Challenges in Timeseries Analysis from Microlensing

R. A. Street

6740 Cortona Drive, Suite 102, Goleta, CA 93117, USA
email: `rstreet@lco.global`

Abstract. Despite a flood of discoveries over the last ∼20 years, our knowledge of the exoplanet population is incomplete owing to a gap between the sensitivities of different detection techniques. However, a census of exoplanets at all separations from their host stars is essential to fully understand planet formation mechanisms. Microlensing offers an effective way to bridge the gap around 1–10 AU and is therefore one of the major science goals of the *Wide Field Infrared Survey Telescope (WFIRST)* mission. *WFIRST*'s survey of the Galactic Bulge is expected to discover ∼20,000 microlensing events, including ∼3000 planets, which represents a substantial data analysis challenge with the modeling software currently available. This paper highlights areas where further work is needed. The community is encouraged to join new software development efforts aimed at making the modeling of microlensing events both more accessible and rigorous.

Keywords. gravitational lensing, methods: data analysis, stars: planetary systems

---

1. Introduction

The technique of gravitational microlensing has produced a range of discoveries over the last ∼20 yrs, most notably the detection of exoplanets in "cool" orbits around 1–10 AU from their host stars and even apparently free-floating planets (Sumi *et al.* 2011). The field is now anticipating a wealth of new discoveries thanks to public data from surveys planned in the near future. In particular, the *WFIRST* mission will image the Galactic Bulge continuously in the NIR for ∼72 days/year for 6 six years from the mid-2020s. This will produce a public catalogue of high-quality lightcurves for millions of stars. Among many other discoveries, Penny *et al.* (2016) predict that ∼20,000 microlensing events will be discovered, including ∼3000 that show planetary anomalies.

This will allow us to complete the census of exoplanetary systems begun by ground-based surveys and exemplified by the *Kepler* mission. Figure 1 presents the distribution of the masses and orbital separations of the current sample of confirmed planets, and confirms that the distribution of planets within ∼1 AU of their host stars is now reasonably well understood. However, the most successful survey techniques to date – radial velocities and transits – are both most sensitive to massive planets that lie close to their host stars. The also require observations over the course of multiple planet orbits in order to confirm them, making it very time consuming to verify planets further from their host stars. The direct imaging technique complements this discovery space somewhat, being most sensitive to massive planets sufficiently far (several AU) from their host star as not to be lost in the glare, but the number of accessible nearby targets is limited. The mass-separation diagram remains sparsely populated in the 1–10 AU region, sometimes referred to as the 'snowline' as it is the region in protoplanetary disks which is cool enough for ices to form. The core accretion model predicts a substantial population of 0.1–10 M_{Earth}-mass planets at these separations (e.g. Ida, Lin & Nagasawa, 2013) so the
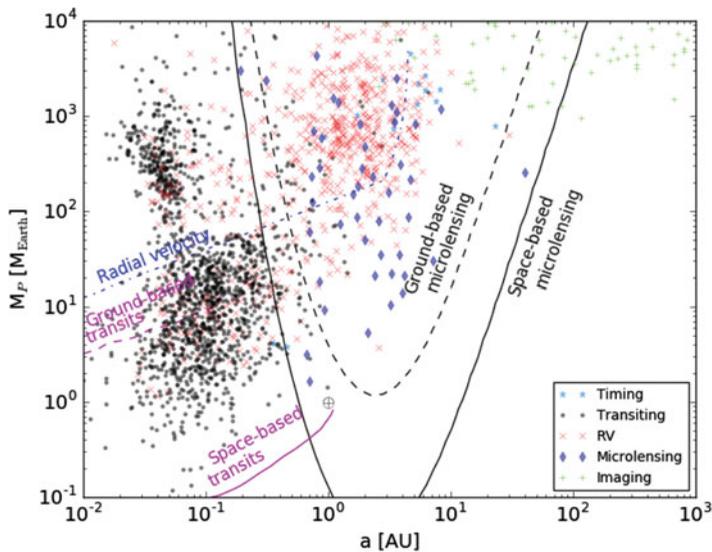
Figure 1. Distribution of the masses and orbital separations of the confirmed exoplanets, superimposed with the sensitivity contours of different detection techniques.

observed distribution will place important constraints on planet formation models. Fortunately, the microlensing technique is sensitive to planets of all mass within this gap, and ground-based observing programs have published more than 40 planets† to date, providing tantalising insights but ultimately constrained by sampling and photometric systematics. *WFIRST*'s space-based survey will provide a much more complete picture of this population.

However, the large number of lightcurves in the *WFIRST* catalogue presents some data analysis challenges. With thousands of events to analyze, now is a good time to review the limitations of existing software and to explore the application of modern statistical and data mining techniques. As the microlensing community is currently relatively small, the astroinformatics is encouraged to collaborate to answer these challenges. The two next sections provide a briefly introduction to the technique of microlensing and some of the challenges of modeling these events. Section 4 then describes the present state-of-the-art in modeling software while Section 5 highlights some examples of unsolved problems.

2. Gravitational Microlensing

When a massive object passes through the line of sight of an observer to a background source star, its gravity bends the light from the source, creating two images of the source and resulting in more light reaching the observer (Refsdal 1964). As the proper motions of all three objects carry them into and out of alignment, the observer sees a distinctive gradual brightening and fading of the source. This can reveal the presence of planets around the middle (lensing) object if the planet happens to lie close to one of the lensed images of the source. In this case, a deviation from the smooth lensing curve is seen, called an anomaly. Anomalous lightcurves can assume a wide range of morphologies, as the source trajectory can intersect the sometimes complex magnification structures or caustics at any angle; one possible example is shown in Figure 2. Microlensing is routinely used to discover single and binary stellar objects in the Galactic Bulge thanks

† http://exoplanetarchive.ipac.caltech.edu/, http://exoplanet.eu
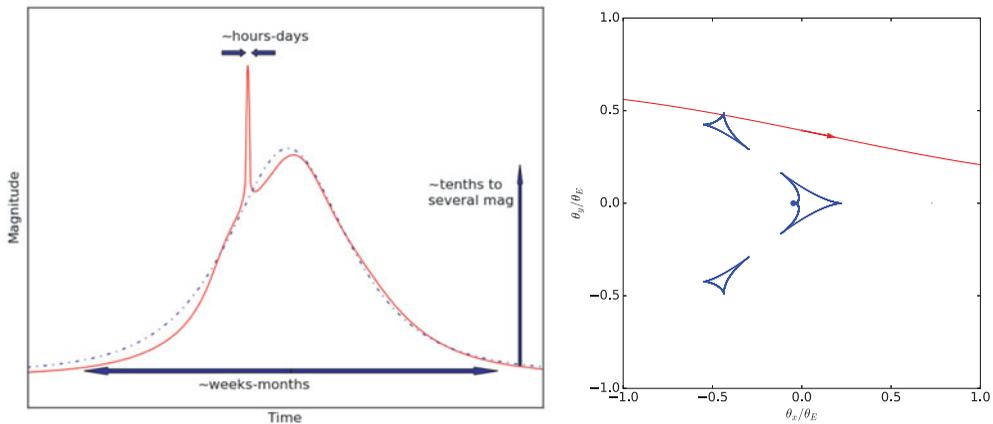
Figure 2. (Left) Schematic lightcurve of a microlensing event, indicating timescales and magnifications typical of lensing events by planet systems in the galactic disk. The dotted blue curve shows the expected lightcurve if the lens were a single star while the solid red curve shows lensing by a planetary system. (Right) Map of magnification in the plane of a binary lens. The lens masses are represented as blue dots and the abrupt changes in lensing magnification they cause are mapped as caustic structures in blue. The motion of the source star relative to the caustic structure during the event is show as an arrowed red line.

to well-established ground-based surveys, and over 40 planets have been published to date. Lensing events can last between 1–hundreds of days, but anomalies are most likely to be discovered around the peak of the event, which means targets require intensive, continuous monitoring during that (multi-day) period. For this reason, ground-based datasets consist of lightcurve data from multiple observatories. Notably, the amplitude of a microlensing event depends on the projected minimum separation of the lens and source rather than the mass of the lens, so the technique is sensitive objects of al masses. For a comprehensive introduction, see Gaudi (2010).

3. Modeling binary lensing events

A microlensing dataset typically consists of a set of irregularly-sampled photometric timeseries, up to years in length, often from multiple instruments. Binary lensing models have at least 7 and sometimes as many 11 parameters, including: t_0, time time of closest projected approach, u_0, the minimum impact parameter, t_E, the time to cross the lens' angular Einstein radius θ_E, α_0, the angular separation of the source trajectory from the binary axis, q, the ratio of binary masses, s_0, the projected separation of the binary ρ, the angular size of source, $\pi_{E,E}$, $\pi_{E,N}$ components of annual parallax, $d\alpha/dt$, the rate of change of α and ds/dt, the rate of change of s. As a result, one of the main challenges in modeling microlensing events is the size of the parameter space that must be searched for the best-fitting model lightcurve.

3.1. *The challenges of real data*

Real data introduces some additional complexity. The rate of occurance, Γ, of microlensing events scales strongly as the product of the surface density of source stars and that of Red Clump stars in the field (Poleski 2016), reaching a maximum in the Galactic Bulge where $\Gamma_{max} = [18.74 \pm 0.91] \times 10^{-6} \exp[(0.53 \pm 0.05)(3 - |b|)]$ star^{-1} yr^{-1}, where b is galactic latitude (Sumi & Penny 2016). Thus the region of highest lensing rate is also one of the most densely-populated regions of the sky. Every pixel in an image has to be assumed

to contain the flux of multiple stars as well as the source, but the pixel scale (and hence blended flux) varies from one telescope (k) to another. The flux from the source, f_s, can only be distinguished from the background stars (f_b) via its magnification, $A(t)$, it undergoes during the event, $f(t, k) = A(t)f_s(k) + f_b(k)$.

Ideally, one would like to sample microlensing lightcurves at a constant regular cadence, but even the best surveys don't produce truly continuous coverage. For microlensing events, data gaps often lead to features being poorly constrained or missed entirely which translates to lower planetary sensitivity. To combat this issue, ground-based observing programs often use multiple telescopes at different longitudes to monitor high priority events at high cadence for extended ($\sim$days-weeks) periods but this then raises the separate issue of combining data from multiple different telescopes while ensuring the photometric errors are handled consistently.

3.2. *Model degeneracies*

While microlensing produces a wide variety of lightcurve morphologies there are also several known degeneracies, where the same (or very similar) lightcurves may be produced from quite different lens/source configurations or trajectories. This is too large a topic to cover in full here, so the reader is referred to the references below for more details. The conceptually simplest example is the physical scale of the lens' Einstein ring, which is determined by its mass (M_L) but which subtends an angular diameter, θ_E, to the observer which also depends on the distances to the lens (D_L) and source (D_S): $\theta_E = \sqrt{\frac{4GM_L}{\mu D_L c^2}}$, where $\mu = 1 + D_L/D_S$. The time scale of an event is determined by the time taken for the source to cross this ring, which can therefore be longer or shorter if the lens is closer or further away, but also depends on the relative velocities of the lens and source. More complex degeneracies occur for binary lenses, where the position of the lens' companion object may intersect either the major or the minor images, which remain unresolved. For a full discussion, see Gaudi & Gould (1997).

Fortunately, there are second-order effects occur which sometimes break these degeneracies and which can be used to determine the physical parameters of the lensing system. The most common is annual parallax, caused either by the motion of the Earth in its orbit during the event, or between simultaneous observations from widely-separated observing platforms. This must be carefully disentangled from the signature of orbital motion of the binary lens (or that of a binary source) during the event. Although microlensing source stars are often at distances of several kiloparsec from Earth, the finite radius of the source stars nevertheless produces a measureable distortion in the lightcurve, resulting in a "rounding over" of the features (a detailed discussion is given in Yoo *et al.* 2004). It is also possible to measure the parallax to an event by observing simultaneously from widely separated platforms on Earth (terrestrial parallax, e.g. Gould *et al.* 2009) or space e.g. Dong *et al.* 2007, Udalski *et al.* 2015, Street *et al.* 2016).

3.3. *Low amplitude signals and red noise*

The magnification, A, of a microlensing event is inversely related to the projected angular separation of the source and lens, u, $A = \frac{u^2+2}{u\sqrt{u^2+4}}$, meaning that the source brightness can increase by several magnitudes in events with low impact parameters. The magnification during a planetary anomaly depends on the source trajectory relative to the magnification caustic and the source size and can range from millimags to magnitudes. Thus microlensing is sensitive to planets of all masses, even at distances of several kiloparsecs. However, the majority of anomalies are of relatively low amplitude and care is needed to

prove that any deviations seen in the lightcurve are physical in origin rather than due to systematic noise, particularly for fainter sources (see Bachelet *et al.* 2015).

4. Current modeling tools

While a few software packages for microlensing modeling currently exist, all are proprietary, with no comprehensive documentation and little testing or verification has been conducted. Combined with limited publicly available data, this has presented a high barrier to entry for newcomers to the field. In anticipation of the public *WFIRST* catalogue, it is appropriate to make available the tools for analysis.

However, the analysis of events has historically been very time consuming owing to the large parameter space that must be searched. Typically a grid search is conducted over physically-reasonable ranges of the parameters s, q and α, with a Markov-Chain, Monte-Carlo algorithm used to find the parameters of the best fitting solution at each grid point. Cluster computing facilities have often been required for this grid search to be conducted within a practical timeframe.

4.1. *PyLIMA*

The need to make the analysis of large numbers *WFIRST* of events both practically feasible and widely accessible has motivated the development of a new open-source package for microlensing modeling. E. Bachelet is leading the development of *pyLIMA*†, a Python-based package which incorporates the fast binary modeling code *VBBinaryLensing*‡ by Bozza (2010). The package adheres to modern coding standards, employing unit testing to ensure reliability as well as regular code reviews by in-house software engineers. It also includes lightcurve simulation tools used to conduct systematic testing of the modeling software (Bachelet *et al.* 2016). Designed from the outset to be user-friendly, the package also includes extensive documentation and Python tutorial workbooks. A second package is also under development by R. Poleski and J. Yee.

5. Outstanding Challenges in Microlensing Modeling

There are a number of areas of microlensing modeling which will benefit from further work. In particular, very little modeling of triple or higher-order multiple lenses has been done to date, including explorations of their particular degeneracies. As the *Kepler* discoveries have demonstrated, multi-planet systems are common and may be identifed in the *WFIRST* catalogue. However it can be difficult to distinguish binary and triple-lenses lightcurves, as demonstrated by the event OGLE-2013-BLG-0723. Udalski *et al.* (2015) reported the discovery of a triple lens: a Venus-mass planet orbiting a brown dwarf, orbiting an M-dwarf star but further analysis by Han *et al.* (2016) identified a binary-lens model which better fit the data. More efficient techniques to thoroughly explore the large parameter space are needed, including for circumbinary systems.

Compounding this issue, there are no widely-recognized criteria to determine when higher-order parameters such as those describing lens orbital motion, need to be introduced. Similarly, while a $\Delta\chi^2$ test is typically applied to determine the "best" model, no rigorous, objective criteria have been applied to determine the minimum $\Delta\chi^2$ required to distinguish between competing solutions, or to test whether an anomalous feature in a lightcurve is genuine or the product of systematic noise.

† https://github.com/ebachelet/pyLIMA
‡ http://www.fisica.unisa.it/gravitationAstrophysics/VBBinaryLensing.htm

5.1. *Data Challenge*

To stimulate work in all of these areas, a microlensing data challenge is planned. This will consist of a publicly-released set of simulated lightcurves, designed to mimic the photometric data products of both ground-based and space-based surveys. Single and binary lens events will be injected into this dataset, and participants will be evaluated on their software's accuracy as well as its speed and hardware requirements.

6. Summary

Microlensing offers an effective way to bridge the gap in the known planet distribution at separations between 1–10 AU. The *WFIRST* mission will deliver an extensive, public catalogue of lightcurves from which we expect $\sim$3000 planets to be discovered in addition to binary (and higher-order multiple) lens systems, but the software currently available is inadequate to perform the analysis of these data. The community is encouraged to join new software development efforts aimed at making the modeling of microlensing events both more accessible and rigorous.

7. Acknowledgments

References

Bachelet, E. 2015, *ApJ*, 812, 136.
Bachelet, E., 2016, *in prep.*
Bozza, V. 2010, *MNRAS*, 408, 2188.
Dong, S. *et al.* 2007, *ApJ*, 664, 862.
Gaudi, B. S. 2010, *Microlensing by Exoplanets in Exoplanets, ed. S. Seager*, p. 79.
Gaudi, B. S. & Gould, A. 1997, *ApJ*, 486, 85.
Gould, A. *et al.* 2009, *ApJ*, 698, 147.
Han, C. *et al.* 2016, *ApJ*, 825, id.8.
Ida, S., Lin, D. N. C., & Nagasawa, M. 2013, *ApJ*, 775, 42.
Penny, M. *et al.* 2016, *in prep.*
Poleski, R. 2016, *MNRAS*, 455, 3656.
Refsdal, S. 1964, *MNRAS*, 128, 29.
Street, R. A. *et al.* 2016, *ApJ*, 819, 93.
Sumi, T. *et al.* 2011, *Nature*, 473, 349.
Sumi, T. & Penny, M. 2016, *ApJ*, 827, id.139.
Udalski, A. *et al.* 2015, *ApJ*, 799, id.237.
Udalski, A. *et al.* 2015, *ApJ*, 812, id.47.
Yoo, J. *et al.* 2004, *ApJ*, 603, 139.