posterior inferences on all source astrometric parameters for all the models considered. The plot is a full version of the densities shown in Fig. 26.

proximation does not allow the correlated noise part \mathcal{M}_{DC} to interact with the deflection part of the model. This produces a biased inference on Θ_E because, as Section 3.2.4 suggests, the correlated noise can mimic and is somewhat degenerate with the deflection signal.

A.3 CONSISTENCY WITH GAIA

This Section contains the full corner plots for both the source (Fig. 60) and lens (Fig. 61) prior and posterior distributions of their astrometric parameters. For both the lens and source, and all models, it can be seen that the posteriors are in good agreement with the GEDR3 priors.

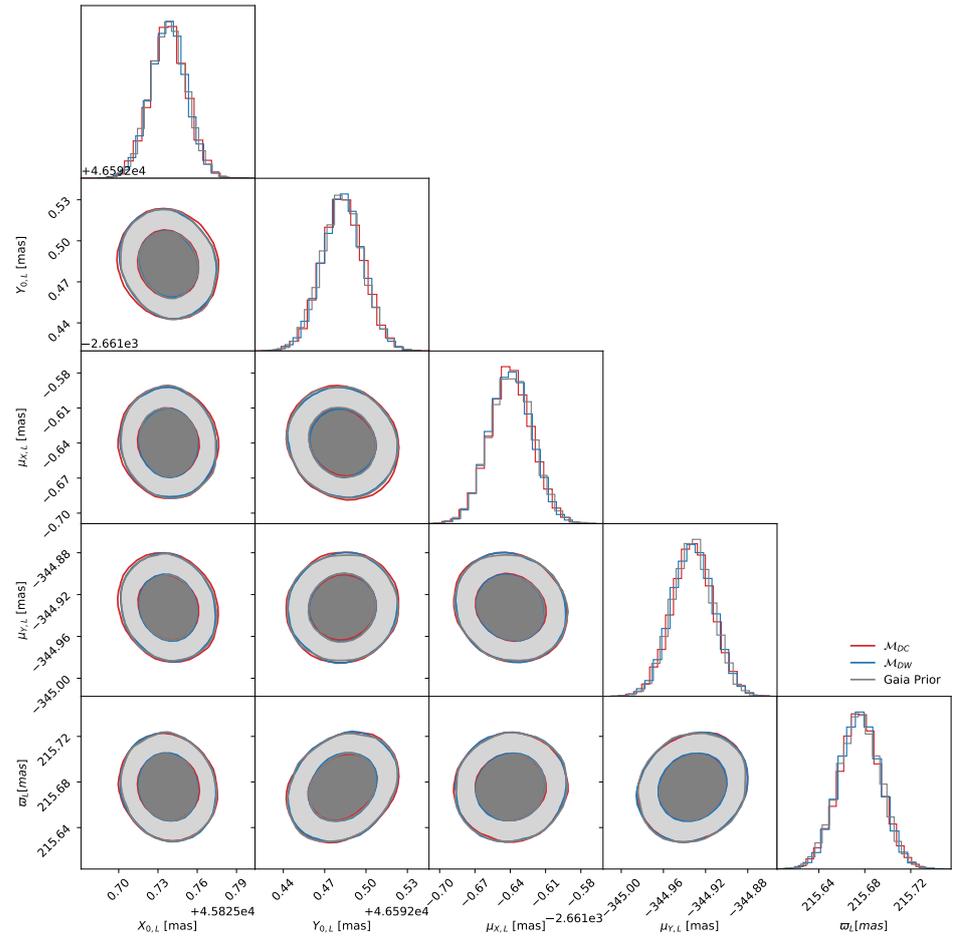


Figure 61: GEDR3 priors versus posterior inferences on all the lens (LAWD 37) astrometric parameters for the two deflection models which contain the astrometry as free parameters. For both these models, the HST data provides no further constraint from GEDR3.

BIBLIOGRAPHY

- Abbas U., Bucciarelli B., Lattanzi M. G., 2019, *MNRAS*, 485, 1147
- Afonso C., et al., 2003, *A&A*, 404, 145
- Ak T., Bilir S., Ak S., Eker Z., 2008, *New Astron.*, 13, 133
- Akaike H., 1981, *Journal of econometrics*, 16, 3
- Alard C., Guibert J., Bienayme O., Valls-Gabaud D., Robin A. C., Terzan A., Bertin E., 1995, *The Messenger*, 80, 31
- Alcock C., et al., 1993, *Nature*, 365, 621
- Alcock C., et al., 1995, *ApJ*, 445, 133
- Alcock C., et al., 1997, *ApJ*, 486, 697
- Alcock C., et al., 2000a, *ApJ*, 542, 281
- Alcock C., et al., 2000b, *ApJ*, 542, 281
- Allard F., 2014, in Booth M., Matthews B. C., Graham J. R., eds, *IAU Symposium Vol. 299, Exploring the Formation and Evolution of Planetary Systems*. pp 271–272, doi:10.1017/S1743921313008545
- Allard F., Homeier D., Freytag B., 2012, *Philosophical Transactions of the Royal Society of London Series A*, 370, 2765
- Alonso-García J., Mateo M., Sen B., Banerjee M., Catelan M., Minniti D., von Braun K., 2012, *AJ*, 143, 70
- Altmann M., Roeser S., Demleitner M., Bastian U., Schilbach E., 2017, *A&A*, 600, L4
- An J. H., et al., 2002, *ApJ*, 572, 521
- Andrae R., Schulze-Hartung T., Melchior P., 2010, arXiv e-prints, p. arXiv:1012.3754
- Angeloni R., et al., 2014, *A&A*, 567, A100
- Arenou F., et al., 2018, *A&A*, 616, A17
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Aubourg E., et al., 1993, *Nature*, 365, 623
- Awiphan S., Kerins E., Robin A. C., 2016, *MNRAS*, 456, 1666

- Baillon P., Bouquet A., Giraud-Heraud Y., Kaplan J., 1993, *A&A*, **277**, 1
- Baraffe I., Chabrier G., Allard F., Hauschildt P. H., 1997, *A&A*, **327**, 1054
- Batista V., Beaulieu J. P., Bennett D. P., Gould A., Marquette J. B., Fukui A., Bhattacharya A., 2015, *ApJ*, **808**, 170
- Bayliss D., et al., 2018, *MNRAS*, **475**, 4467
- Bayo A., Rodrigo C., Barrado Y Navascués D., Solano E., Gutiérrez R., Morales-Calderón M., Allard F., 2008, *A&A*, **492**, 277
- Beaulieu J. P., et al., 2006, *Nature*, **439**, 437
- Bédard A., Bergeron P., Brassard P., Fontaine G., 2020, *ApJ*, **901**, 93
- Bellini A., Anderson J., Bedin L. R., 2011, *PASP*, **123**, 622
- Belokurov V. A., Evans N. W., 2002, *MNRAS*, **331**, 649
- Belokurov V., Evans N. W., Du Y. L., 2003, *MNRAS*, **341**, 1373
- Belokurov V., et al., 2020, *MNRAS*, **496**, 1922
- Bennett D., 1998, *Phys. Rep.*, **307**, 97
- Bennett D. P., 2005, *ApJ*, **633**, 906
- Bennett D. P., et al., 2018, *AJ*, **156**, 113
- Bergeron P., Leggett S. K., Ruiz M. T., 2001, *ApJS*, **133**, 413
- Best W. M. J., Liu M. C., Magnier E. A., Dupuy T. J., 2020, *AJ*, **159**, 257
- Bevington P. R., Robinson D. K., 2003, McGraw Hill, New York
- Blouin S., Dufour P., 2019, *MNRAS*, **490**, 4166
- Blouin S., Dufour P., Allard N. F., 2018, *ApJ*, **863**, 184
- Boden A. F., Shao M., Van Buren D., 1998, *ApJ*, **502**, 538
- Bond I. A., et al., 2004, *ApJ*, **606**, L155
- Bond H. E., et al., 2015, *ApJ*, **813**, 106
- Bond H. E., et al., 2017a, *ApJ*, **840**, 70
- Bond H. E., Bergeron P., Bédard A., 2017b, *ApJ*, **848**, 16
- Bond H. E., Schaefer G. H., Gilliland R. L., Vandenberg D. A., 2020, *ApJ*, **904**, 112
- Borucki W. J., et al., 2010, *Science*, **327**, 977
- Boubert D., Everall A., 2020, *MNRAS*, **497**, 4246

- Boubert D., Everall A., Holl B., 2020, *MNRAS*, 497, 1826
- Bramich D. M., 2018, *A&A*, 618, A44
- Bramich D. M., Nielsen M. B., 2018, *Acta Astron.*, 68, 183
- Branch M. A., Coleman T. F., Li Y., 1999, *SIAM Journal on Scientific Computing*, 21, 1
- Breiman L., 2001, *Machine Learning*, 45, 5
- Breiman L., Friedman J. H., Olshen R. A., Stone C. J., 1984, *Classification and regression trees*. The Wadsworth & Brooks/Cole statistics/probability series, Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA, <https://cds.cern.ch/record/2253780>
- Broyden C. G., 1970, *IMA Journal of Applied Mathematics*, 6, 76
- Bürkner P.-C., Gabry J., Vehtari A., 2020, *Computational Statistics*, pp 1–19
- Cabral J. B., Ramos F., Gurovich S., Granitto P. M., 2020, *A&A*, 642, A58
- Calchi Novati S., et al., 2005a, *A&A*, 443, 911
- Calchi Novati S., et al., 2005b, *A&A*, 443, 911
- Calchi Novati S., et al., 2015, *ApJ*, 804, 20
- Casertano S., et al., 2016, *ApJ*, 825, 11
- Chabrier G., Baraffe I., 1997, *A&A*, 327, 1039
- Chandra V., Hwang H.-C., Zakamska N. L., Cheng S., 2020, *ApJ*, 899, 146
- Chandrasekhar S., 1935, *MNRAS*, 95, 207
- Chang K., Refsdal S., 1979, *Nature*, 282, 561
- Chaplin W. J., et al., 2011, *Science*, 332, 213
- Chwolson O., 1924, *Astronomische Nachrichten*, 221, 329
- Claudi R., et al., 2019, *A&A*, 622, A96
- Congdon A. B., Keeton C., 2018, *Principles of Gravitational Lensing: Light Deflection as a Probe of Astrophysics and Cosmology*. Springer
- Coutu S., Dufour P., Bergeron P., Blouin S., Loranger E., Allard N. F., Dunlap B. H., 2019, *ApJ*, 885, 74
- Crotts A. P. S., 1992, *ApJ*, 399, L43
- Di Stefano R., 2008a, *ApJ*, 684, 46

- Di Stefano R., 2008b, *ApJ*, 684, 59
- Di Stefano R., Perna R., 1997, *ApJ*, 488, 55
- Dittmann J. A., et al., 2017, *Nature*, 544, 333
- Dominik M., Sahu K. C., 2000, *ApJ*, 534, 213
- Dong S., et al., 2019, *ApJ*, 871, 70
- Dotter A., Chaboyer B., Jevremović D., Kostov V., Baron E., Ferguson J. W., 2008, *ApJS*, 178, 89
- Dravins D., Lindegren L., Madsen S., 1999, *A&A*, 348, 1040
- Drummond J. D., Christou J. C., Fugate R. Q., 1995, *ApJ*, 450, 380
- Dufour P., Bergeron P., Fontaine G., 2005, *ApJ*, 627, 404
- Dyson F. W., Eddington A. S., Davidson C., 1920, *Philosophical Transactions of the Royal Society of London Series A*, 220, 291
- Einstein A., 1916, *Annalen der Physik*, 354, 769
- Einstein A., 1936, *Science*, 84, 506
- Eker Z., et al., 2018, *MNRAS*, 479, 5491
- Erben T., et al., 2013, *MNRAS*, 433, 2545
- Evans N. W., 1994, *ApJ*, 437, L31
- Evans N. W., 2003, in Valls-Gabaud D., Kneib J.-P., eds, *Gravitational Lensing: A unique tool for Cosmology*. ([arXiv:astro-ph/0304252](https://arxiv.org/abs/astro-ph/0304252))
- Evans N. W., Belokurov V., 2002, *ApJ*, 567, L119
- Evans D. W., et al., 2018, *A&A*, 616, A4
- Everall A., Boubert D., Kuposov S. E., Smith L., Holl B., 2021, *MNRAS*, 502, 1908
- Fabrizius C., et al., 2016, *A&A*, 595, A3
- Falcon R. E., Winget D. E., Montgomery M. H., Williams K. A., 2010, *ApJ*, 712, 585
- Feibelman W. A., 1966, *Science*, 151, 73
- Feibelman W. A., 1986, *PASP*, 98, 1199
- Fong E., Holmes C., 2020, *Biometrika*, 107, 489
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, *ApJ*, 795, 64

- Foreman-Mackey D., et al., 2021, *The Journal of Open Source Software*, 6, 3285
- Freudling W., Romaniello M., Bramich D. M., Ballester P., Forchi V., García-Dabló C. E., Moehler S., Neeser M. J., 2013, *A&A*, 559, A96
- Gaia Collaboration et al., 2016a, *A&A*, 595, A1
- Gaia Collaboration et al., 2016b, *A&A*, 595, A2
- Gaia Collaboration et al., 2018, *A&A*, 616, A1
- Gaia Collaboration et al., 2021a, *A&A*, 649, A1
- Gaia Collaboration et al., 2021b, *A&A*, 649, A6
- Gaidos E., et al., 2014, *MNRAS*, 443, 2561
- Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, *Bayesian data analysis*. CRC press
- Gelman A., Hwang J., Vehtari A., 2014, *Statistics and computing*, 24, 997
- Giammichele N., Bergeron P., Dufour P., 2012, *ApJS*, 199, 29
- Gillon M., et al., 2016, *Nature*, 533, 221
- Godines D., Bachelet E., Narayan G., Street R. A., 2019, *Astronomy and Computing*, 28, 100298
- Golovich N., et al., 2020, arXiv e-prints, p. arXiv:2009.07927
- Goodman J., Weare J., 2010, *Communications in Applied Mathematics and Computational Science*, 5, 65
- Gould A., 1994, *ApJ*, 421, L71
- Gould A., 1995, *ApJ*, 446, L71
- Gould A., 2000, *ApJ*, 532, 936
- Gould A., 2001, *PASP*, 113, 903
- Gould A., Yee J. C., 2013, *ApJ*, 764, 107
- Griest K., 1991, *ApJ*, 366, 412
- Griest K., et al., 1991, *ApJ*, 372, L79
- Hamada T., Salpeter E. E., 1961, *ApJ*, 134, 683
- Han C., Gould A., 2003, *ApJ*, 592, 172
- Hansen B. M. S., et al., 2002, *ApJ*, 574, L155
- Harding A. J., Stefano R. D., Lépine S., Urama J., Pham D., Baker C., 2018, *MNRAS*, 475, 79

- Harris C. R., et al., 2020, *Nature*, **585**, 357
- Hastie T., Tibshirani R., Friedman J., 2009, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media
- Henderson C. B., et al., 2016, *PASP*, **128**, 124401
- Hertzsprung E., 1923, *Bull. Astron. Inst. Netherlands*, **2**, 15
- Higson E., Handley W., Hobson M., Lasenby A., 2019, *Statistics and Computing*, **29**, 891
- Hodgkin S. T., et al., 2021, arXiv e-prints, p. [arXiv:2106.01394](https://arxiv.org/abs/2106.01394)
- Hog E., Novikov I. D., Polnarev A. G., 1995, *A&A*, **294**, 287
- Hogg D. W., Myers A. D., Bovy J., 2010, *ApJ*, **725**, 2166
- Hojjatpanah S., et al., 2019, *A&A*, **629**, A80
- Holberg J. B., Sion E. M., Oswalt T., McCook G. P., Foran S., Subasavage J. P., 2008, *AJ*, **135**, 1225
- Holz D. E., Wald R. M., 1996, *ApJ*, **471**, 64
- Homan M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, **15**, 1593–1623
- Honma M., 2001, *PASJ*, **53**, 233
- Hunter J. D., 2007, *Computing in Science & Engineering*, **9**, 90
- Hurley J. R., Tout C. A., Pols O. R., 2002, *MNRAS*, **329**, 897
- Husseiniova A., McGill P., Smith L. C., Evans N. W., 2021, *MNRAS*, **506**, 2482
- Iben I. J., Tutukov A. V., 1984, *ApJ*, **282**, 615
- Irwin M. J., Webster R. L., Hewett P. C., Corrigan R. T., Jedrzejewski R. I., 1989, *AJ*, **98**, 1989
- Jao W.-C., Nelan E. P., Henry T. J., Franz O. G., Wasserman L. H., 2016, *AJ*, **152**, 153
- Jones E., Oliphant T., Peterson P., et al., 2001, *SciPy: Open source scientific tools for Python*, <http://www.scipy.org/>
- Joyce S. R. G., Barstow M. A., Holberg J. B., Bond H. E., Casewell S. L., Burleigh M. R., 2018, *MNRAS*, **481**, 2361
- Kains N., Bramich D. M., Sahu K. C., Calamida A., 2016, *MNRAS*, **460**, 2025
- Kains N., et al., 2017, *ApJ*, **843**, 145

- Kains N., Calamida A., Sahu K. C., Anderson J., Casertano S., Bramich D. M., 2018, *ApJ*, 867, 37
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- Karaali S., Bilir S., Karataş Y., Ak S. G., 2003, *Publ. Astron. Soc. Australia*, 20, 165
- Karolinski N., Zhu W., 2020, *MNRAS*, 498, L25
- Kayser R., Refsdal S., Stabell R., 1986, *A&A*, 166, 36
- Kerins E., Robin A. C., Marshall D. J., 2009, *MNRAS*, 396, 1202
- Kervella P., Arenou F., Mignard F., Thévenin F., 2019, *A&A*, 623, A72
- Khakpash S., Pepper J., Penny M., Gaudi B. S., Street R. A., 2021, *AJ*, 161, 132
- Kilic M., Bergeron P., Kosakowski A., Brown W. R., Agüeros M. A., Blouin S., 2020, *ApJ*, 898, 84
- Kim S.-L., et al., 2016, *Journal of Korean Astronomical Society*, 49, 37
- Kiraga M., Paczynski B., 1994, *ApJ*, 430, L101
- Klüter J., Bastian U., Demleitner M., Wambsganss J., 2018a, *A&A*, 615, L11
- Klüter J., Bastian U., Demleitner M., Wambsganss J., 2018b, *A&A*, 620, A175
- Klüter J., Bastian U., Wambsganss J., 2020, *A&A*, 640, A83
- Kneib J.-P., Natarajan P., 2011, *A&ARv*, 19, 47
- Koester D., Weidemann V., 1982, *A&A*, 108, 406
- Konacki M., Muterspaugh M. W., Kulkarni S. R., Hełminiak K. G., 2010, *ApJ*, 719, 1293
- Koposov S., Bartunov O., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, *Astronomical Society of the Pacific Conference Series Vol. 351, Astronomical Data Analysis Software and Systems XV*. p. 735
- Kozłowski S., Woźniak P. R., Mao S., Wood A., 2007, *ApJ*, 671, 420
- Kumar R., Carroll C., Hartikainen A., Martin O., 2019, *The Journal of Open Source Software*, 4, 1143
- Lam C. Y., Lu J. R., Hosek Matthew W. J., Dawson W. A., Golovich N. R., 2020, *ApJ*, 889, 31
- Lemeshko S., 2006, *Measurement Techniques*, 49, 962
- Lépine S., DiStefano R., 2012, *ApJ*, 749, L6

- Lépine S., Shara M. M., 2005, *AJ*, 129, 1483
- Lépine S., Rich R. M., Shara M. M., 2007, *ApJ*, 669, 1235
- Li S. S., et al., 2019, *MNRAS*, 488, 3308
- Liebes S., 1964, *Physical Review*, 133, 835
- Lindegren L., Dravins D., 2021, arXiv e-prints, p. [arXiv:2105.09014](https://arxiv.org/abs/2105.09014)
- Lindegren L., et al., 2018, *A&A*, 616, A2
- Lindegren L., et al., 2021, *A&A*, 649, A2
- Lintott C., et al., 2011, *MNRAS*, 410, 166
- Lynds R., Petrosian V., 1989, *ApJ*, 336, 1
- Mao S., 2012, *Research in Astronomy and Astrophysics*, 12, 947
- Mao S., Paczynski B., 1991, *ApJ*, 374, L37
- Mao S., Paczynski B., 1996, *ApJ*, 473, 57
- Marrone D. P., et al., 2018, *Nature*, 553, 51
- McDonald I., et al., 2021, *MNRAS*, 505, 5584
- McGill P., Smith L. C., Evans N. W., Belokurov V., Smart R. L., 2018, *MNRAS*, 478, L29
- McGill P., Smith L. C., Evans N. W., Belokurov V., Zhang Z. H., 2019a, *MNRAS*, 483, 4210
- McGill P., Smith L. C., Evans N. W., Belokurov V., Lucas P. W., 2019b, *MNRAS*, 487, L7
- McGill P., Everall A., Boubert D., Smith L. C., 2020, *MNRAS*, 498, L6
- Medford M. S., Lu J. R., Dawson W. A., Lam C. Y., Golovich N. R., Schlafly E. F., Nugent P., 2020, *ApJ*, 897, 144
- Medina N., et al., 2018, *ApJ*, 864, 11
- Michalik D., Lindegren L., Hobbs D., 2015, *A&A*, 574, A115
- Minniti D., 2018, in Gionti G., Kikwaya Eluo J.-B., eds, Vol. 51, The Vatican Observatory, Castel Gandolfo: 80th Anniversary Celebration. p. 63, [doi:10.1007/978-3-319-67205-2_4](https://doi.org/10.1007/978-3-319-67205-2_4)
- Minniti D., et al., 2010, *New Astron.*, 15, 433
- Miralda-Escude J., 1996, *ApJ*, 470, L113
- Miyamoto M., Yoshii Y., 1995, *AJ*, 110, 1427
- Miyazaki S., Johnson S. A., Sumi T., Penny M. T., Koshimoto N., Yamawaki T., 2021, *AJ*, 161, 84

- Morton T. D., 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library (ascl:1503.010)
- Mróz P., 2020, *Acta Astron.*, 70, 169
- Mróz P., et al., 2017, *Nature*, 548, 183
- Mróz P., et al., 2019, *ApJS*, 244, 29
- Mróz P., et al., 2020a, *ApJS*, 249, 16
- Mróz P., et al., 2020b, *ApJ*, 903, L11
- Mustill A. J., Davies M. B., Lindegren L., 2018, *A&A*, 617, A135
- Navarro M. G., Minniti D., Contreras Ramos R., 2017, *ApJ*, 851, L13
- Navarro M. G., Minniti D., Contreras-Ramos R., 2018, *ApJ*, 865, L5
- Navarro M. G., Minniti D., Pullen J., Ramos R. C., 2020a, *ApJ*, 889, 56
- Navarro M. G., Contreras Ramos R., Minniti D., Pullen J., Capuzzo-Dolcetta R., Lucas P. W., 2020b, *ApJ*, 893, 65
- Navarro M. G., Minniti D., Contreras Ramos R., 2020c, *ApJ*, 902, 35
- Nemiroff R. J., Wickramasinghe W. A. D. T., 1994, *ApJ*, 424, L21
- Nielsen M. B., Bramich D. M., 2018, *Acta Astron.*, 68, 351
- Nucita A. A., de Paolis F., Ingrassio G., Giordano M., Manni L., 2017, *International Journal of Modern Physics D*, 26, 1741015
- Ofek E. O., 2018, *ApJ*, 866, 144
- Ofek E. O., et al., 2012, *PASP*, 124, 854
- Oguri M., Rusu C. E., Falco E. E., 2014, *MNRAS*, 439, 2494
- Paczynski B., 1986a, *ApJ*, 301, 503
- Paczynski B., 1986b, *ApJ*, 304, 1
- Paczynski B., 1991, *ApJ*, 371, L63
- Paczynski B., 1995, *Acta Astron.*, 45, 345
- Paczynski B., 1996a, *ARA&A*, 34, 419
- Paczynski B., 1996b, *Acta Astron.*, 46, 291
- Paczynski B., Stanek K. Z., Udalski A., Szymanski M., Kaluzny J., Kubiak M., Mateo M., Krzeminski W., 1994, *ApJ*, 435, L113
- Park B. G., et al., 2004, *ApJ*, 609, 166
- Parsons S. G., et al., 2016, *MNRAS*, 458, 2793

- Pasquini L., Pala A. F., Ludwig H. G., Leao I. C., de Medeiros J. R., Weiss A., 2019, *A&A*, **627**, L8
- Paulin-Henriksson S., et al., 2002, *ApJ*, **576**, L121
- Paulin-Henriksson S., et al., 2003, *A&A*, **405**, 15
- Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, **12**, 2825
- Pelletier C., Fontaine G., Wesemael F., Michaud G., Wegner G., 1986, *ApJ*, **307**, 242
- Penny M. T., Gaudi B. S., Kerins E., Rattenbury N. J., Mao S., Robin A. C., Calchi Novati S., 2019, *ApJS*, **241**, 3
- Popowski P., et al., 2001, in Menzies J. W., Sackett P. D., eds, *Astronomical Society of the Pacific Conference Series Vol. 239, Microlensing 2000: A New Era of Microlensing Astrophysics*. p. 244 ([arXiv:astro-ph/0005466](https://arxiv.org/abs/astro-ph/0005466))
- Price-Whelan A. M., et al., 2014, *ApJ*, **781**, 35
- Proft S., 2016, PhD thesis, University of Heidelberg
- Proft S., Demleitner M., Wambsganss J., 2011, *A&A*, **536**, A50
- Provencal J. L., Shipman H. L., Wesemael F., Bergeron P., Bond H. E., Liebert J., Sion E. M., 1997, *ApJ*, **480**, 777
- Quinlan J. R., 2014, *C4. 5: programs for machine learning*. Elsevier
- Rajpurohit A. S., Reyl e C., Allard F., Homeier D., Schultheis M., Bessell M. S., Robin A. C., 2013, *A&A*, **556**, A15
- Refsdal S., 1964, *MNRAS*, **128**, 295
- Refsdal S., 1966, *MNRAS*, **134**, 315
- Reid I. N., Hawley S. L., Gizis J. E., 1995, *AJ*, **110**, 1838
- Robinson E. L., 1976, *ARA&A*, **14**, 119
- Roeser S., Demleitner M., Schilbach E., 2010, *AJ*, **139**, 2440
- Rybicki K. A., Wyrzykowski  ., Klencki J., de Bruijne J., Belczyński K., Chru lińska M., 2018, *MNRAS*, **476**, 2013
- Sahu K. C., Chaney E., Graham J., Kane S., Wieldt D., 1998, in *American Astronomical Society Meeting Abstracts #192*. p. 07.01
- Sahu K. C., Bond H. E., Anderson J., Dominik M., 2014, *ApJ*, **782**, 89
- Sahu K. C., et al., 2017, *Science*, **356**, 1046

- Sahu K. C., et al., 2020, Mass determination of an extreme halo M subdwarf through astrometric and photometric microlensing, HST Proposal
- Sajadian S., Poleski R., 2019, *ApJ*, 871, 205
- Salim S., Gould A., 2000, *ApJ*, 539, 241
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Computer Science*, 2, e55
- Schechter P. L., Mateo M., Saha A., 1993, *PASP*, 105, 1342
- Schlafly E. F., et al., 2018, *ApJS*, 234, 39
- Schmidt H., 1996, *A&A*, 311, 852
- Schroeder D. J., et al., 2000, *AJ*, 119, 906
- Seager S., Deming D., 2010, *ARA&A*, 48, 631
- Seager S., Mallén-Ornelas G., 2003, *ApJ*, 585, 1038
- Serenelli A., Weiss A., Cassisi S., Salaris M., Pietrinferni A., 2017, *A&A*, 606, A33
- Serenelli A., et al., 2021, *A&ARv*, 29, 4
- Shan Y., Yee J. C., Bailey V. P., Close L. M., Hinz P. M., Males J. R., Morzinski K. M., 2021, *ApJ*, 908, 240
- Shvartzvald Y., Bryden G., Gould A., Henderson C. B., Howell S. B., Beichman C., 2017, *AJ*, 153, 61
- Sion E. M., Holberg J. B., Oswalt T. D., McCook G. P., Wasatonic R., 2009, *AJ*, 138, 1681
- Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, American Institute of Physics Conference Series Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. pp 395–405, doi:10.1063/1.1835238
- Skilling J., et al., 2006, Bayesian analysis, 1, 833
- Skowron J., et al., 2015, *ApJ*, 804, 33
- Skrutskie M. F., et al., 2006, *AJ*, 131, 1163
- Smith M. C., Mao S., Woźniak P., 2002, *MNRAS*, 332, 962
- Smith M. C., Belokurov V., Evans N. W., Mao S., An J. H., 2005, *MNRAS*, 361, 128
- Smith M. C., Woźniak P., Mao S., Sumi T., 2007, *MNRAS*, 380, 805

- Smith L. C., et al., 2018, *MNRAS*, 474, 1826
- Sokolovsky K. V., et al., 2017, *MNRAS*, 464, 274
- Soldner J., 1804, *Berliner Astronomisches Jahrbuch*, 161, 1804
- Soubiran C., Bienaymé O., Mishenina T. V., Kovtyukh V. V., 2008, *A&A*, 480, 91
- Speagle J. S., 2020, *MNRAS*, 493, 3132
- Specht D., Kerins E., Awiphan S., Robin A. C., 2020, *MNRAS*, 498, 2196
- Spiegelhalter D. J., Best N. G., Carlin B. P., Van Der Linde A., 2002, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583
- Storn R., Price K., 1997, *Journal of global optimization*, 11, 341
- Strunov V., 2006, *Measurement Techniques*, 49, 755
- Subasavage J. P., Jao W.-C., Henry T. J., Bergeron P., Dufour P., Ianna P. A., Costa E., Méndez R. A., 2009, *AJ*, 137, 4547
- Sumi T., Penny M. T., 2016, *ApJ*, 827, 139
- Sumi T., et al., 2003, in *American Astronomical Society Meeting Abstracts*. p. 112.27
- Sumi T., et al., 2006, *ApJ*, 636, 240
- Sumi T., et al., 2013, *ApJ*, 778, 150
- Sundararajan S., Keerthi S. S., 2001, *Neural computation*, 13, 1103
- Surot F., Valenti E., Gonzalez O. A., Zoccali M., Sökmen E., Hidalgo S. L., Minniti D., 2020, *A&A*, 644, A140
- Sutherland W., et al., 2015, *A&A*, 575, A25
- Thiemann H. B., Norton A. J., Dickinson H. J., McMaster A., Kolb U. C., 2021, *MNRAS*, 502, 1299
- Tisserand P., et al., 2007, *A&A*, 469, 387
- Torres G., 2006, *AJ*, 131, 1022
- Torres G., Andersen J., Giménez A., 2010, *A&ARv*, 18, 67
- Tremblay P. E., Bergeron P., 2008, *ApJ*, 672, 1144
- Tremblay P.-E., et al., 2017, *MNRAS*, 465, 2849
- Tsapras Y., 2018, *Geosciences*, 8, 365
- Tyson J. A., Valdes F., Wenk R. A., 1990, *ApJ*, 349, L1

- Udalski A., Szymanski M., Kaluzny J., Kubiak M., Mateo M., 1992, *Acta Astron.*, **42**, 253
- Udalski A., et al., 1994, *Acta Astron.*, **44**, 165
- Udalski A., Kubiak M., Szymanski M., 1997, *Acta Astron.*, **47**, 319
- Udalski A., et al., 2005, *ApJ*, **628**, L109
- Udalski A., Szymański M. K., Szymański G., 2015, *Acta Astron.*, **65**, 1
- Urban S. E., Seidemann P. K., 2014, in *American Astronomical Society Meeting Abstracts #223*. p. 247.20
- Vanderriest C., Schneider J., Herpe G., Chevreton M., Moles M., Wierick G., 1989, *A&A*, **215**, 1
- Vehtari A., Simpson D., Gelman A., Yao Y., Gabry J., 2015, arXiv preprint arXiv:1507.02646
- Vehtari A., Gelman A., Gabry J., 2017, *Statistics and computing*, **27**, 1413
- Vehtari A., Gelman A., Simpson D., Carpenter B., Bürkner P.-C., 2021, *Bayesian analysis*, **1**, 1
- Vernet J., et al., 2011, *A&A*, **536**, A105
- Vietri M., Ostriker J. P., 1983, *ApJ*, **267**, 488
- Virtanen P., et al., 2020, *Nature Methods*, **17**, 261
- Wales D. J., Doye J. P. K., 1997, *Journal of Physical Chemistry A*, **101**, 5111
- Walker M. A., 1995, *ApJ*, **453**, 37
- Walsh D., Carswell R. F., Weymann R. J., 1979, *Nature*, **279**, 381
- Weidemann V., Koester D., 1995, *A&A*, **297**, 216
- Witt H. J., Mao S., 1994, *ApJ*, **430**, 505
- Woźniak P., Paczyński B., 1997, *ApJ*, **487**, 55
- Wyrzykowski Ł., Mandel I., 2020, *A&A*, **636**, A20
- Wyrzykowski Ł., Udalski A., Mao S., Kubiak M., Szymanski M. K., Pietrzynski G., Soszynski I., Szewczyk O., 2006, *Acta Astron.*, **56**, 145
- Wyrzykowski Ł., et al., 2009, *MNRAS*, **397**, 1228
- Wyrzykowski Ł., et al., 2011a, *MNRAS*, **413**, 493
- Wyrzykowski Ł., et al., 2011b, *MNRAS*, **416**, 2949

- Wyrzykowski Ł., et al., 2015, *ApJS*, 216, 12
- Wyrzykowski Ł., et al., 2016, *MNRAS*, 458, 3012
- Wyrzykowski Ł., et al., 2020, *A&A*, 633, A98
- York D. G., et al., 2000, *AJ*, 120, 1579
- Zhang Z. H., Homeier D., Pinfield D. J., Lodieu N., Jones H. R. A., Allard F., Pavlenko Y. V., 2017, *MNRAS*, 468, 261
- Zhang Z. H., et al., 2018, *MNRAS*, 480, 5447
- Zhang K., Bloom J. S., Gaudi B. S., Lanusse F., Lam C., Lu J. R., 2021, *AJ*, 161, 262
- Zurlo A., et al., 2018, *MNRAS*, 480, 236
- Zwicky F., 1937, *Physical Review*, 51, 290
- van Leeuwen F., 2007, *A&A*, 474, 653
- von Neumann J., 1941, *The Annals of Mathematical Statistics*, 12, 367
- von Neumann J., 1942, *The Annals of Mathematical Statistics*, 13, 86

COLOPHON

This document was typeset using the typographical look-and-feel classicthesis developed by André Miede.

Final Version as of September 5, 2021 (classicthesis version 4.2).