A Reanalysis of Public Galactic Bulge Gravitational Microlensing Events from OGLE-III and -IV

Nathan Golovich[1] , William Dawson[1] , Fran Bartolić[2,3] , Casey Y. Lam[4] , Jessica R. Lu[4] , Michael S. Medford[4,5] ,
Michael D. Schneider[1] , George Chapline[1], Edward F. Schlafly[1] , Alex Drlica-Wagner[6,7] , and Kerianne Pruett[1]

[1] Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA; golovich1@llnl.gov
[2] Centre for Exoplanet Science, School of Physics and Astronomy, University of St. Andrews, St. Andrews KY16 9SS, UK
[3] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 6th Floor, New York, NY 10010, USA
[4] University of California, Berkeley, Department of Astronomy, Berkeley, CA 94720, USA
[5] Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
[6] Wilson Fellow, Fermi National Accelerator Laboratory, Kirk Road and Pine St, Batavia, IL 60510, USA
[7] Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
Received 2020 June 22; revised 2021 November 30; accepted 2021 December 15; published 2022 April 21

Abstract

Modern surveys of gravitational microlensing events have progressed to detecting thousands per year, and surveys are capable of probing Galactic structure, stellar evolution, lens populations, black hole physics, and the nature of dark matter. One of the key avenues for doing this is the microlensing Einstein radius crossing time (t_E) distribution. However, systematics in individual light curves as well as oversimplistic modeling can lead to biased results. To address this, we developed a model to simultaneously handle the microlensing parallax due to Earth's motion, systematic instrumental effects, and unlensed stellar variability with a Gaussian process model. We used light curves for nearly 10,000 OGLE-III and -IV Milky Way bulge microlensing events and fit each with our model. We also developed a forward model approach to infer the t_E distribution by forward modeling from the data rather than using point estimates from individual events. We find that modeling the variability in the baseline removes a source of significant bias in individual events, and the previous analyses overestimated the number of $t_E > 100$ day events due to their oversimplistic model ignoring parallax effects. We use our fits to identify the hundreds filling a regime in the microlensing parameter space that are 50% pure of black holes. Finally, we have released the largest-ever catalog of Markov Chain Monte Carlo parameter estimates for microlensing events.

Unified Astronomy Thesaurus concepts: Gravitational microlensing (672); Astrophysical black holes (98); Time domain astronomy (2109); Astronomy data modeling (1859)

1. Introduction

Gravitational microlensing was first proposed as a means to study the dark halo of the Milky Way by Paczynski (1986), and has since yielded direct constraints on the population of dark objects (Alcock et al. 1996; Renault et al. 1997; Alcock et al. 1998, 2001; Afonso et al. 2003; Tisserand et al. 2007). These were achieved through time-domain photometric surveys of the densest stellar regions of the night sky (the Milky Way bulge and Magellanic Clouds) with optical telescopes (Udalski et al. 1992; Alcock et al. 1993; Arnaud et al. 1994; Muraki et al. 1999). The field of microlensing has pushed dense field photometry, difference imaging photometry, transient alerts, and many more technologies that will be used by upcoming surveys (e.g., the Legacy Survey of Space and Time and the Wide Field Infrared Space Telescope, WFIRST) for a variety of science goals. Though the heritage of microlensing was in the identification of dark objects to understand the nature of dark matter, most practitioners are on the hunt for exoplanets today (see Gaudi 2012; Penny et al. 2019, for a review and a projection, respectively).

Setting binary systems and exoplanets aside, modeling even a simple point-source, point-lens (PSPL) microlensing event is challenging. First, given that most surveys point toward extremely dense stellar regions, it is virtually impossible to measure the flux of the source and lens alone for most events. Given typical seeing from the ground, the baseline flux of most microlensing events are likely composed of several stars' light (Sajadian & Poleski 2019), and given the frequency at which stars exhibit intrinsic stellar variability, the baseline is likely to be variable itself. On top of this, since the Galactic bulge is near the ecliptic, it is near the Sun periodically. The seasonal gaps in the light curve lead to aliasing, and the variations of atmospheric seeing, weather patterns, and other systematics must be considered, especially for long-duration events that span many of these cycles and seasonal gaps. Additionally, the motion of the Earth around the Sun complicates the modeling. The Earth's orbital parallax effect (caused by variable magnification due to Earth's motion around the Sun) is also on a yearly timescale, so care must be taken to not allow this physical signal to mix with the systematics.

Most Galactic bulge microlensing events are short (measurable magnification for at most a couple months) such that the Earth's motion can be either outright ignored or assumed as a constant acceleration. A number of degeneracies arise with the latter assumption (see, e.g., Gould 2004). However, microlensing parallax measurements are useful given they help constrain the mass-distance degeneracy (Wyrzykowski et al. 2016). Microlensing parallax is also useful for exoplanet studies allowing numerous degeneracies in binary lens caustics to be resolved (Yee et al. 2015). Given the usefulness, microlensing parallax is frequently modeled; however, given

the potential for correlations with troublesome systematics, there has been no systematic study of microlensing parallax across entire surveys.

As microlensing surveys began detecting upwards of thousands of microlensing events, distributions of microlensing parameters were studied, especially the global t_E (Einstein crossing time) distributions. The t_E distribution has become an important proxy for physical inference on the Galactic structure. These studies probed several details regarding the population of lenses—including at the short t_E end, the existence of free-floating planets (Sumi et al. 2011; Mróz et al. 2017). The global shape of the distribution has been studied to compare observations (Wyrzykowski et al. 2015; Mróz et al. 2019) to simulations that probe the present-day mass function of lenses (Mao & Paczynski 1996). The shape of the distribution has also been used to probe the matter profile in the inner Galaxy (Mao & Paczynski 1996; Wegg et al. 2016, 2017; Lam et al. 2020). Wyrzykowski et al. (2015) used it to probe the spatial distribution of the optical depth in the Galactic bulge region. Mróz et al. (2019) used it to show the optical depth and microlensing event rate over roughly 40 square degrees with higher spatial resolution and showed that the bulge event rates compared well with the expected microlensing event rate predicted by a Besançon stellar model (see, e.g., Czekaj et al. 2014). In general, however, the t_E distribution is dependent on the line-of-sight distribution of the source and lens parallax and relative proper motion (Mróz et al. 2019).

The t_E distribution is also a rough proxy of the mass distribution of lenses (see, e.g., Mao & Paczynski 1996), and can be used to constrain the fraction of intermediate mass ($\sim$10–1000$M_\odot$) primordial black hole dark matter (Gould 1992; Lu et al. 2019). Lu et al. (2019) showed that OGLE could begin to probe this with the t_E distribution; however, Lam et al. (2020) showed in a more thorough analysis of PopSyCLE simulations that utilizing parallax and astrometric measurements would greatly increase the purity of black hole populations from microlensing. We will explore this question in more detail in two follow-up studies (W. Dawson et al. 2022, in preparation; K. Pruett et al. 2022, in preparation).

Black holes in microlensing surveys need not be primordial in nature. The Milky Way has 10^8–10^9 black holes of stellar origin (Agol & Kamionkowski 2002). The exact number is uncertain, and methods to detect them, other than microlensing, are limited, with most incapable of detecting isolated black holes (i.e., those not in binaries; Remillard & McClintock 2006). Given the clear existence of black holes in the $\sim$5–30 $M_\odot$ mass range from binary detections, it is reasonable to expect that isolated black holes have likely been detected already in microlensing surveys. The typical timescale of these events would be $\sim$100 days,[8] and events with this duration have been observed frequently. The problem is that estimating the mass from the photometric signal that microlensing surveys probe is challenging. Assuming the same source distance, the mass–distance degeneracy ensures that a high-mass and far-away lens can provide a similar observed lensing effect to a low-mass and nearby lens. This degeneracy is broken with the parallax effect due to the Earth's orbit (which causes

asymmetric magnification as well as *wiggles* in the light curve) and with astrometric observations of the lensing behavior (which causes unresolved multiple images of the source and appears as a deviation from the linear proper motion of the flux centroid). Parallax is notoriously challenging to model, and the systematic observation of all events with astrometry is not possible given the time-consuming observations required. There is also a degeneracy between massive, fast-moving lenses and low-mass, slow-moving lenses that is difficult to disambiguate without high-resolution follow-up over years.

If the route of carefully modeling the Earth's orbital parallax is taken, then additional challenges will arise. Black holes are more massive than typical stars, so the relative parallax between the source and lens is more likely to be small. A small parallax manifests a less obvious signal—rather than the obvious periodic wiggles in the photometric light curve of a long-duration microlensing event, the rise and fall of the microlensing signal is very slightly asymmetric. When we consider that microlensing signal in seeing-limited surveys is typically highly blended with unlensed flux due to numerous neighbor stars and consider the frequency that stars exhibit in measurable photometric variability, there is often a time-varying signal due to baseline variability in either the source or blended flux. This signal makes constraining the parallax challenging when it is intrinsically small ($\pi_E < 0.1$; see Section 3 for definition of π_E). The timescale of the microlensing event (the Einstein crossing time, t_E) is also useful to analyze in the context of black holes, as it contains information about the mass of the lens and distances to and relative proper motion of the lens and source. Individual events with long timescales have been scrutinized as black hole candidates (Agol et al. 2002; Bennett et al. 2002; Mao et al. 2002; Poindexter et al. 2005). In principle long-duration and low-parallax microlensing events are most likely to be black holes (Lam et al. 2020); however, given the systematics described above, picking out and constraining the parallax to a small value with the signal correlated over several observing seasons is a challenge.

Gaussian process (GP) models are often used for problems like these. This additional time-domain-correlated noise has been fit with GP models across time-domain astronomy including the transiting exoplanet observations with Kepler (see, e.g., Brewer & Stello 2009; Foreman-Mackey et al. 2017). This method has only very recently been used in microlensing studies of an individual event (Li et al. 2019), and has not been widely deployed on an entire survey. This is unsurprising given the computational requirements to invert large matrices repeatedly in order to infer the hyperparameters of the GP kernel.

A recent theoretical study of OGLE-like microlensing survey simulations using a powerful new tool suggests that many microlensing events are indeed black holes (PopSyCLE; Lam et al. 2020); however, given the aforementioned challenges, it seems a population approach is needed. We undertook the considerable task of analyzing all public Galactic bulge OGLE-III and -IV microlensing light curves and estimating posteriors for the microlensing model parameters (including parallax) as well as inferring GP hyperparameters in order to subtract away the time-domain-correlated noise that afflicts proper inference on the microlensing parameters. In Section 2, we describe the input data for our individual light-curve modeling. In Section 3, we describe the model and sampling algorithm to estimate the

---

[8] See below for a full discussion of the microlensing math, but assuming 4 kpc to the lens, 8 kpc to the source, and 200 km s^{-1} transverse velocity of the lens to the line of sight, 5−30 $M_\odot$ gives lensing event durations of $\sim$78–191 days.

posterior probability density function (PDF) for all of the model parameters. In Section 4, we describe a powerful method of combining all of the information from our parameter posterior samples and demonstrate a useful analysis on the Einstein crossing time global distribution. In Section 5, we present our individual light-curve results as well as our results from the global Einstein crossing time analysis before we discuss and conclude our study in Section 6. We are releasing our Markov Chain Monte Carlo (MCMC) parameter estimate chains for all 9350 light curves analyzed at https://gdo-microlensing.llnl.gov/.

Throughout this paper, all timescale parameters have the units of days (measured in modified Julian days), and flux quantities assume $m_I = 22$ for a zero-point in the OGLE published magnitude system. We will discuss throughout three different microlensing models. The math for each of these will be developed in Sections 2 and 3. First is the standard five-parameter Paczynski (1996) model, which assumes a point-source, point-lens (PSPL) microlensing with no parallax and a constant baseline with blended, unlensed flux. We will refer to this model as PSPL throughout. Second is the parallax model, still assuming PSPL but with the added physics of Earth's motion. We will refer to this as PSPL+parallax model. Third is the most complicated model, which includes both parallax and a GP model for the variability in the baseline. We will refer to this as the PSPL+parallax+GP model. Finally, we will discuss a PSPL+GP model, but only for plotting purposes in Section 5.1.

2. Data

2.1. OGLE-III

The OGLE-III survey spanned the years of 2002–2009 and was operated from Las Campanas Observatory in Chile on the 1.3 m Warsaw University Telescope. The observations were collected with eight (4×2) 2048×4096 CCD chips with $0''.26$ pixel^{-1} with a total field of view of $35' \times 35'$. For more details, we refer readers to the survey paper Udalski (2003). OGLE-III published alerts online as part of the OGLE Early Warning System[9] for each of their eight seasons with 4000 events on their Early Warning System. OGLE-III began the era of statistical inference on populations of thousands of microlensing events. Wyrzykowski et al. (2015) published 3718 unique microlensing events, which at the time was the largest-ever catalog of identified microlensing events. In their analysis, they fit each event with a Paczynski (1996) model for a PSPL;

$$A = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} \quad (1)$$

is the amplification of the baseline source flux, and

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

is the projected angular separation between the lens and source in units of the Einstein angle θ_E:

$$\theta_E = \sqrt{\frac{4GM}{c^2} \frac{D_S - D_L}{D_L D_S}}, \quad (3)$$

which is dependent on the lens mass (M) and distances to the lens (D_L) and source (D_S); I_0 is the baseline magnitude of the light curve in the OGLE I-band. It is convenient to define

$$\tau = \frac{t - t_0}{t_E}, \quad (4)$$

where t_0 and u_0 are defined as $u(t_0) \equiv u_0$ in the heliocentric frame, and t_E is the Einstein crossing time, or the time it takes the source to traverse the lens's angular Einstein radius:

$$t_E = \frac{\theta_E}{\mu_{rel}}, \quad (5)$$

where μ_{rel} is the projected lens–source relative proper motion. The modeled parameters are then

$$\boldsymbol{\theta}_{PSPL} = \{t_0, t_E, u_0, I_0, f_S\}. \quad (6)$$

Wyrzykowski et al. (2015) first fit this model to all light curves using a χ^2 minimization code in order to use these values as features in a random-forest classifier for event detection along with many other light-curve characteristics. There was also an upper limit of $t_E = 400$ days applied to all the OGLE-III data. The full OGLE-III catalog was then reduced to 3560 events, which were then modeled with a MCMC analysis, which estimates the posterior distributions for the parameters described above.

In this paper, we will work with the 3560 I-band light curves as published in Wyrzykowski et al. (2015), perform new fits, and we will draw comparisons to their analysis of the distribution of fit parameters of their events.

2.2. OGLE-IV

The OGLE-IV survey commenced in 2010 and continues today with a 32 chip (2048×4096 pixel) mosaic camera with $0''.26$ pixel^{-1} with a total field of view of 1.4 square degrees. The upgraded camera also had a much shorter readout time allowing for an order of magnitude increase in survey efficiency over OGLE-III. The most recent full-scale analysis of OGLE-IV was presented in Mróz et al. (2019), which analyzes eight years of OGLE-IV data and amasses a catalog of 8000 microlensing events toward the Galactic bulge fields of the OGLE-IV survey. This increase in statistical power allowed for higher-resolution modeling of the Galactic structure toward the bulge and higher-fidelity population statistics of the microlensing parameters. Similar to the OGLE-III analysis (Wyrzykowski et al. 2015), this study published the individual light curves for microlensing events they detected, as well as Paczynski (1996) curve fits to each event. Similarly to Wyrzykowski et al. (2015), they also imposed an upper limit on the timescale ($t_E < 300$ days).

We will use the I-band light curves as published in our analysis to follow. We note here that this is not a comprehensive list of OGLE-IV events. It is missing the central-most Galactic fields, which were published separately in Mróz et al. (2017) as a stand-alone analysis of the high-cadence light curves in search for free-floating planets. These 2617 light curves were not made public, and so they are left out of our study. Additionally we did not include 630 OGLE-IV events in the Galactic plane recently made public and analyzed in Mróz et al. (2020).

---

[9] http://ogle.astrouw.edu.pl/ogle4/ews/ews.html

2.3. PopSyCLE Simulation

A third set of microlensing events that we will use for comparisons to guide the analysis are from a recent simulation that serves as a detailed mock OGLE-type survey (Lam et al. 2020). For details on the simulation and recording of microlensing events, we refer readers to Lam et al. (2020), but in short, PopSyCLE (Population Synthesis for Compact object Lensing Events) is an open-source tool for simulating microlensing surveys starting with a stellar model of the Milky Way, evolving stars through stellar evolution models across an initial-final mass relation, stochastically assigning compact object populations to stars that evolve to that stage, and simulating a microlensing survey on the final Galactic catalog. We make use of the mock OGLE-III and -IV simulations presented in Lam et al. (2020) for comparisons to our analysis (v2—see their Appendix A).

3. Event Modeling

In this section, we describe the modeling framework that we will carry out in order to fit all OGLE-III and OGLE-IV events described in Section 2.

3.1. Annual Parallax

In Section 2.1, we described the basic Paczynski (1996) microlensing model used by both Wyrzykowski et al. (2015) and Mróz et al. (2019) to model individual events. This model is an approximation that ignores the motion of the Earth around the Sun and models the measured flux as the sum of an unlensed and lensed flux as well as additional complexities such as binary sources and/or lenses, and finite source effects (see, e.g., Dominik 1998).

In this subsection, we introduce the effect of observing microlensing from the Earth. Long t_E events or good black hole candidates are sensitive to parallax effects. This is coupled with multiple degeneracies in the model space; i.e., multiple combinations of parameters can yield approximately the same light curve to within the modeling precision. These degeneracies arise from numerous situations including blending of lensed and unlensed light (Woźniak & Paczyński 1997; Dominik 2009), the geometry of gravitational lensing generically (the so-called mass–distance degeneracy; see, e.g., Saha 2000), and the fact that Paczynski (1996) parameters are heliocentric in nature but estimated in the geocentric frame (Smith et al. 2003; Gould 2004). The last of these, observing distant lenses and sources from a moving reference frame, especially when the timescale of the microlensing event is long, is particularly challenging.

The annual parallax was first observed in a microlensing event by the MACHO collaboration (Alcock et al. 1995). Our formulation for microlensing parallax is based on a combination of Alcock et al. (1995) and Gould (2013). We first reconciled Equation (2) from Alcock et al. (1995) with Equation (4) from Gould (2013) in order to add the eccentricity of Earth's orbit to the formalism of Gould (2013).[10] We then modeled the parallax as follows.

The observed data of microlensing is a flux time-series or *light curve* that is typically constructed either from point-spread
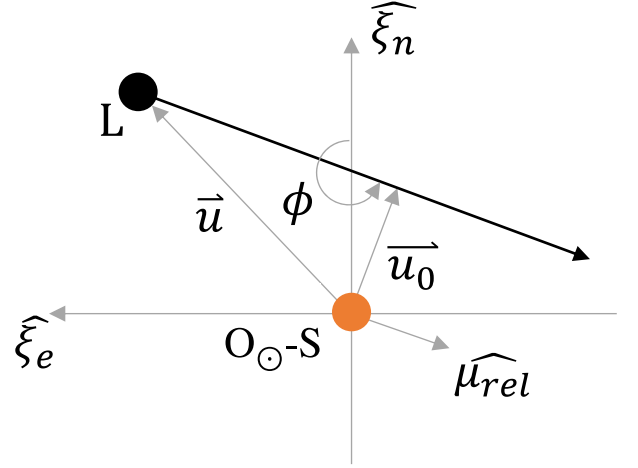
---

[10] In our reconciliation of Alcock et al. (1995) and Gould (2013), we discovered that Equation (2) from Alcock et al. (1995) is missing an additive term $2\tau u_0 \sin(2\phi)$.


Figure 1. The geometry in the plane of the sky normal to the heliocentric observer-source axis. The lens (black circle) passes to the north and west of the source with a relative proper-motion direction characterized by the angle ϕ with a minimum angular separation of u_0. These are each model parameters in our analysis. The unit vectors $\hat{\xi}_n$ and $\hat{\xi}_e$ are aligned with the ecliptic coordinate latitude and longitude, respectively. The unit vector $\hat{\mu}_{rel}$ is in the direction of the relative motion of the lens with respect to the $O_\odot$–S coordinate.

function (PSF) fit photometry or from differential imaging analysis. Most generally, the flux in the light curve is the sum of all sources that lie along the line of sight inside of the extraction profile. However, since the projected angular extent of the detectable lensing effect is much smaller than the typical PSF full width at half maximum, it is difficult to determine which of these sources is the one being lensed. We separate the flux in the light curve into two categories: S for the source, B for the blended unlensed flux (this includes lens flux). Then the flux may be modeled as follows:

$$F(t) = F_B(t) + F_S(t)A(t), \tag{7}$$

where the individual, unlensed fluxes are time-varying due to stellar variability in general, and from General Relativity (Einstein 1936), $A(t)$ is given by Equation (1); u is still the magnitude of the lens–source separation in the plane normal to the observer–source axis in units of θ_E, but it is modified by Earth's motion.

Accounting for Earth's motion, the lens–source separation in geocentric coordinates, $\boldsymbol{u}$, is a vector with ecliptic components given by the following:

$$\boldsymbol{u} = [\tau \sin \phi + u_0 \cos \phi + \xi_e \, \pi_E \sin \Omega] \hat{\xi}_e$$
$$+ [\tau \cos \phi - u_0 \sin \phi + \xi_n \, \pi_E \cos \Omega] \hat{\xi}_n, \tag{8}$$

see Figure 1. For circular orbits,

$$\xi_e = \frac{a}{AU} \equiv \xi; \; \xi_n = \frac{b}{AU} = \xi \sin \beta, \tag{9}$$

where a is the semimajor axis of Earth's orbit, b is the semiminor axis of the circular orbit's projected ellipse on the plane perpendicular to the $O_\odot$–S axis, and β is the ecliptic latitude; see Figure 2. ϕ is the angle between the ecliptic north and the direction of the source-lens relative proper motion in the heliocentric frame. Accounting for the eccentricity, ϵ, of Earth's orbit generalizes this to the following:

$$\xi_e = \xi[1 - \epsilon \cos [\Omega_0(t - t_p)]] \tag{10}$$
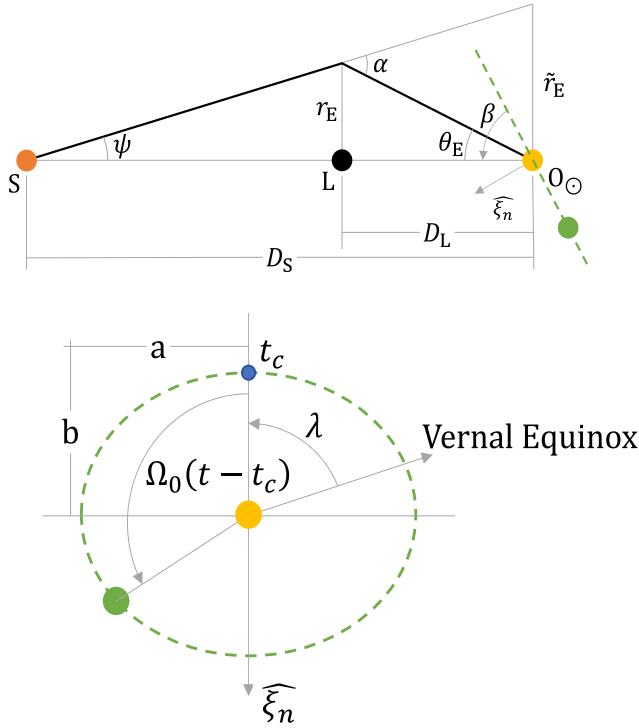
Figure 2. Top: the microlensing formalism of Gould (2000) augmented by microlensing parallax relevant parameters. The bold line represents the path of light from the source (S) to the heliocentric observer ($O_\odot$), which is deflected by the lens (L). D_L and D_S are the assumed Euclidean distances to the lens and source from $O_\odot$, respectively. The deflection angle α is a function of the Einstein radius $r_E (\alpha = 4GM/r_E c^2)$, while θ_E is the Einstein angular radius on the sky. The Einstein radius projected onto the S–O normal plane is defined as $\tilde{r}_E$. In this projection, Earth's orbit (green dashed line) is angled by the ecliptic latitude of the source star β. Bottom: the orbit of Earth (dashed green ellipse) projected onto the S-$O_\odot$ normal plane. Note that the ellipticity and corresponding semimajor and semiminor axes a and b, respectively, are representative of a nearly circular orbit projected on the plane. The frame is oriented such that the time, t_c, at which the Earth is closest to the $O_\odot$–S line is at the top. Within the orbital plane of Earth, λ is the angle of t_c from the vernal equinox. Note that t_c can be determined by the time of vernal equinox, t_{VE} and the ecliptic longitude of S, λ ($t_c = \frac{\lambda}{2\pi} T + t_{VE}$), where T is the Earth's orbital period in days. The angle β in the left figure is simply the ecliptic latitude of S, and ξ_n is in the plane of the page and corresponds to the northern celestial coordinate unit vector. The phase angle of Earth at time t is given by $\lambda + \Omega_0 * (t - t_c)$, where $\Omega_0 = 2\pi/T$ is the conversion factor for the orbital time to radians.

$$\xi_n = \xi \sin\beta [1 - \epsilon \cos[\Omega_0(t - t_p)]], \tag{11}$$

where t_p is the time of the perihelion, and Ω_0 is the angular frequency of Earth's orbit:

$$\Omega_0 = \frac{2\pi}{T}, \tag{12}$$

where T is the period of the Earth's orbit around the Sun;

$$\Omega = \Omega_0(t - t_c) + 2\epsilon \sin[\Omega_0(t - t_p)], \tag{13}$$

where t_c is the time of the Earth's closest approach to the Sun–source axis:

$$t_c = \frac{\lambda}{2\pi}T + t_{VE}, \tag{14}$$

where λ is the equatorial longitude of the source, and t_{VE} is the time of the vernal equinox; see bottom panel of Figure 2.

Finally, the microlensing parallax π_E is

$$\pi_E = \frac{\pi_{rel}}{\theta_E}, \tag{15}$$

where π_{rel} is the relative lens–source parallax.

3.2. Intrinsic Variability of the Baseline

In most microlensing analyses to date, it is assumed that F_B, F_S, and F_L are constant in time thus rendering the baseline ($A(t) \to 1$) flat at large $|t - t_0|$, with the exception of Wyrzykowski et al. (2006), who investigated periodic variability in the baseline of OGLE data, and Li et al. (2019), who modeled the baseline variability of a single event using a GP model. This assumption is never perfect as all stars are variable at some level due to the complex physics of each star, and all light curves are afflicted with seasonal throughput variations due to weather and other systematics; however, it may be a reasonable assumption for the average microlensing event that has $t_E \sim 20$ days. Since we are just as interested in long-timescale events, where stellar variability is more likely to be a significant systematic, we desire a generic model that can fit all OGLE-III and OGLE-IV events.

To model the time-variability of the baseline of the light curve, we use a GP model. For a good summary of GP models and their use in statistical modeling and machine learning, we refer readers to Rasmussen & Williams (2005). In short, a GP is a probability distribution over possible functions. This is an important aspect that allows for the use of Bayes' Theorem and thus integrates the formalism into our preferred probabilistic sampling algorithms. More concretely, for the problem of modeling a microlensing light curve, we use a GP model to handle the variable baseline and correlated noise in the light curve. This is necessary because of the myriad of sources of systematic noise in our observed data (e.g., detector peculiarities, stellar variability, weather, sky conditions, among countless others). The use of a GP may be thought of (and appears in our plots) as fitting the residual of the physical model, but we caution against this viewpoint. In actuality, we are simultaneously fitting both the GP model and the physical model parameters, i.e.,

$$F(t) = \text{physics}(t) + \text{GP}(t) + \text{white noise.} \tag{16}$$

Conditioned on a finite number of observations, a GP is just a multivariate Gaussian distribution with a dense covariance matrix. We must define the μ and Σ—mean and covariance matrix, respectively—for this multivariate Gaussian to set up such a distribution. GP models are simplest if we center the data (i.e., allow $\mu = 0$). In principle we want to allow for relationships between points in our light curve (this is simple to imagine in two dimensions). The covariance matrix is generated by evaluating the kernel function κ, which takes two points t_i and t_j and returns the similarity measure between them as a scalar:

$$\kappa: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}; \ \Sigma_{ij} = \kappa(\tau_{ij}) + K^2 \sigma_i^2 \delta_{ij}, \tag{17}$$

where $\tau_{ij} = |t_i - t_j|$, σ_i is the typical uncertainty in the ith data point, and δ_{ij} is the Kronecker-δ. We have also included a scaling factor, K, on the uncertainty of each measurement, which helps handle underestimated uncertainty in flux measurements. This means that choosing a kernel function strongly influences the range of shapes the function can take.

We refer readers to Duvenaud (2014) for an overview of constructing kernels.

Computationally, adding a GP to our likelihood function means we now must invert and compute the determinant of a covariance matrix. To avoid an unsophisticated $\mathcal{O}(N^3)$ implementation (for N flux measurements in the light curve), we make use of the celerite python package (Foreman-Mackey et al. 2017), which scales as $\mathcal{O}(N)$ through a number of simplifications on the kernel choices that are physically motivated for stellar flux time-series measurements.

The variability in the baseline is complex. It is due to a number of causes including the variations of photometric quality over the course of the year and the asteroseismic oscillations and other internal phenomena that occur at timescales related to the internal structure of the stellar system but are also damped by dissipation. For this reason, we use a damped simple harmonic oscillator (SHO) term for the kernel provided by celerite. Foreman-Mackey et al. (2017) show the underlying kernels that are motivated by the Fourier transform solutions to a stochastically driven, damped SHO (the derivation proceeds through their Equations (19)–(31)). The natural kernels that come from this physical picture are shown to be quasiperiodic and described with exponential kernels, which celerite exploits to vastly speed up the evaluation. The power-spectral density of this kernel is given by the following:

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0\,\omega_0^4}{(\omega^2 - \omega_0^2)^2 + \omega_0^2\,\omega^2/Q^2}, \qquad (18)$$

where ω_0 is the undamped frequency, Q is the quality factor of the oscillator, and S_0 is a constant proportional to the undamped spectral power. We set $Q = 1/\sqrt{2}$ since leaving it unconstrained gave the GP model unlimited freedom to fit the microlensing event itself. This also dramatically improves the speed of hyperparameter estimation and reduces the kernel to the following:

$$\kappa_{\text{SHO}}(\tau_{ij}) = S_0\,\omega_0 \exp\left(-\frac{1}{\sqrt{2}}\,\omega_0\,\tau_{ij}\right)$$
$$\times \cos\left(\frac{\omega_0\,\tau_{ij}}{\sqrt{2}} - \frac{\pi}{4}\right). \qquad (19)$$

This kernel has been used for the background granulation noise in time-series flux measurements of stars (Harvey 1985; Michel et al. 2009; Kallinger et al. 2014; Foreman-Mackey et al. 2017). The source of this noise has been described as a combination of several effects including convection, acoustic oscillations, magnetic activity, and rotation (Cranmer et al. 2014); these are all studied in the rich literature of astroseismology (see Aerts et al. 2010, for a comprehensive review).

In Figure 3, we show a portion of an example light curve from OGLE-III (see Section 2.1). There is clearly a quasiperiodic baseline in the wing of the microlensing event. The SHO kernel term should handle the periodic fluctuations, but we must also handle the general trend of the baseline. This low-order polynomial trend is extremely troublesome for sampling the posterior of the microlensing parameters of interest (e.g., π_E and t_E). Without handling the time-varying baseline, the model is forced into extremely unlikely regions of parameter space in order to explain both the true underlying microlensing event that varies over a small fraction of the whole light curve and the nearly linear trend over the several

years that flux measurements were made. To handle this type of light curve, we add a Matérn-3/2 kernel. Note that we may construct kernels as the sum or product of two valid kernels (Durrande et al. 2011). The Matérn-3/2 kernel is given by the following:

$$\kappa_{\text{M3/2}}(\tau_{ij}) = \sigma^2\left(1 + \frac{\sqrt{3}\,\tau_{ij}}{\rho}\right)\exp\left(-\frac{\sqrt{3}\,\tau_{ij}}{\rho}\right), \qquad (20)$$

where ρ and σ are the modeled hyperparameters.

All together, our data is modeled with a covariance structure according to the following:

$$\Sigma_{ij} = \kappa_{\text{SHO}}(\tau_{ij}) + \kappa_{\text{M3/2}}(\tau_{ij}) + K^2\,\sigma_i^2\,\delta_{ij}, \qquad (21)$$

and the model flux vector is as follows:

$$\boldsymbol{\mu} = \boldsymbol{\Phi}\,\boldsymbol{w}, \qquad (22)$$

where $\boldsymbol{\Phi}$ is as follows:

$$\boldsymbol{\Phi} = \begin{pmatrix} A(t_1) & 1 \\ A(t_2) & 1 \\ \vdots & \vdots \\ A(t_N) & 1 \end{pmatrix} \qquad (23)$$

for N flux measurements, where A is given in Equation (1), and

$$\boldsymbol{w} = \begin{pmatrix} F_S \\ F_B \end{pmatrix}. \qquad (24)$$

It is convenient to define the baseline, unlensed flux:

$$F_{\text{base}} = F_S + F_B \qquad (25)$$

and the source-flux-fraction:

$$b_{\text{sff}} = \frac{F_S}{F_S + F_B}, \qquad (26)$$

which we will sample in our modeling and can be used to select dark lenses (Wyrzykowski et al. 2016; Wyrzykowski & Mandel 2020).

Thus, for observed flux data $\boldsymbol{F}$, the above formulation allows for a Gaussian likelihood function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

$$p(\boldsymbol{F}|\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{F};\,\boldsymbol{\mu},\,\boldsymbol{\Sigma}), \qquad (27)$$

where $\mathcal{N}$ indicates a normal distribution, and $\boldsymbol{\theta}$ is the set of microlensing and GP hyperparameters:

$$\boldsymbol{\theta} = \{F_{\text{base}}, b_{\text{sff}}, t_0, t_E, u_0, \pi_E, \phi, S_0, \omega_0, \rho, \sigma, K\}. \qquad (28)$$

3.3. Priors

In addition to the likelihood of the previous section, we must also define the priors for our model parameters in order to properly sample the posterior PDF. We discuss these priors in this subsection individually; however, the key information is contained in Table 1.

3.3.1. Flux Parameters: F_{base} and b_{sff}

The measured flux of a typical microlensing event is an unknown combination of source, lens, and blended light. It is possible to disambiguate the three using a combination of high-resolution follow-up and parallax measurements, but in principle we do not know flux ratios a priori. Modeling the
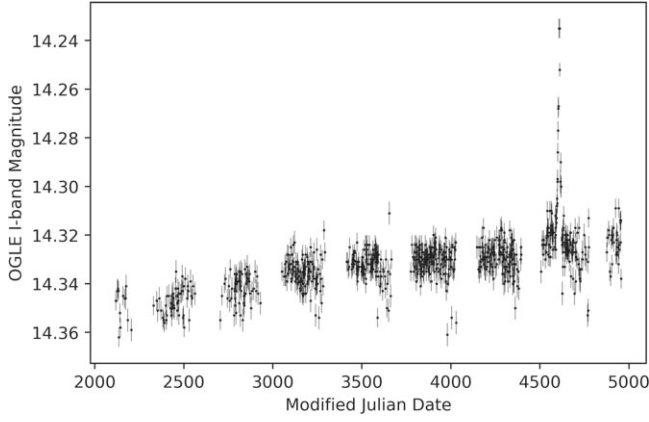
Figure 3. Example OGLE-III light curve (OGLE BLG 156.7.141434) showing the baseline of the light curve. The baseline is clearly rising over the course of several years in addition to a quasiperiodic oscillatory behavior. We have designed our GP kernel to handle these systematics specifically since these are frequent in OGLE data.

Table 1
Summary of Prior Probability Density Functions for Modeled Parameters

Parameter (1)	Distribution (2)	Inputs (3)
F_{base}	$\mathcal{N}$	$\mu = \text{med}(\boldsymbol{F})$, $\sigma = \sigma_{\boldsymbol{F}}$
b_{sff}	$\mathcal{U}$	lower $= 0$, upper $= 1 + \epsilon_{NB}$
t_0	$\mathcal{N}$	$\mu = t_{0,OGLE}$, $\sigma = 100$ days
$\log_{10} t_E$	$\mathcal{N}$	$\mu = 1.13435$, $\sigma = 0.67502$[a], truncated between $0.5 < t_E < 3000$ days
$\log_{10} u_0$	$\mathcal{N}$	$\mu = -1$, $\sigma = 2$, truncated between $10^{-5} < u_0 < 3$
$\log_{10} \pi_E$	$\mathcal{N}$	$\mu = -0.884205$, $\sigma = 0.072755$,[a] truncated between $10^{-5} < \pi_E < 3$
ϕ	$\mathcal{U}$	lower $= 0$, upper $= 2\pi$
$\log(S_0 \omega_0^4)$	$\mathcal{N}$	$\mu = \sigma_{\boldsymbol{F}}^2$, $\sigma = 5$
$\log \omega_0$	$\mathcal{N}$	$\mu = 0$, $\sigma = 5$
$\log \sigma$	$\mathcal{N}$	$\mu = 0$, $\sigma = 5$
ρ	Γ^{-1}	see Section 3.3.8 for details
$\log(K^2)$	$\mathcal{N}$	$\mu = 0.0953$, $\sigma = 1$

Note. Column 1 gives the parameters in our PSPL+parallax+GP model. Column 2 gives the distribution assumed for the prior PDF for each parameter. $\mathcal{N}$, $\mathcal{U}$, and Γ^{-1} indicate a normal, uniform, and inverse-gamma distribution, respectively. Column 3 gives the input values for each of these distributions. For the normal distributions, a mean (μ) and standard deviation (σ) are listed. In some cases, these normal distributions are truncated, with zero probability outside the specified range. In the case of uniform distributions, the lower and upper limits are listed; outside this range, the prior PDF is zero. We refer readers to the corresponding text for details of the inverse-gamma distribution.
[a] Based on PopSyCLE output.

flux parameters is a challenge since these are highly correlated with several other microlensing parameters. We thus use a MCMC to sample in the total baseline flux and the source-flux-fraction with the following priors:

$$p(F_{base}) \propto \mathcal{N}(\text{median}(\boldsymbol{F}), \sigma_{\boldsymbol{F}}), \quad (29)$$

where the median is taken over the flux time-series, and $\sigma_{\boldsymbol{F}}$ refers to the standard deviation of the flux time-series;

$$p(b_{sff}) \propto \mathcal{U}(0, 1 + \epsilon_{NB}), \quad (30)$$

where $\mathcal{U}(a, b)$ describes a uniform distribution between a and b that is zero elsewhere; ϵ_{NB} allows for a small amount of "negative blending" (Park et al. 2004; Smith et al. 2007), where

the source apparently contributes more than 100% of the observed flux. This is, strictly speaking, nonphysical, but it becomes possible in the parameters due to systematics in the modeling of the pixels, e.g., if the difference imaging procedure improperly attributes flux from a neighbor to the source. We restrict the amount of negative blending to the amount of flux from a star of $m_I = 20.5$. This follows the analysis of Mróz et al. (2019), where a comparison of OGLE fields was made to high-resolution Hubble Space Telescope images and pixel-level simulations (see Figures 8 and 9 of Mróz et al. 2019).

3.3.2. Time of Closest Approach: t_0

The time of closest approach is defined in the heliocentric frame; i.e., it is the time when, as viewed from the Sun, the lens and source are nearest in their trajectories on the sky. This is troublesome for modeling light curves with large microlensing parallax signals because then the peak of the light curve is different than t_0; thus we cannot set a tight prior on t_0 to the peak of the light curve. Since a high microlensing parallax is rare, and most light curves have good agreement between t_0 and the peak of the light curve, we use a Gaussian prior, centered on the reported t_0 values from OGLE (Wyrzykowski et al. 2015; Mróz et al. 2019);

$$p(t_0) \propto \mathcal{N}(t_{0,OGLE}, 100 \text{ days}). \quad (31)$$

3.3.3. Einstein Crossing Time: t_E

The time it takes for the source to transit the angular Einstein radius, θ_E, is again defined in the heliocentric reference frame, which is parameterized by t_E. Dominik (2009) showed reparameterizations that are useful for fitting Earth-based microlensing observations including defining the effective-timescale $t_{eff} = t_E u_0$ and modeling in u_0 and t_{eff}. We use t_E and u_0 parameters in our analysis since t_{eff} is most useful for shorter events, and we want to be capable over a large range in t_E. The timescale distribution of microlensing surveys is often reported as a histogram of t_E point estimates from the population of events. For OGLE-III and -IV, this distribution has proven to have a peak roughly at 20 days with tails toward short timescales as short as 1 day and as long as several hundred days. Recently, Lam et al. (2020) published the first simulated OGLE-like microlensing survey that largely confirmed these findings (though the peak of the t_E distribution is slightly lower). We use the output from the Lam et al. (2020) OGLE-like PopSyCLE simulation to develop our t_E prior. However, we inflate the distribution to allow for the possibility of longer timescale events in the OGLE database. We sample in $\log_{10} t_E$, so our t_E prior is given by the following:

$$p(\log_{10} t_E) \propto \mathcal{N}(\mu_{\log_{10} t_E}, 2\sigma_{\log_{10} t_E}), \quad (32)$$

where the mean and standard deviation refer to the values from the PopSyCLE OGLE simulation ($\mu_{\log_{10}} t_E = 1.13435$ and $\sigma_{\log_{10}} t_E = 0.33751$). In addition, we truncate this prior to only allow for $0.5 < t_E < 3000$ days.

3.3.4. Impact Parameter: u_0

The impact parameter is defined as $u_0 \equiv u(t_0)$. It is again defined in the heliocentric frame and often reparameterized as a factor of the effective timescale ($t_{eff} = t_E u_0$). Geometrically, given the random sources and lenses with random motion on

the sky, the distribution of impact parameters for any pair is roughly uniform out to large radii. However, when modeling a population of microlensing events, this is weighted by the detection sensitivity of the survey. Large impact parameter microlensing events are difficult to detect since the peak magnification gets buried in the noise of the flux measurements. We sample in $\log_{10} u_0$ under the prior:

$$p(\log_{10} u_0) \propto \mathcal{N}(-1, 2) \tag{33}$$

truncated between $10^{-5} < u_0 < 3$. Note that we are only modeling the positive u_0 solutions. We discuss this choice in more detail in Section 4. We checked that there is no bias in the population modeling made by this choice.

3.3.5. Microlensing Parallax, π_E

The microlensing parallax is a measure of the effect of the Earth's orbit relative to the size of the projected angular Einstein radius. No population distribution of this parameter has ever been constructed from data, so we again turn to the PopSyCLE simulation results. We sample in $\log_{10} \pi_E$ under the prior:

$$p(\log_{10} \pi_E) \propto \mathcal{N}(\mu_{\log_{10} \pi_E}, \sigma_{\log_{10} \pi_E}) \tag{34}$$

truncated between $10^{-5} < \pi_E < 3$. Here, again, the mean and standard deviation values are computed from the PopSyCLE output ($\mu_{\log_{10} \pi_E} = -0.884205$ and $\sigma_{\log_{10} \pi_E} = 0.072755$).

3.3.6. Relative Proper-motion Angle: ϕ

In most characterizations of microlensing parallax, it is treated as a vector quantity with a magnitude equal to the relative parallax between the lens and source in units of the angular Einstein radius. The direction of the vector is, by convention, aligned with the lens–source relative motion vector and often quoted in ecliptic coordinate components. In our parameterization, we instead model the magnitude of the parallax and, separately, the angle ϕ measured from the ecliptic north to the direction of the projected relative proper motion of the source and lens. This proper motion is in the heliocentric coordinate frame. We sample under a uniform parallax angle prior:

$$p(\phi) \propto \mathcal{U}(0, 2\pi). \tag{35}$$

In practice, sampling in a periodic angle can cause issues since the sampler sees a discontinuity, so we sampled in $\sin \phi$ and $\cos \phi$ following the method developed in the exoplanet python package (Foreman-Mackey et al. 2019).

3.3.7. Simple Harmonic Oscillator Gaussian-process Kernel Priors: S_0 and ω_0

Our model includes a stochastically driven, damped SHO covariance function as part of the covariance matrix for the likelihood function. This GP term has two hyperparameters S_0 and ω_0. These are related to the power-spectral density at the characteristic frequency of the SHO described by the kernel. They are often highly correlated, and since the power-spectral density of this kernel is a function of $S_0 \omega_0^4$, we sample in $\log(S_0 \omega_0^4)$ and $\log \omega_0$ with priors:

$$p(\log(S_0 \omega_0^4)) \propto \mathcal{N}(\sigma_F^2, 5), \tag{36}$$

$$p(\log(\omega_0)) \propto \mathcal{N}(0, 5), \tag{37}$$

where σ_F^2 is the variance of the flux time-series, following the celerite tutorials for time-series data.

3.3.8. Matérn-3/2 Gaussian-process Kernel Priors: σ and ρ

Our model also includes a Matérn-3/2 kernel term, which is parameterized by ρ and σ, which are the characteristic timescale and amplitude of the flux variability, respectively. We allow for a wide range of flux variability:

$$p(\log(\sigma)) \propto \mathcal{N}(0, 4). \tag{38}$$

For the timescale hyperparameter, we use an inverse-gamma prior with 1% of the probability below the typical data spacing and above the length of the time-series:[11]

$$p(\rho) \propto \Gamma^{-1}(a, b). \tag{39}$$

3.3.9. Diagonal Covariance Scale Factor: K

Finally, we model an additional parameter that linearly scales the uncertainty of the flux measurements. This is sampled following the prior:

$$p(\log(K^2)) \propto \mathcal{N}(\log(1.1), 1). \tag{40}$$

We note that more sophisticated methods of scaling the uncertainty estimates have been carried out, especially for high magnification photometric measurements where each data point caries much information (e.g., Yee et al. 2012; Li et al. 2019); however, since our primary interest is in longer timescale events with no caustic-crossings, we implement this simpler method here since it readily is modeled simultaneously with our MCMC sampling.

3.4. Sampling

Our model has 12 parameters and many known degeneracies and correlations between subsets of the model parameters. Thus, in order to sample from the posteriors, we must make use of an efficient sampling algorithm. In the previous section, we listed one by one the prior PDFs we will use. The likelihood function is given in Equation (27). Each likelihood evaluation is computationally expensive, given the GP in our model. We make use of celerite and exoplanet packages to not only speed up this portion of our analysis but also because these packages were built into the framework of PyMC3 (Salvatier et al. 2016), which is a Python package for Bayesian statistical modeling. We make use of the No-U-Turn Sampler (NUTS; Hoffman & Gelman 2011), which is related to the Hamiltonian Monte Carlo (HMC) algorithm. HMC is a MCMC algorithm that uses gradient information to converge with complicated posteriors more quickly than Metropolis or Gibbs sampling (for a review, see Betancourt 2017). NUTS is an improvement on HMC because it computes important tuning parameters as it samples and removes the potential for user error. The NUTS implementation in exoplanet computes the mass matrix internally through a series of tuning steps that enable more efficient sampling of the dense regions of high-dimensional posterior PDFs (Foreman-Mackey et al. 2019).

---

[11] Following https://betanalpha.github.io/assets/case_studies/gp_part1/part1.html.

After we tune the sampler with 4000 tuning steps, we sample all 3560 OGLE-III light curves and 5790 OGLE-IV light curves under the described model for 40,000 samples (20,000 steps of burn-in and 20,000 saved samples). We provide the chains for all 9350 posterior samples for all parameters in an online data set[12] totaling roughly 23 GB.

4. Population Modeling

In the microlensing literature, many authors present distributions of point estimates over each of their fits. This is especially prevalent with the t_E (Sumi et al. 2011; Wyrzy-kowski et al. 2015; Mróz et al. 2017, 2019). Given that microlensing modeling is riddled with degeneracies and correlations, there is reason to believe these distributions are potentially biased given that they were modeled without parallax, and the variability of the baseline was not modeled. In this section, we will discuss the proper Bayesian methods for combining the full posteriors from each light curve.

4.1. A Hierarchical Model

Often, MCMC methods are used to sample the complicated posteriors for binary lens and source systems throughout the recent literature, and there are now open-source modeling packages available to the community, e.g., pyLIMA (Bachelet et al. 2018) and MulensModel (Poleski & Yee 2019).

As Bayesian methods have become more popular, authors have provided a varying amount of this information in the literature. Some authors do not report posteriors and only the maximum likelihood, maximum a posteriori (MAP), or the marginal median values for each parameter along with typical confidence intervals on modeled parameters (often 68%, 90%, or 95% credible intervals). Some others provide graphical representations including corner plots, which show the marginal two-dimensional posteriors and correlations in addition to the marginal one-dimensional distributions. Others still provide all the samples. We have chosen to provide all the samples, so our posteriors can be maximally useful for comparisons and extensions to our work as well as any subsequent analysis.

In this subsection, we will demonstrate a use case with the t_E chains. We follow the method described by Hogg et al. (2010), which develops the probability theory for combining the samples of eccentricity posteriors from the analysis of individual exoplanet orbits from radial velocity data. We refer readers to that paper for the full details. In short the method sets out to forward model the distribution of the parameter of interest itself, rather than estimate the distribution from histograms of single point estimates or summed MCMC samples, which can result in biased estimates.

We have N microlensing events labeled $n (1 \leqslant n \leqslant N)$ with M_n flux measurements at times t_j, labeled $j (1 \leqslant j \leqslant M_n)$), F_{nj}. From these data and the likelihood function given in Equation (27) and the prior PDFs ($p(\boldsymbol{\theta}_n)$) in Section 3.3, we have computed samples from the the individual posterior PDFs:

$$p(\boldsymbol{\theta}_n|F_n) = \frac{p(F_n|\boldsymbol{\theta}_n)p(\boldsymbol{\theta}_n)}{Z_n}, \qquad (41)$$

where Z_n is a normalization constant. Then, assuming all of the light curves are independent (note this is mostly true, but

sources of telescope and detector systematic error make it not fully true), the total likelihood for all of the light-curve parameters is the product of each likelihood function. Instead we want the likelihood for the set of parameters, $\boldsymbol{\alpha}$, that describe the distribution of interest over the individual chains, $f_{\boldsymbol{\alpha}}(t_E)$. We can think of this distribution as a better prior PDF for the parameter of interest. Mathematically this amounts to a change of variables and integrating out the individual light-curve parameters. We follow Hogg et al. (2010):

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\alpha}} &\equiv p(\{F_n\}_{n=1}^N|\boldsymbol{\alpha}) \\ &= \prod_{n=1}^N \int d\boldsymbol{\theta}_n\, p(F_n|\boldsymbol{\theta}_n)p(\boldsymbol{\theta}_n|\boldsymbol{\alpha}),\end{aligned} \qquad (42)$$

where,

$$p(\boldsymbol{\theta}_n|\boldsymbol{\alpha}) = \frac{f_{\boldsymbol{\alpha}}(t_E)p(\boldsymbol{\theta}_n)}{p(t_E)}. \qquad (43)$$

The integrals are taken over $12N$ dimensions, and yield a marginal likelihood or probability of the data given the parameters $\boldsymbol{\alpha}$. This integral is not practical in most cases, but since we have already generated K-element samplings with elements $\boldsymbol{\theta}_{nk}$,

$$\int d\boldsymbol{\theta}_n\, p(\boldsymbol{\theta}_n|F_n)f(\boldsymbol{\theta}_n) \approx \frac{1}{K}\sum_{k=1}^K f(\boldsymbol{\theta}_{nk}), \qquad (44)$$

where the posterior PDF in the integrand represents the posteriors from the individual fits, and $f(\boldsymbol{\theta}_n)$ is an arbitrary function of the parameters. Since all of the integrals in Equation (44) can be approximated by sums over samples of the individual posteriors, the marginal likelihood for $\boldsymbol{\alpha}$ is as follows:

$$\mathcal{L}_{\boldsymbol{\alpha}} \approx \prod_{n=1}^N \frac{1}{K}\sum_{k=1}^K \frac{f_{\boldsymbol{\alpha}}(t_{E_{nk}})}{p(t_{E_{nk}})}. \qquad (45)$$

Here the sum is over the ratio of the new prior; we want to infer over the prior we used for t_E in the individual light-curve model. Importantly, we must make sure that the old prior has support over the full range of interest for the new distribution. Hogg et al. (2010) recommend uninformative priors for the individual fits; however, as long as there is support over the full domain of $f_{\boldsymbol{\alpha}}(t_E)$, this need not be entirely uninformative. In fact, uninformative priors are often informative in ways that the modeler did not intend.[13] We note that our prior is supported in the range, 0.5 days < t_E < 3000 days with a smooth normal distribution in $\log_{10}(t_E)$. With this likelihood function, we can then multiply by a prior PDF for $\boldsymbol{\alpha}$, normalized and sampled to obtain a posterior PDF distribution for $\boldsymbol{\alpha}$ fully forward modeled from the individual light curves.

Now we turn our attention to choosing a functional form for $f_{\boldsymbol{\alpha}}(t_E)$ and a prior PDF for $\boldsymbol{\alpha}$. We make relatively simple choices for each of these since our intention is to demonstrate the power of our publishing of chains for all public Galactic bulge OGLE-III and OGLE-IV microlensing events from Wyrzykowski et al. (2015) and Mróz et al. (2019), respectively. We model the distribution as a step-function following Hogg et al. (2010). This is useful to compare to back to the traditional histograms used in point-estimate distributions presented in

---

[12] https://gdo-microlensing.llnl.gov/

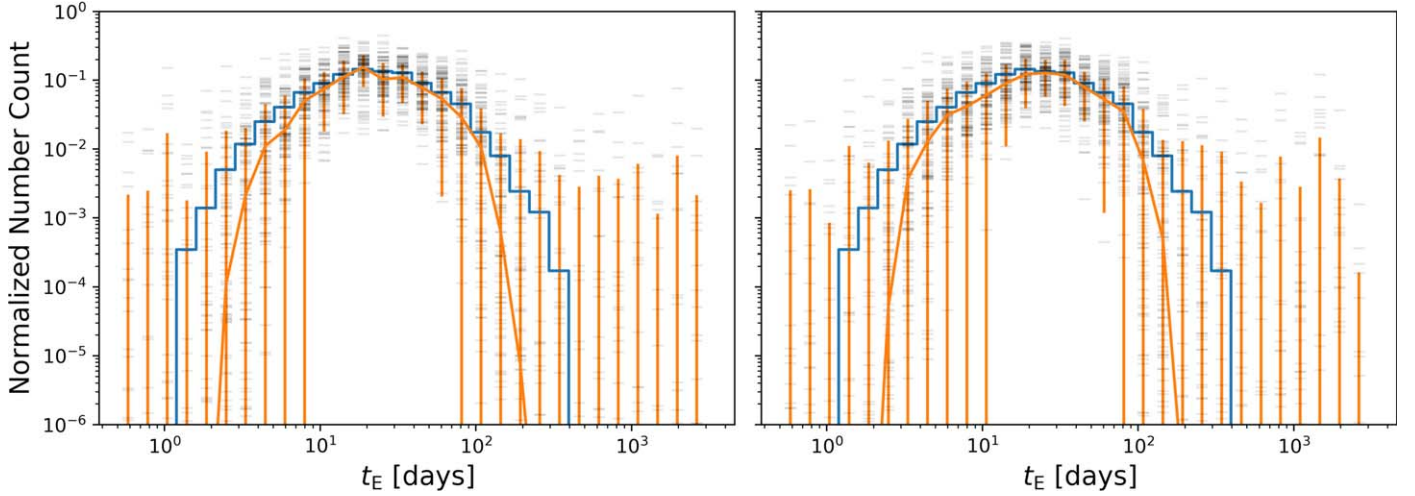[13] See Bailer-Jones (2015) for an illustrative example.

Figure 4. Graphical representation for the prior probability density function for the population modeling of the Einstein crossing time distribution for OGLE-III (left panel) and OGLE-IV (right panel). The blue histogram in each panel shows the normalized bin heights for the MAP t_E values for our individual fits. The orange bars show the 68% confidence interval, about the mean (solid orange curve), for random samples from the Dirichlet-prior distribution (shown in gray ticks). The Dirichlet priors are for $b = 20$ and $c = 0.02$.

Wyrzykowski et al. (2015), Mróz et al. (2019). It is sort of a nonparameteric model (the parameters are the bin heights themselves) where we do not have to assume an overly simplified functional form.

$$f_\alpha(\log_{10}(t_E)) \equiv \sum_{m=1}^{M} \alpha_m \, s\left(t_E; \frac{m-1}{M}, \frac{m}{M}\right), \qquad (46)$$

where

$$s(x; L, H) \equiv \begin{cases} 0, & \text{for } x < L \\ (H - L)^{-1}, & \text{for } L \leqslant x \leqslant H, \\ 0, & \text{for } H < x \end{cases} \qquad (47)$$

and

$$\sum_{m=1}^{M} \alpha_m = 1. \qquad (48)$$

This can be thought of as a normalized histogram with M equal width bins in $\log_{10}(t_E)$. For the prior PDF on α (i.e., the bin amplitudes), we use a Dirichlet prior:

$$p(\alpha) \propto \text{Dir}(\alpha|a). \qquad (49)$$

Here the Dirichlet distribution is parameterized by the Dirichlet concentration vector a, which has the same number of elements M as α. The Dirichlet prior describes a vector-valued multinomial distribution, which has the convenient property in that it naturally satisfies Equation (48). We construct a by conditioning on the weights (w) for a histogram of MAP t_E values for all events, from the respective survey, and compute the Dirichlet concentration parameters as follows:

$$a = (w - \min(w))b + c; \qquad (50)$$

b tunes the allowable bin amplitude variance about the prior, where the larger b, the less variance there will be for each bin amplitude m away from w_m, c influences how uniform the bin height prior is across the M bins, where $c < 1$ will more closely follow w, and $c > 1$ will approach uniform. A full Bayesian analysis would open these up for modeling; however, we tested a large range for both of these parameters and found negligible

difference in the results for the t_E distribution. We set $b = 20$ and $c = 0.02$. The priors for both OGLE-III and OGLE-IV are shown in Figure 4.

4.2. Sampling

For sampling this model, we again make use of PyMC3 and the NUTS algorithm. We sample the model over all t_E chains for OGLE-III and OGLE-IV separately. We sample for 3200 steps, burning the first 1600. We again ensured that the number of effective samples was sufficiently large to avoid biased results from autocorrelation. The results of this analysis are presented in Section 5.2.

5. Results

In Section 3, we described in detail the model and sampling algorithm that we used to generate posterior samples for 12 individual event microlensing and GP hyperparameters simultaneously for 3560 OGLE-III and 5790 OGLE-IV microlensing events. In Section 4, we described the advantage of modeling the population distribution of important microlensing parameters by importance sampling the individual samples generated in Section 3 and estimating the distribution parameters in a Bayesian, hierarchical framework. In this section, we will present the key results from these sections separately.

5.1. Individual Events

Each light curve has been sampled with the likelihood function given in Equation (27) (the PSPL+parallax+GP model) and prior PDFs on 12 parameters given in Section 3.3. We also computed the MAP values for all 12 parameters as well as the representative MAP estimates for the PSPL+parallax +GP model and the PSPL model. The MAP models were computed with a Broyden–Fletcher–Goldfarb–Shanno algorithm (see, e.g., Byrd et al. 1987). The posterior PDF for 12 parameters for all of the publicly available OGLE-III (Wyrzykowski et al. 2015) and OGLE-IV (Mróz et al. 2019) Galactic bulge microlensing events were sampled, and all samples are available for download with this publication. In this subsection, we will present the results for a few representative examples.
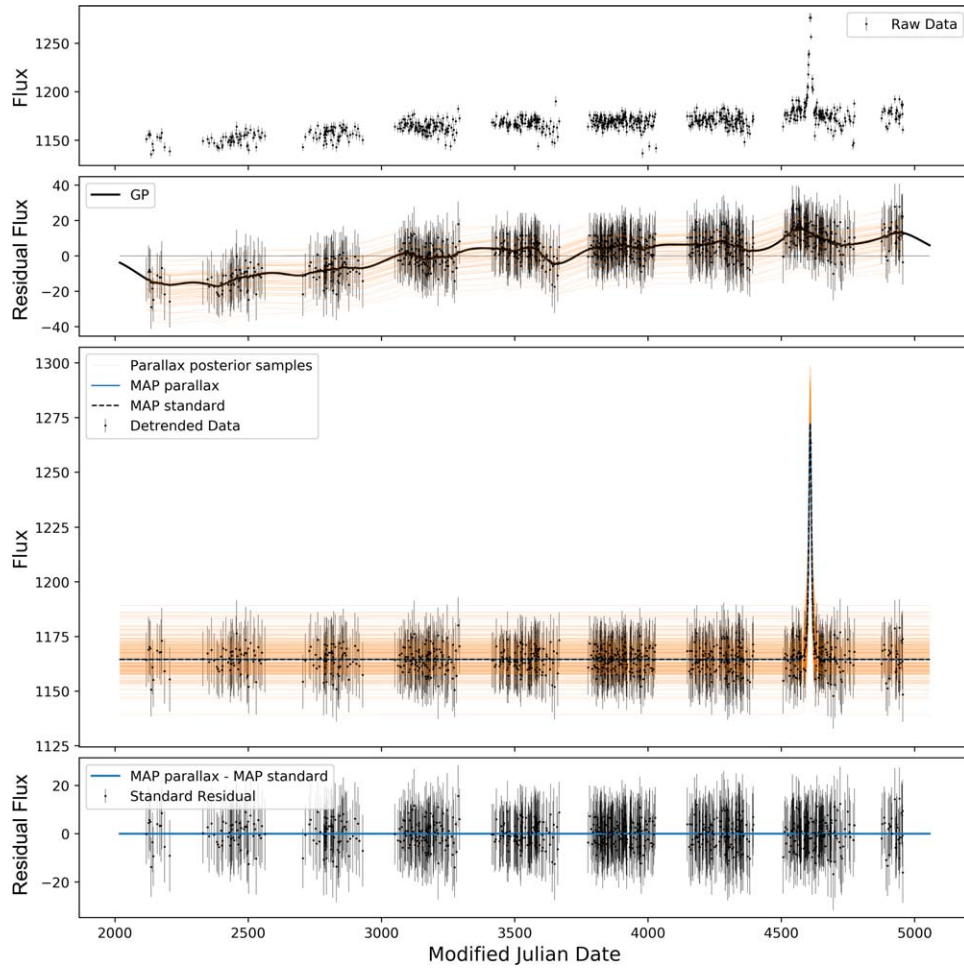
Figure 5. Example OGLE-III light curve (OGLE BLG 156.7.141434) with results from our MCMC analysis. First panel: raw data published by Wyrzykowski et al. (2015); second panel: GP component of the PSPL+parallax+GP model with the MAP in thick black and 50 random draws from our posteriors in orange. The data has the MAP PSPL+parallax subtracted from the raw data in the first panel and rescaled uncertainty on each measurement by the MAP value of K in our model. Third panel: 200 microlensing posterior samples (orange) and MAP parallax model (solid blue) subtracting away the MAP GP from the panel above. The MAP PSPL+GP model is also plotted for comparison. The data is detrended, meaning the MAP GP model realization (black curve in the second panel) is subtracted away. Fourth panel: residual flux after subtracting the PSPL+GP model. The blue curve shows the difference between PSPL+parallax+GP and PSPL+GP models. In this case, the two are similar (PSPL+parallax+GP to PSPL+GP likelihood-ratio = 1.6).

In Figure 3, we presented the type of time-domain-correlated noise that can cause troubles with fitting microlensing events with a model that has a constant baseline (such as the PSPL or PSPL+parallax models). In Figure 5, we show the MAP model (with the GP and PSPL+parallax components separated) along with samples from the posterior in the data-space for this same event. Though technically the microlensing parameters and GP-kernel hyperparameters are fit simultaneously, this plotting technique helps visualize how the model is working. The spread in the baseline flux for the random posterior samples is indicative of the combined uncertainty in the GP model and the baseline flux; however, the microlensing parameters are all well sampled with minimal autocorrelation. The PSPL+GP and PSPL+parallax+GP models are in good agreement for this event.

In Figure 6, we show (analogously to Figure 5) a light curve that has a much flatter baseline. Here the posterior samples are tightly clustered around the MAP in both the microlensing and GP panels, and the sampler has sufficiently mapped out the posterior for all parameters. This is due to the relatively high signal-to-noise ratio (S/N) of the data. Similar to the previous example, there is not a substantial difference between the

PSPL+parallax+GP and PSPL+GP models. This is largely due to the short timescale of these events.

In Figure 7, we show (analogously to Figure 6) another light curve that has a much flatter baseline flux; however, in this example, there is a clear parallax signal evident in the asymmetry of the microlensing event. This relatively long event with the parallax signal demonstrates a limitation of observing Galactic bulge events, which are always unobservable for a portion of the year since the bulge is close to the ecliptic and therefore periodically close to the Sun; the seasonal viewing means there are large gaps in the time-series. This yearly pattern ensures larger uncertainty in the level of microlensing parallax since the residual from nonparallax models is also periodic on yearly timescales. Readers will notice the increased spread in the posterior samples (orange curves) in the third panel of Figure 7. In Section 6.5, we discuss how this presents a challenge for selecting black hole microlensing events with photometric microlensing data. This event has a much lower S/N and has little signal in the GP model.

In Figure 8, we show (analogously to Figure 7) a light curve that has both GP signal and a parallax signal. Once again, the
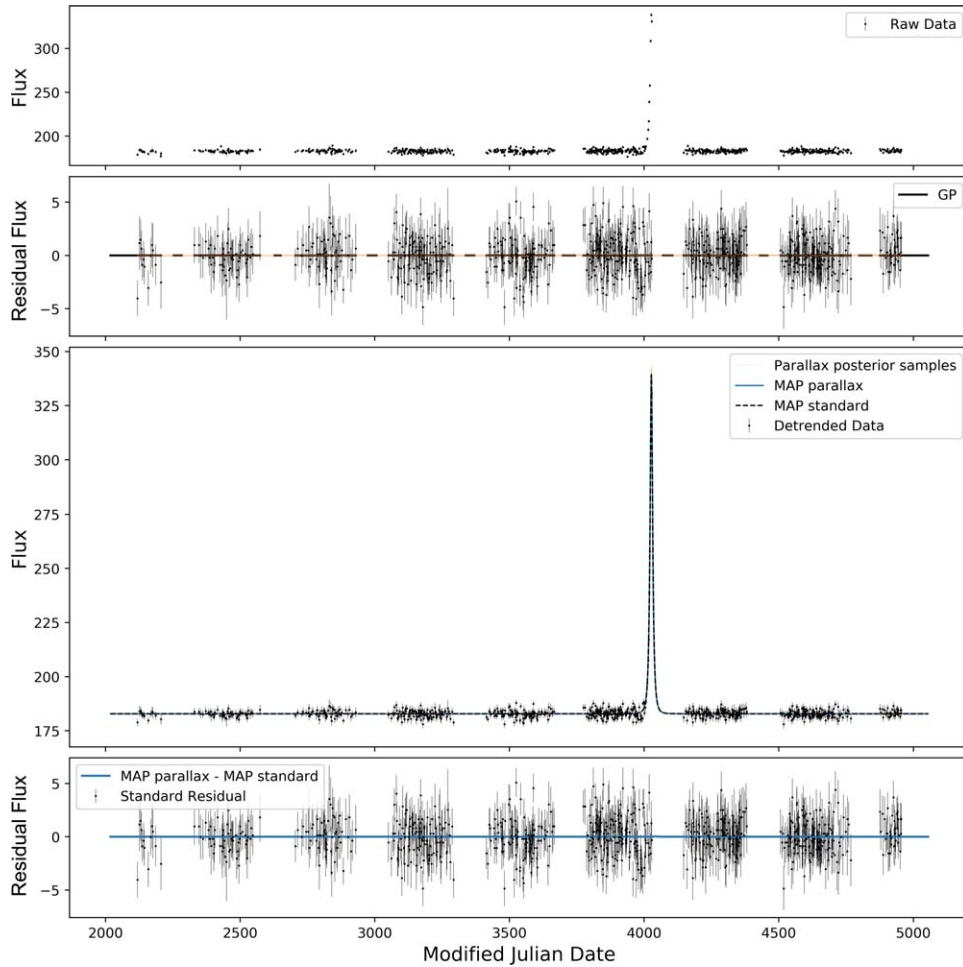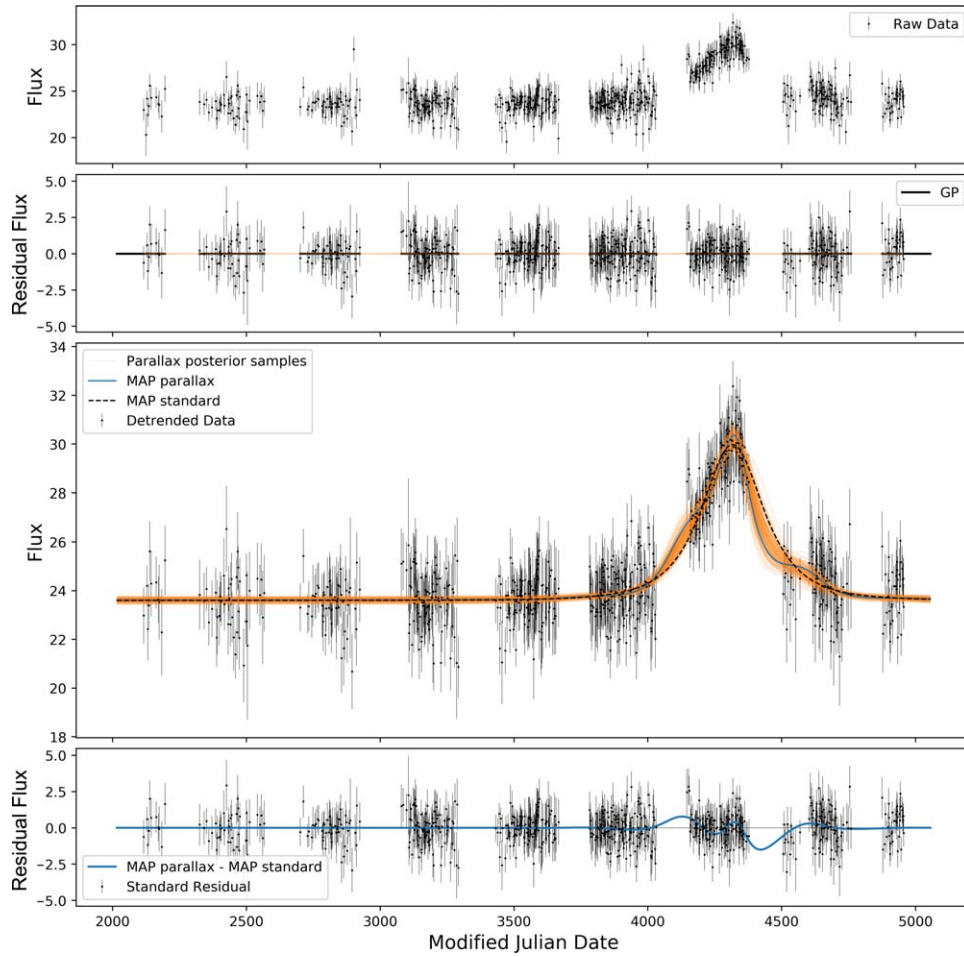
Figure 6. Example OGLE-III light curve (OGLE BLG 155.1.13052) with results from our MCMC analysis. This figure is analogous to Figure 5 but with a much simpler GP component. Again there is little parallax signal (PSPL+parallax+GP to PSPL+GP likelihood-ratio = 1.1).

spread in the orange curves of the third panel are largest in the seasonal observing gaps. This event demonstrates the power of the simultaneous GP and microlensing inference. The second and fourth panels show that the flux of the stellar variability and microlensing parallax signals are similar in magnitude as well as coincident in time. Nevertheless, the full model is again well sampled and unaffected by autocorrelation. The parallax signal is clearly detectable over at least three years.

In Figures 9 and 10, we present the longest timescale events in OGLE-III and -IV from our modeling, respectively. For the OGLE-III event, most of the parallax signal is present in the seasonal observing gap; however, there remains a parallax signal across consecutive observing seasons due to the long timescale. There is minimal baseline variability in this light curve. For the OGLE-IV event in Figure 10, the majority of the posterior samples suggest a dramatic parallax signal with the highest peak occurring in the gap between observing seasons, once again demonstrating this challenging aspect of observing long-duration microlensing events toward the Galactic bulge that exhibit strong parallax from a single site. This light curve also shows a discrepancy between the MAP and the majority of the posterior samples. There are only about 1% of the posterior samples trending down toward the MAP realization, which prefers a less dramatic peak in the data gap. Further inspection in the second panel of Figure 10 suggests that the GP signal in the MAP model is stealing a fraction of the parallax signal.

Again here, the vast majority of the posterior samples disagree with this picture. We discuss this in detail in Section 6.1.

The samples for all 12 model parameters for 3560 OGLE-III and 5790 OGLE-IV are available online.

5.2. Population

In Section 4, we described our model the global t_E distribution based on our samples for the individual microlensing event posterior PDFs. We chose a relatively simple model, a step-function (i.e., histogram), over $\log_{10} t_E$. We fixed the number of bins to 30 log-spaced bins between 0.5 and 3000 days for linear t_E. We used a Dirichlet prior and sampled the relative bin heights for the OGLE-III and OGLE-IV populations separately. This is a necessary choice since the two surveys are not coincident on the sky and therefore will not necessarily have the same intrinsic t_E distribution.

The results for our OGLE-III and OGLE-IV population t_E distribution modeling are presented in Figures 11 and 12, respectively. For OGLE-III, the results are plotted as a series of confidence intervals on the bin heights in comparison to the histogram of t_E values published by Wyrzykowski et al. (2015) for their PSPL model fits to the OGLE-III events (orange bins). We also plot the histogram of MAP t_E values for our PSPL+parallax+GP model fits (blue histogram). The bins are in general agreement for short-timescale events but start to diverge at greater than 2σ around 100 days, with PSPL fits
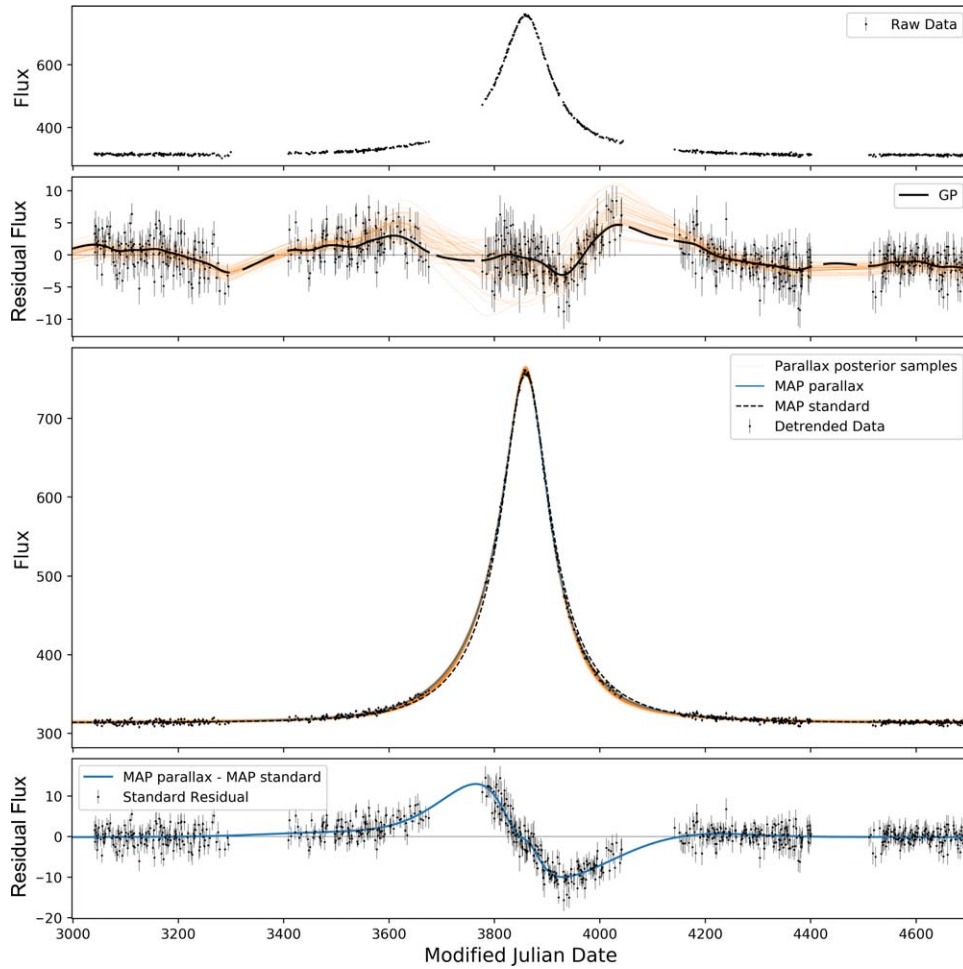
Figure 7. Example OGLE-III light curve (OGLE BLG 122.6.83113) with results from our MCMC analysis. This figure is analogous to Figure 5 but with a much simpler GP component and a clear parallax signal. Despite most of the parallax signal lying in the seasonal gaps in observability, the parallax model is still a better fit (PSPL+parallax+GP to PSPL+GP likelihood-ratio = 4.8).

showing more events with longer timescales. This remains true for all bins corresponding to $t_E > 100$ days. The hierarchical modeling is in general agreement with the MAP t_E histogram for our PSPL+parallax+GP fits across the domain.

For OGLE-IV (Figure 12), the results are plotted similarly except the orange histogram corresponds to the distribution of PSPL model t_E point estimates from Mróz et al. (2019). The results for our OGLE-IV analysis largely follow the same trends as the OGLE-III results. We find evidence for fewer long-timescale events as compared to the histogram of point estimates from the previous OGLE analysis, which again assumed a simple PSPL model.

6. Discussion

In this section, we have categorized some of the main points that were mentioned as worthy of further discussion throughout the text above as well as a handful of key concerns that the astute reader may have raised thus far.

6.1. Gaussian-process Model Fitting

The model parameters we have sampled include both the microlensing parallax and GP hyperparameters. These were fit simultaneously, meaning we can estimate correlations between them. This is important because one of the key worries of

utilizing GP models in microlensing is the concern that the GP model might absorb some or all of the actual microlensing signal. This is of further concern for our model since the yearly parallax signal causes a quasiperiodic deviation from the PSPL model curve, and our GP was designed to handle quasiperiodic signals in the baseline. In Figure 8, we showed an example fit with a temporally coincident GP and a parallax signal of approximately the same magnitude. In Figure 13, we show the posterior corner plot for this event, which shows the full set of one- and two-dimensional marginal posterior PDFs for all samples in the chain. By and large, the GP hyperparameters are more well behaved than the microlensing parameters; however, there is a slight correlation between t_E and the Matérn 3/2 parameters. This despite no obvious acceptance of a parallax signal by the GP model in Figure 8.

In order to explore the effect of including the GP model in our likelihood function, we injected and fit microlensing events. Since OGLE only published light curves with microlensing events, we selected the 50 events with the shortest t_E, so that when we removed the data points with a significant microlensing signal, we were left with as much data as possible. We masked the data between $t_0 \pm 5t_E$ and injected microlensing events drawn randomly from our prior distributions onto the masked baseline. We rescaled the uncertainty in the remaining flux measurements assuming Poisson noise. We
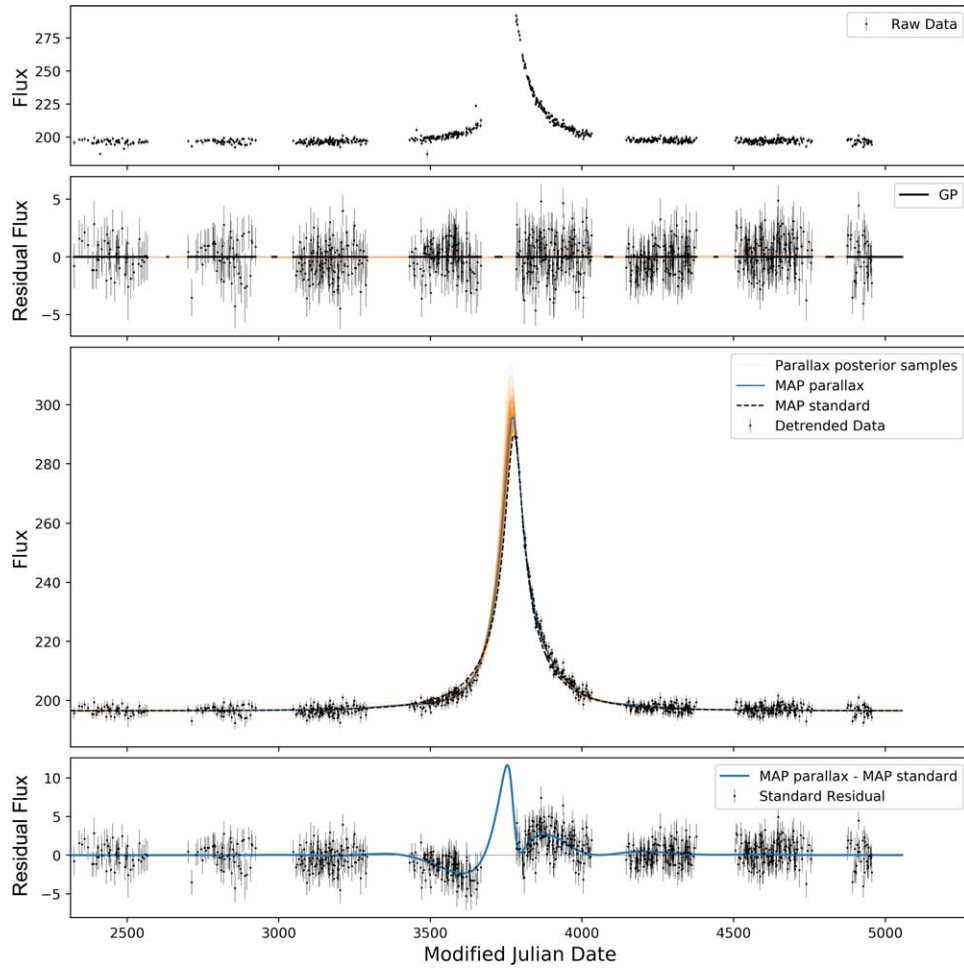
Figure 8. Example OGLE-III light curve (OGLE BLG 102.7.44461) with results from our MCMC analysis. This figure is analogous to Figure 5 but with a much simpler GP component and a clear parallax signal (PSPL+parallax+GP to PSPL+GP likelihood-ratio = 25).

fit these synthetic data with our model and computed the relative bias as follows:

$$\mathrm{bias} = \frac{\theta_{\mathrm{med}} - \theta_{\mathrm{inj}}}{\sigma_{\theta}}, \qquad (51)$$

where the difference in the numerator is between the marginal inferred median value and the injected truth for each parameter, and the denominator indicates the standard deviation in the chains for each parameter.

We present the results from this analysis in Table 2 as the parameters of a Gaussian fit to the bias values for each of the microlensing parameters. If there is no inherent bias, we expect the bias values to be normally distributed with zero mean and unit variance. We find that for most of the microlensing parameters this holds approximately true.

There is a slight positive bias for b_{ssf}, and the spread in relative bias for F_{base} is much smaller than expected. The relative bias in b_{sff} is largely due to poor inference in highly blended events. That is, when the injected value of b_{sff} is close to zero, the posterior tends to have a large spread, and the median value tends toward the median of the prior, which is closer to 0.5. On the other hand, the small spread in the bias for F_{base} is due to an enhancement in the denominator of Equation (51) for that parameter due to the GP model. An example of this is present in Figure 5, where the uncertainty in

the GP model manifests additional uncertainty in the baseline flux. The positive bias in u_0 may also be of concern, but this is again due to a one-sided prior (bounded at zero), and a tendency for low-S/N events to trend toward the median of the prior. Note that if we had included negative u_0 models, there would be a dichotomy in this bias (negative for $u_0 < 0$ injected models and positive for $u_0 > 0$). We again note that we checked for a bias in our t_{E} population modeling by including only $u_0 > 0$ and found none. This follows from the geometry.

Another test for our GP inference is to remove it and fit the same data with only the microlensing model. For the majority of our events, there is good agreement, since the baseline flux is generally smooth; although, in some cases, there are catastrophic breakdowns, which was the impetus for developing this GP in the first place. We did this for the OGLE-III events shown in Figures 5, 6, and 8 to demonstrate a range of scenarios. In Figure 14, we show overlapping posteriors for models with (orange) and without (green) the GP component of our model for the light curve shown in Figure 5. This demonstrates the importance of including a GP for events with strong variability in the baseline flux. Without the GP model, the microlensing model is forced into a corner of parameter space that has nearly zero flux from the source, a near-zero impact parameter, and an extremely long timescale despite this clearly not being the case upon visual inspection. It is also clear that the modeling including the GP model is handling parallax better. Given the short timescale of the
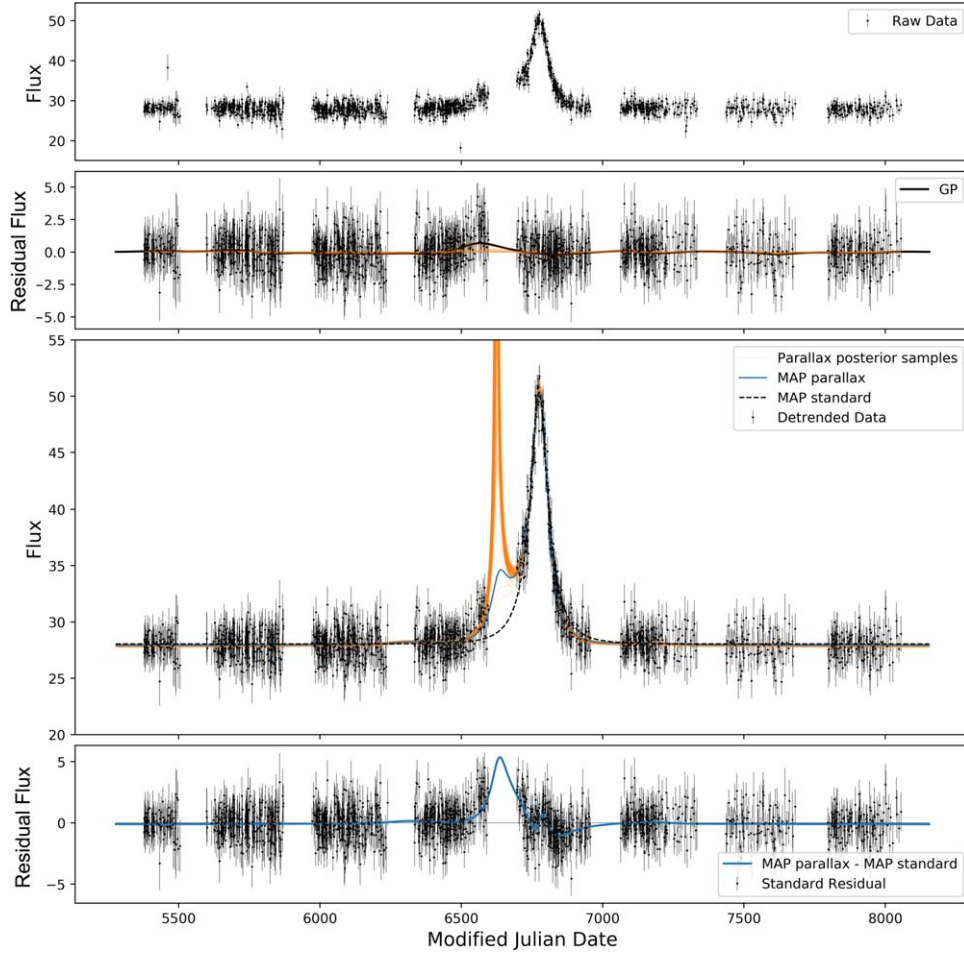
Figure 9. Example OGLE-III light curve (OGLE-III BLG 122.1.184151) with results from our MCMC analysis. This figure is analogous to Figure 5 and shows the results for the OGLE-III event with the largest Einstein crossing time. There is strong parallax (PSPL+parallax+GP to PSPL+GP likelihood-ratio = 8.8), but again much is apparently missing as it occurs in a gap in the data.

event, there is little ability to estimate π_E or ϕ, so the sampler explored the entire prior distribution. In Figure 15, which corresponds to the light curve in Figure 6, we show that when there is no baseline variability, the GP and non-GP posteriors are in excellent agreement. As a final example, we present Figure 16 corresponding to the light curve shown in Figure 8, which has both baseline variability (albeit much less than the case shown in Figure 5) and parallax signal. Here we found the posterior PDFs for all parameters were inflated with the inclusion of the GP into the model. This may seem counterintuitive since one might expect the more-complicated GP model to be able to estimate physical parameters more tightly, but it is expected since the GP model propagates additional uncertainty into the model. Similar to Figure 14, there is a significant bias in the no-GP model, however to a lesser degree, suggesting that the bias is correlated with the level of intrinsic stellar variability.

Figures 14–16 show the posterior PDFs for the two models for the microlensing parameters in common. There is an additional parameter we left out, $\log(K^2)$, since it was vastly different for the two models in light curves with baseline variability. In the absence of the GP model, this parameter is inflated in an attempt to mask over the variable baseline. The additional modeling flexibility of the GP is *more correct* in turning model biases into additional propagated uncertainty (the bias/variance trade-off). Also evident is the inflated

uncertainty in the F_{base} parameter in the GP model, especially in Figure 14. Cases such as this caused the low σ value for that parameter in Table 2. Figures 14 and 16 show the value of adding a GP to model the variable baseline. Without it, the model is overly confident and often biased (as is the case here–note the green posteriors in Figure 16 are often at the edge of the larger orange posteriors). In more catastrophic scenarios, the model is forced into extremely unlikely scenarios, as in Figure 14. Only when there is an intrinsically flat baseline, as in the best-case scenario (Figure 15), does the GP inference agree with the non-GP modeling.

6.2. Analysis of the Model Residuals

We performed three tests on the residuals of the model fits. In all cases, we computed the MAP model for two cases: one with and one without the GP portion of the model. We then subtracted that respective model from the data to examine the residuals. We retrieved a random subset of 100 light curves from both the OGLE-III and OGLE-IV data sets and carried out three tests on the two different residuals:

1. Augmented Dickey–Fuller test, which tests the null hypothesis that a time-series is nonstationary (i.e., it has time-dependent structure or nonconstant variance over time). This tests for stationarity in the residuals.

Figure 10. Example OGLE-IV light curve (OGLE-IV BLG 514.15.53029) with results from our MCMC analysis. This figure is analogous to Figure 5 except with OGLE-IV data from Mróz et al. (2019). It shows the results for the OGLE-IV event with the largest Einstein crossing time. This event has the most dramatic apparent parallax signal; however, the vast majority is fit within the observing gaps (PSPL+parallax+GP to PSPL+GP likelihood-ratio = 3.0).
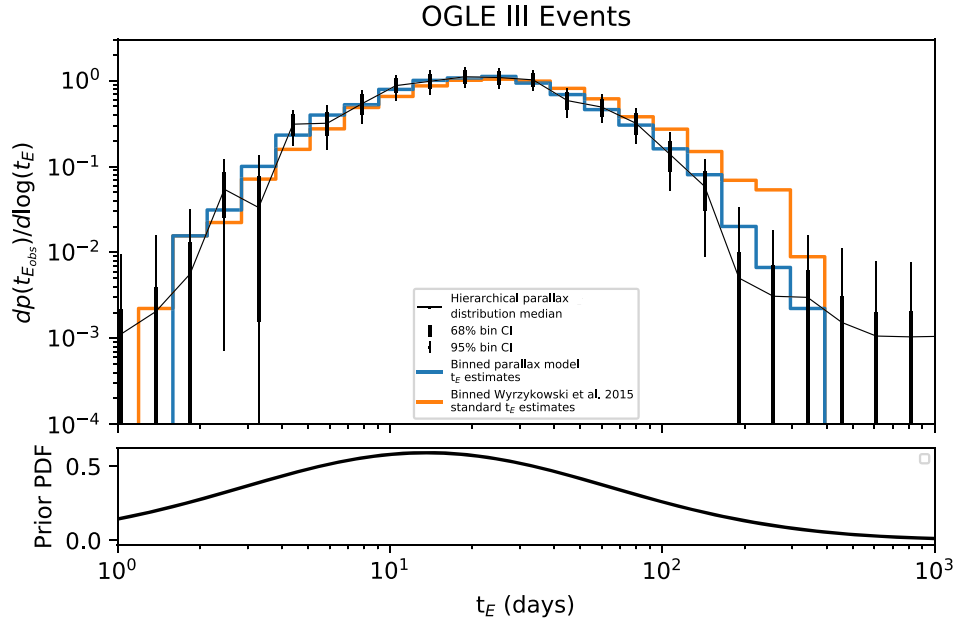
Figure 11. OGLE-III population results for the global observed t_E distribution (no survey efficiency corrections have been applied to any of the distributions). The black thin and thick bars show the 68% and 95% confidence intervals on the bin heights in our hierarchical model of the global distribution. The orange histogram shows the binned values presented in Wyrzykowski et al. (2015) for comparison, and the blue histogram shows the MAP-point estimates from our individual PSPL +parallax+GP model fits. In the lower panel, we show the prior PDF on t_E given by Equation (32). Note that the effect of our broadening of the t_E prior is apparent by the amplitude that varies from $\sim$0.1 to 0.5, whereas the actual distribution amplitudes vary by approximately 3 dex.

16
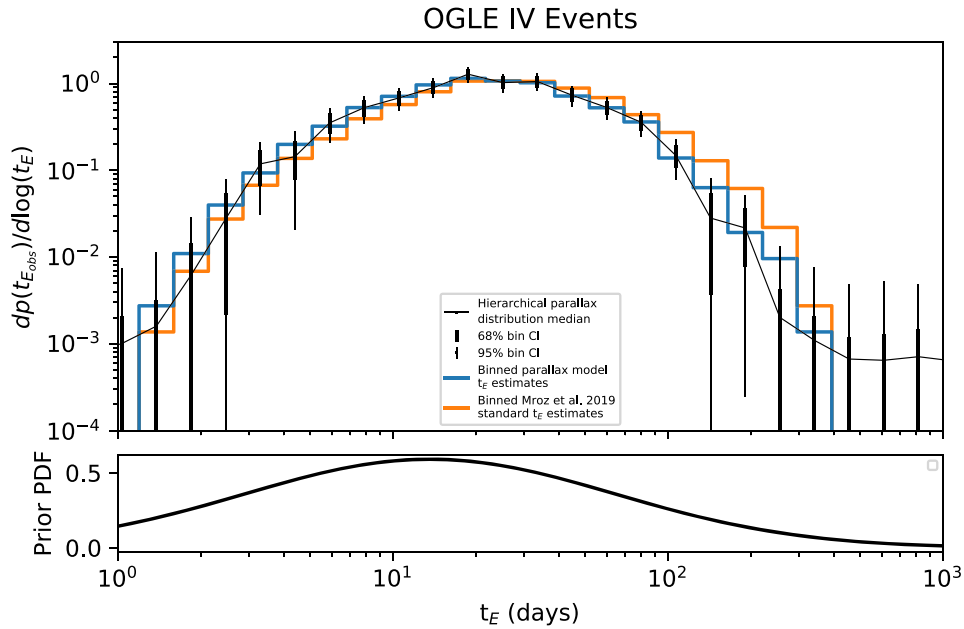
Figure 12. OGLE-IV population results for the global observed t_E distribution (no survey efficiency corrections have been applied to any of the distributions). The black thin and thick bars show the 68% and 95% confidence intervals on the bin heights in our hierarchical model of the global distribution. The orange histogram shows the binned values presented in Mróz et al. (2019) for comparison, and the blue histogram shows the MAP-point estimates from our individual PSPL+parallax +GP model fits. In the lower panel, we show the prior PDF on t_E given by Equation (32). Note that the effect of our broadening of the t_E prior is apparent by the amplitude that varies from ∼0.1 to 0.5, whereas the actual distribution amplitudes vary by approximately 3 dex.

2. Ljung–Box test, which tests the null hypothesis that a time-series is independently distributed (i.e., that the correlations are only the result of randomness in the sampling process). This tests for autocorrelations in the residuals.
3. Anderson–Darling test, which tests for departures from normality in the residuals.

In all tests, the model including the GP performs as well as or better than the model without the GP. In the cases where the GP performed significantly better than the model without the GP, it appears to be predominantly due to stars with high stellar variability suggesting the GP is performing as intended. Only one event with the GP model failed the augmented Dickey–Fuller test, and this was for an event where only the first part of the event (i.e., before peak) was observed, and the MAP did not well model the actual event. While the with-GP model performed better than the no-GP model in all events for the Ljung–Box test, there are still 20% of the with-GP residuals that have a p-value <0.05 suggesting that significant autocorrelations persist in the residuals. This is not surprising since our GP model is still a constrained model that is not optimized for various types of stellar variability. Finally, we find that 15% of the no-GP residuals fail the Anderson–Darling test, with critical values computed at the 99% level, while 13% of the with-GP residuals fail the test. These failures are correlated with the Ljung–Box failures and events with large intrinsic variability where the GP did not fully model the variability. While the GP model is not perfect at modeling the stellar variability and other systematics, it does improve the fit relative to the model excluding the GP component.

6.3. Multimodality of Individual Posterior PDFs

It is well documented that the parallax signal in photometric microlensing data leads to a multimodal likelihood (Smith et al. 2002; Gould 2004; Poindexter et al. 2005; Smith et al. 2005).

There are numerous degeneracies that have different values for microlensing parameters. These issues mostly arise for short-timescale events. It is typical in the literature to identify these degeneracies and initialize sampling algorithms in these subpeaks and present the maximum likelihood models for each of the degenerate solutions. In our analysis, we have not gone through this procedure since we are not interested in the peculiarities of any single event.

The degeneracies emerge from ambiguities on the components of the relative proper-motion vector parallel and perpendicular to the Earth's acceleration. Since this is purely geometric and randomly distributed over the thousands of OGLE-III and OGLE-IV events, we allow our sampler to settle in a random mode for individual events by initializing the chains randomly under the prior PDF. The differences in the modes of any single event do not affect the population analysis, especially for long-duration events ($t_E > 100$ days) where the difference in t_E for different modes is fractionally small, especially for low-π_E events, which are most likely to be black holes (see Section 6.5). However, we do caution against downloading the chains for an individual event and making inference on that event without a fuller exploration of the posterior.

6.4. The t_E Distribution

The peaks of the t_E distribution that we estimated using the samples of all of our fits largely agree with the histograms of point estimates presented by Wyrzykowski et al. (2015), Mróz et al. (2019)—this despite our prior having a lower peak timescale. Our prior was constructed from the OGLE-like survey simulated by Lam et al. (2020) where it was found that two different Galactic bar angles could match either the observed stellar density map or the microlensing event rates; but not both. In all cases simulated, the t_E distribution peaked at shorter timescales than the observed distributions (see also Medford et al. 2020). As can be seen from the bottom panels of
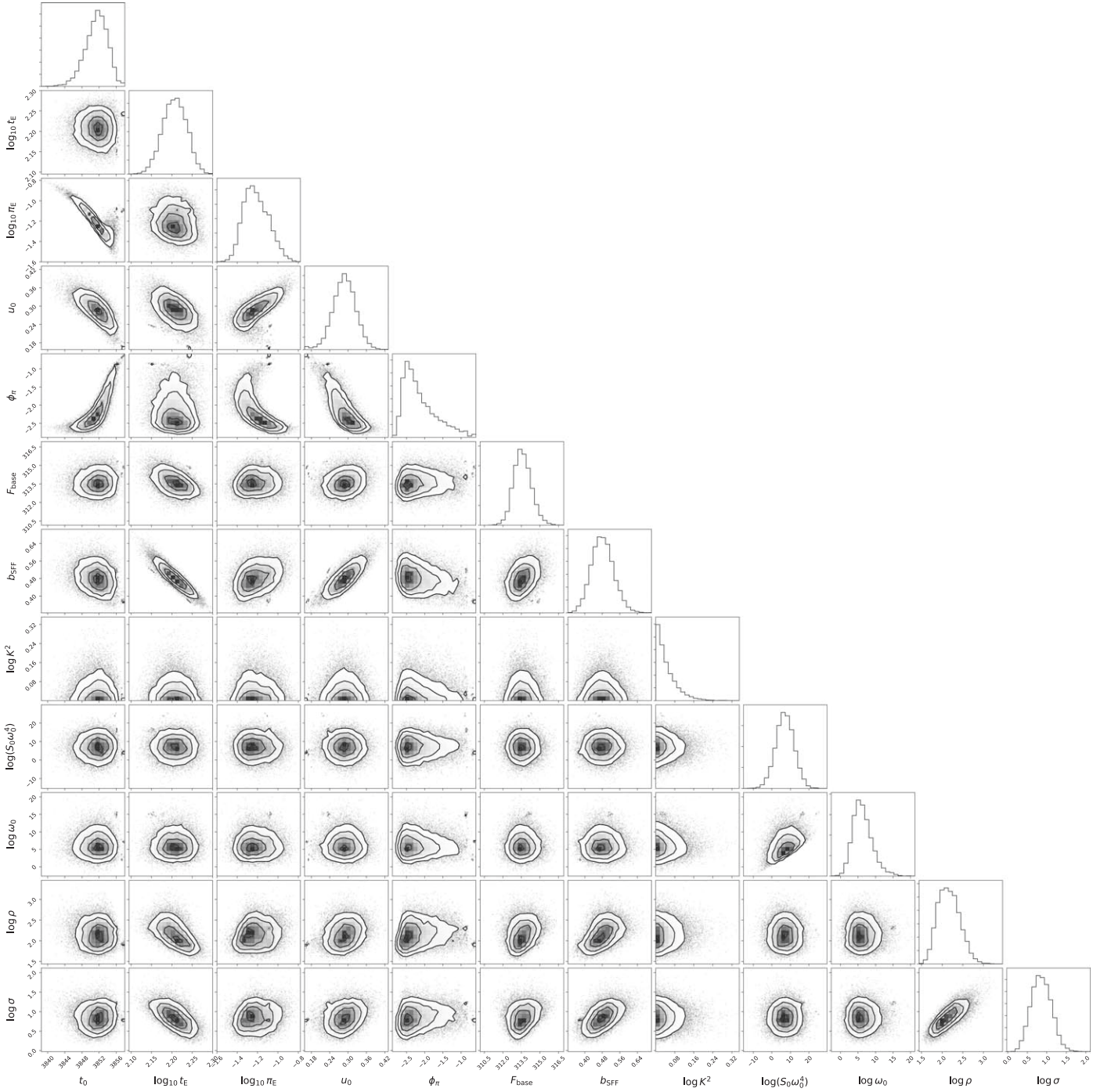
Figure 13. Posterior corner plot for OGLE-III BLG 102.7.44461. The strongest correlations are between microlensing parameters. There is a slight correlation between t_E and the Matérn-3/2 kernel parameters; however, in general, the GP hyperparameters are much more well-behaved than the microlensing parameters.

Figures 11 and 12, we significantly broadened the PopSyCLE-based prior to reduce the impact of this discrepancy.

There are significantly fewer long-timescale events across a range of timescales above 100 days in our t_E distribution as compared to both the OGLE-III and -IV results.[14] However,

---

[14] To investigate whether the discrepancy between our results is due to the t_E prior decreasing at longer t_E (see, e.g., the bottom panel of Figure 11), we importance sampled the individual light-curve samples, replacing the prior of Table 1 with a uniform prior in $\log(t_E)$ space from 1 day $< t_E < 1000$ days. As can be seen from the results of this exercise in Appendix, the decrement of high-t_E events relative to the distributions of Wyrzykowski et al. (2015), Mróz et al. (2019) is not due to our choice of prior.

Table 2
Summary of Injection Analysis

Parameter	μ_{bias}	σ_{bias}
F_{base}	-6.8×10^{-4}	0.040
b_{sff}	0.32	1.1
t_0	-7.8×10^{-3}	1.4
$\log_{10} u_0$	0.16	1.3
$\log_{10} t_E$	-0.12	0.52
$\log_{10} \pi_E$	0.11	0.77
ϕ_π	-0.029	0.97

Note. Here μ and σ refer to the mean and standard deviation of the bias (see Equation (51)) for each parameter listed.
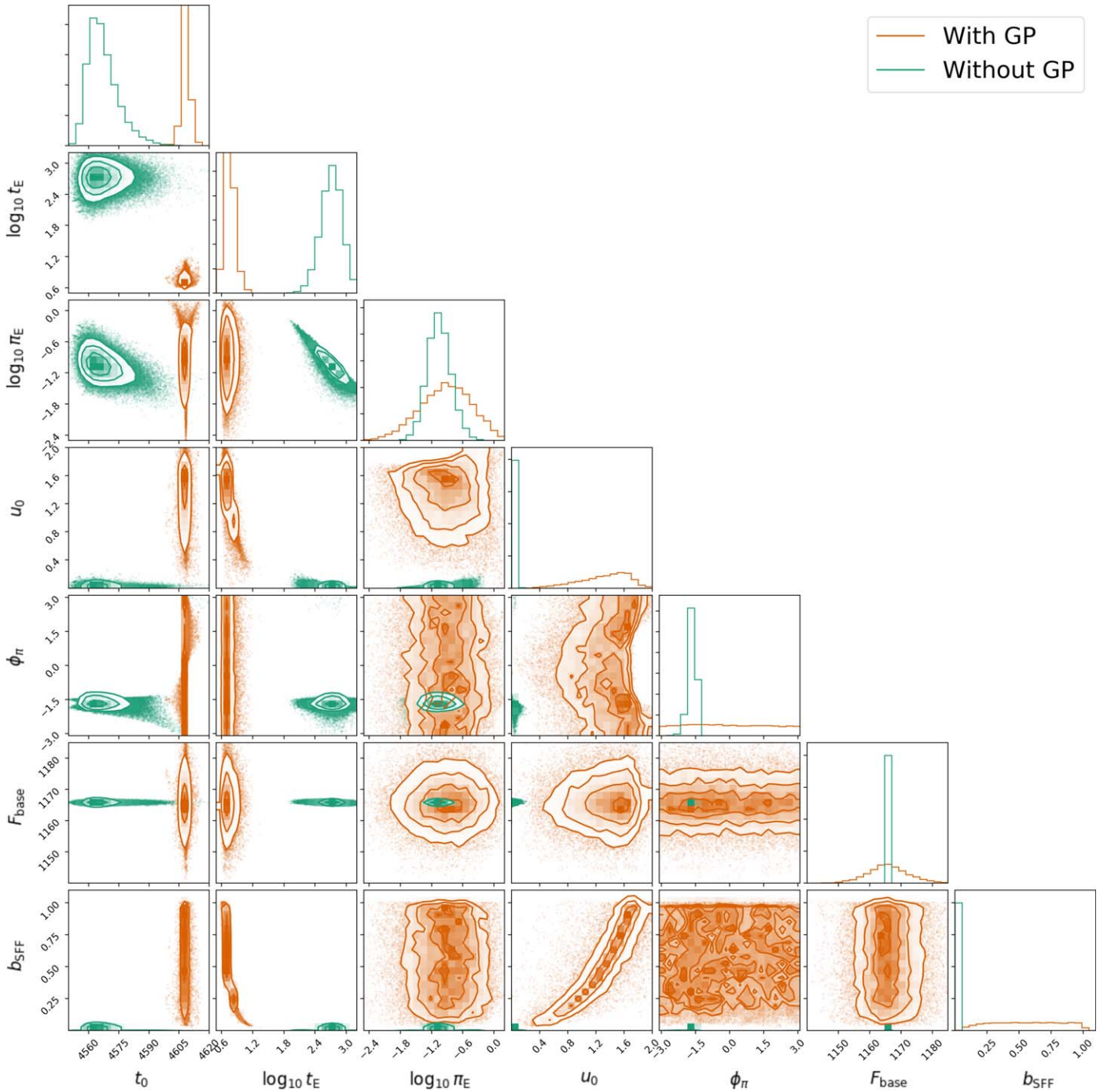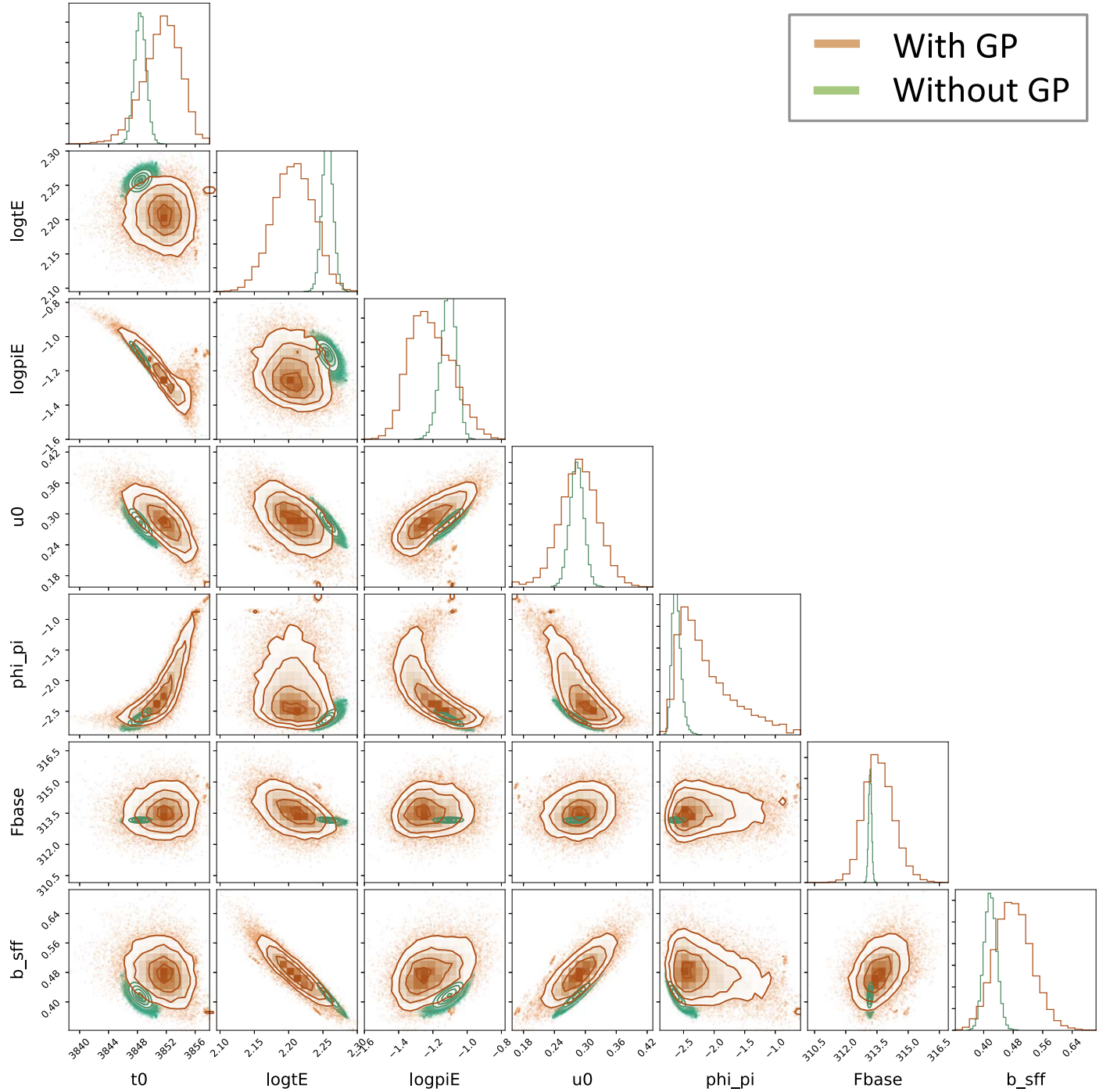
Figure 14. Posterior corner plot for OGLE-III BLG 156.7.141434 (see Figure 5) covering the microlensing model parameters with and without the GP component of the model.

both distributions also suggest that there are multiple events with $t_E > 400$ days despite none being reported in either point-estimate histogram. This is due to the accumulation of high-t_E tails adding weight to these bins. Note that both Wyrzykowski et al. (2015), Mróz et al. (2019) apply a cap on the maximum PSPL t_E point estimates of a few hundred days during the event detection and distillation process. We recommend not applying such long t_E selection limits in the future, as they greatly impact the ability to conduct black hole studies.

Previous studies of the t_E distribution often make reference to the slope of the short- and long-timescale rise of the distribution and the position of the peak. They especially note agreement with theory (e.g., Mao & Paczynski 1996) if the

slopes of the log t_E distribution tails are ± 3 (e.g., Wyrzykowski et al. 2015; Mróz et al. 2019). This "expectation" for $\log t_E = \pm 3$ slopes is an oversimplification of the potential Mao & Paczynski (1996) model diversity. While Mao & Paczynski (1996) show that this is true for many choices of present-day mass function and galactic velocity distributions. They also show that it is not a universal tendency (see, e.g., Figures 1 and 3 of their paper). The choice of slope (α) for the present-day mass function can greatly affect both the high- and low-end t_E slope, with the potential for the effect to increase with wider mass distributions. Note though that, for single power-law mass models, varying α usually only causes t_E slope deviations from ± 3 on one end or the other. For example, a
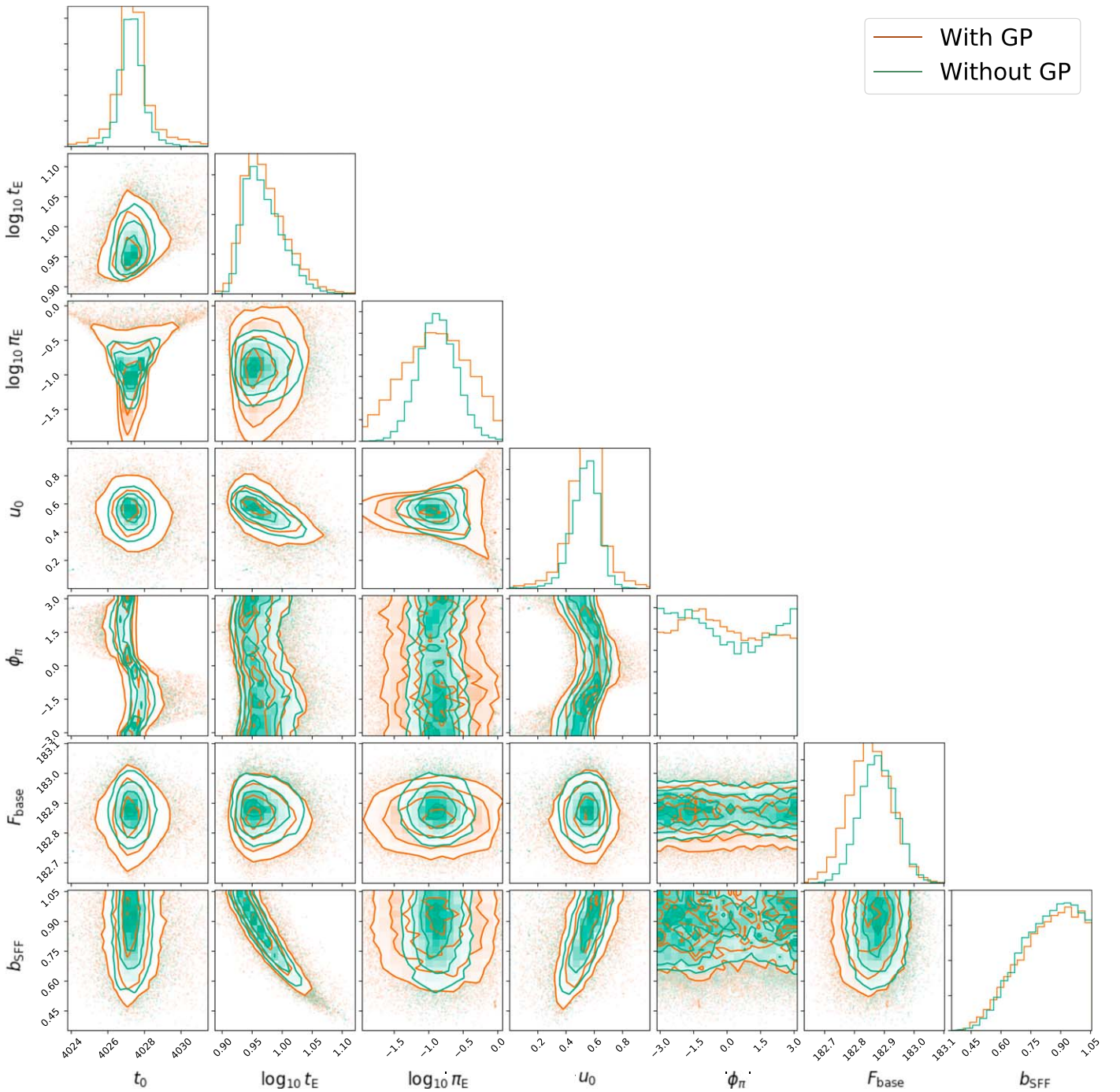
Figure 15. Posterior corner plot for OGLE-III BLG 155.1.13052 (see Figure 6) covering the microlensing model parameters with and without the GP component of the model.

power-law mass distribution with a steeper negative slope (e.g., $\alpha = -2.5$; i.e., more total mass in low-mass objects) will maintain the $+3$ slope on the low-t_E end; however, this will result in a shallower negative slope on the high-t_E end (i.e., slope > -3). Notably, their model with $\alpha = -2.5$, $\beta = 2$ shifted the peak of the t_E distribution to the left, which is evident in the models presented by Lam et al. (2020). More-complex mass distributions, such a bi-modal distributions, can also cause deviations in the slopes (see Figure 17; and Lu et al. 2019).

Based on Mao & Paczynski (1996) and our arguments above, it is not clear how it is physically possible to produce a

slope steeper than ± 3.[15] Note that the high-t_E slopes of our distributions are consistent with -3 although favoring slightly steeper slopes. This may be indicative of bias due to the long t_E

---

[15] Wyrzykowski et al. (2015), Mróz et al. (2019) only provide t_E efficiency corrections based on point-source, point-lens models without parallax, meaning we are not able to reliably apply their efficiency corrections to our parallax-based t_E distribution (additionally, as Mróz et al. (2019) note, their corrected distribution will be biased at the high-t_E end). Furthermore, without access to the full OGLE data sets, we are unable to develop appropriate efficiency corrections. Thus, care should be taken when interpreting our t_E distributions in terms of the Mao & Paczynski (1996) formalism. That said, both Wyrzykowski et al. (2015), Mróz et al. (2019) suggest rather flat efficiency corrections at the high-t_E end, so it seems reasonable to consider the high-t_E slope with respect to the theoretical expectations for the slope.

Figure 16. Posterior corner plot for OGLE-III BLG 102.7.44461 (see Figures 8 and 13) covering the microlensing model parameters with and without the GP component of the model.

selection cuts of Wyrzykowski et al. (2015), Mróz et al. (2019) applied to the t_E Paczynski (1986) model parameter. Mao & Paczynski (1996) also note that the long tails of the t_E distribution will also be modified by the motion of the Earth. However, when we use the Paczynski (1986) model with GP, we find a distribution more consistent with our parallax+GP model distribution, suggesting that the bias may be primarily due to not modeling the intrinsic stellar variability, as we discussed in Section 6.1.

6.5. Selecting Black Hole Candidates

Black hole microlensing has not been mainstream science since the early constraints on MACHO dark matter in the 1990 s and 2000 s. There have been searches for black holes in the MACHO Survey (Agol et al. 2002; Bennett et al. 2002; Poindexter et al. 2005) and OGLE-III (Lu et al. 2016; Wyrzykowski et al. 2016; Wyrzykowski & Mandel 2020) since then, but as MACHOs went out of style as a dark matter
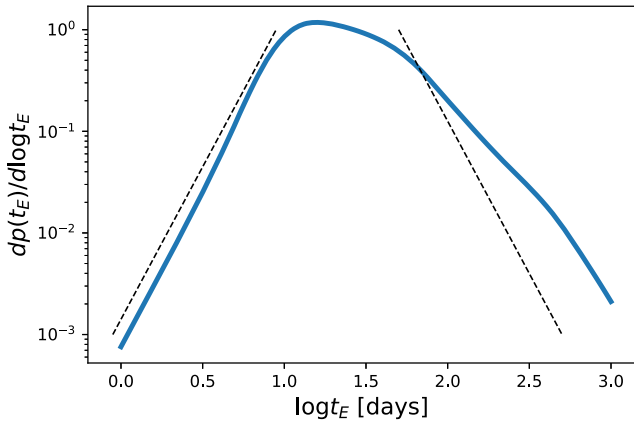
Figure 17. An example t_E distribution from the Mao & Paczynski (1996) formalism, modified for a mixture model composed of two separate power-law mass distributions (blue curve). Mixture component 1 has a min and max mass range of 0.1–3 $M_\odot$ and power-law slope of −1.5, while mixture component 2 has a min and max mass range of 1–100 $M_\odot$ and power-law slope of −1.5. The relative proportion of each mixture is 0.8 and 0.2, respectively. The black dashed lines plot ±3 power-law slopes. There is clearly deviation from the −3 slope due to the second mass component.

candidate, the focus instead was placed on detecting astrophysical black holes. Interest further heightened for both types of black holes with LIGO as it revealed both higher-than-expected black hole masses (Bird et al. 2016) without resolving the possible mass gap between the neutron star and black hole populations (Littenberg et al. 2015).

For either type of black hole, as the mass increases the event moves down and to the right in $t_E − \pi_E$ space, meaning searching for longer timescale events with less microlensing parallax is the best way to identify black hole lenses photometrically. This has not been fully appreciated in any black hole search to date, although Lam et al. (2020) pointed it out in their Figure 13. Estimating by eye from their Figure 13, the purity of astrophysical black holes is very high with events satisfying the following:

$$\log_{10} \pi_E < 0.8 \log_{10} t_E + b, \qquad (52)$$

where b is an arbitrary intercept. We varied the intercept of this inequality until the mock OGLE-III and -IV catalogs (separately) from Lam et al. (2020) yielded a 50% purity of black holes. We found $b = −2.67$ for OGLE-III and $b = −2.57$ for OGLE-IV. We then checked our fits for these two surveys. Using the median marginal $\log_{10} t_E$ and $\log_{10} \pi_E$ values, 107 and 283 events lie within the region of 50% purity (see Figure 18).

We note that this does not mean that there are necessarily 50% of 107 and 283 black holes detected in OGLE-III and -IV, respectively. In fact, there are likely more. The systematics involved with these data (seasonal gaps in the observability from one hemisphere, noisy ground-based data, time-domain variability of the baseline, unresolved blended stars that are not well characterized, to name a few) make it challenging to estimate small π_E for many black hole events. It is likely that π_E values from our modeling are overestimated for many low-S/N events with intrinsically small π_E. In those cases, the posterior is more heavily weighted by the prior PDF than the likelihood (the data), and thus the median value of the π_E marginal posterior samples is skewed high. Our prior was developed from the PopSyCLE simulations and is strongly driven by stellar microlensing events with higher π_E than the

typical black hole microlensing event toward the Galactic bulge. This is also evident in Table 2 where the mean of $\log_{10} \pi_E$ is biased high—when the sampler can identify the parallax signal, it fits it well. This occurs more frequently for intrinsically high-π_E microlensing events, but when it cannot fit for π_E well, it reverts to the prior. This is especially evident in Figure 18, for $t_E < 30$ days, where the parallax signal is small no matter the intrinsic value. This is the cause of the ridge of events that align with $\pi_E \approx 0.1$. Since we are plotting median values, we are effectively reporting the median of the prior for these events.

This, coupled with the fact that the black hole purity increases as t_E increases and π_E decreases and there are a significant number of events offset in this direction from the 50% purity line, means that 50% of 107 and 283 black holes detected in OGLE-III and -IV is likely a lower limit based on the Lam et al. (2020) model. This is supported by the lower panels in Figure 18, where the posteriors for individual events near the cut line clearly extend to both sides. This is the expectation though; the mass-distance degeneracy in photometric microlensing data typical makes inferring individual events as black holes to be very challenging. We have presented this figure to motivate population studies of photometric microlensing. Further estimation on our samples is beyond the scope of our analysis here, but it will be explored in detail in a follow-up paper (W. Dawson et al. 2022, in preparation).

Another avenue for selecting black hole candidates is to make cuts on b_{sff} (e.g., Wyrzykowski et al. 2016). However, this parameter is notoriously challenging to fit with OGLE data as the vast majority of events are highly blended, and it is difficult to find events where all of the flux is from the source star. Wyrzykowski & Mandel (2020) have shown that when Gaia parallax and proper motion can be coupled with the microlensing fits for select events with high b_{sff}, this complexity can be reduced, enabling this potentially powerful method.

Even without blending cuts, a relatively pure set of candidates could be generated to follow up with extensive astrometric observations to model the astrometric shift in the photometric centroid. We are actively carrying out such a survey with Keck adaptive optics (PI: Lu).

6.6. Conclusions

We have modeled all public Galactic bulge microlensing events from OGLE-III and -IV, yielding the most comprehensive public catalog of MCMC fits for microlensing light curves to date. Our model includes both microlensing parallax due to Earth's orbital motion and a GP model for variable baseline flux due to intrinsic stellar variability. This was a computationally intensive task, and our modeling used roughly 1 million CPU-hours on Lawrence Livermore National Laboratory's high-performance computing resources.[16] Below we offer a list of the main takeaways of our analysis.

1. We obtained 3560 from Wyrzykowski et al. (2015) and 5790 microlensing events from Mróz et al. (2019).
2. Our PSPL+parallax+GP model was able to capture and separate the yearly parallax signal and the intrinsic variability in the baseline from flux of the source and/or

---

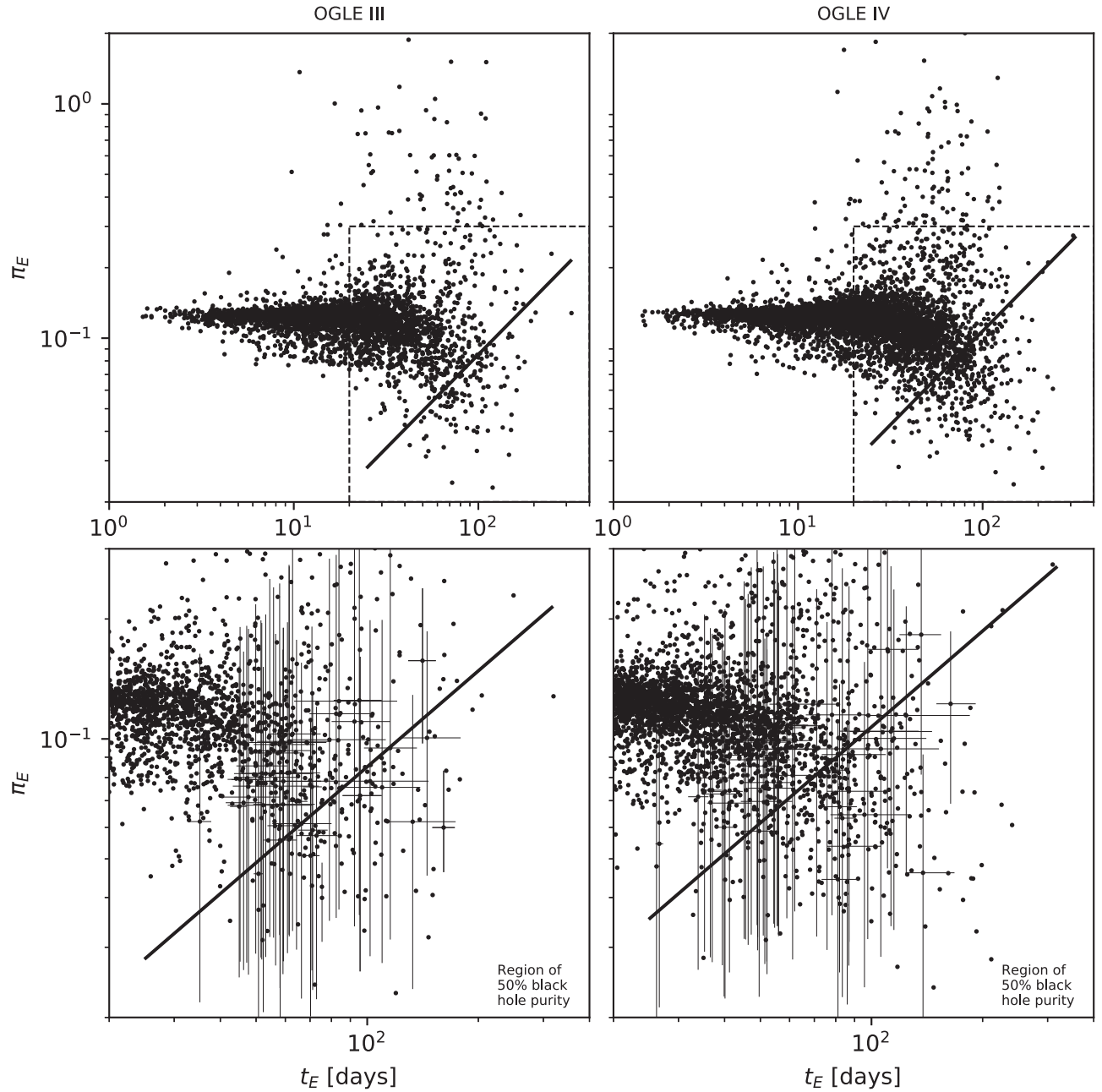[16] We used the Quartz computing system: https://hpc.llnl.gov/hardware/platforms/Quartz.

Figure 18. Top row: marginal t_E–π_E values for 3560 OGLE-III (left panel) and 5760 OGLE-IV (right panel) microlensing events. The points represent the median posterior values for the two parameters. The lines demarcate 50% purity of astrophysical black holes in the PopSyCLE OGLE-III and -IV simulations (Lam et al. 2020); i.e., a random event below the line is roughly 50% likely to be a black hole conditioned on the PopSyCLE simulations. There are 107 and 283 events below the lines for OGLE-III and -IV, respectively. The bulk of the short-timescale events lie along the median of the prior PDF for the timescale since there are few points to allow the likelihood to pull the posterior PDF away from the prior. The posterior PDF is wide for any given event here (see the lower panels for typical uncertainty in a given event). The dashed box is presented as an inset below. Bottom row: zoom into the likely black hole regime. We have presented the same data as the above panels and added 68% confidence intervals to a random selection of 50 points to show that the confidence in any one point is low; however, the bulk of the posterior is below the cut for many events. Furthermore, even events well above the line have tails that extend well into the region below the cut line. Thus, the numbers of point estimates below the line (107 and 283 for OGLE-III and -IV on the left and right, respectively) are likely conservative.

blended light. This represents the largest-ever catalog of microlensing events modeled for the parallax signal due to Earth's motion around the Sun.

3. We have made public 20,000 samples for all 12 parameters in our model (7 microlensing and 5 GP) for all 9350 OGLE-III and -IV events we analyzed.

4. Nonidentifiability is a common critique of GP analyses. We completed an analysis for bias in our results and explained our findings in terms of our model. We did not

identify any issues with the GP model systematically overfitting the parallax signal that should have been modeled by the microlensing model.

5. We developed a Bayesian hierarchical forward model for the t_E distribution following the sample reweighting method developed by Hogg et al. (2010). We found that the parallax and GP modeling favors fewer events with $t_E > 100$ days than were reported by Wyrzykowski et al. (2015), Mróz et al. (2019). Motivated from theoretical

modeling (Mao & Paczynski 1996), we showed how different physical model choices can manifest starkly different t_E distributions (Figure 17). One could follow our population analysis method and use our published samples to forward the model physical characteristics of the Galaxy.

6. Using PopSyCLE (Lam et al. 2020), we demonstrate how relatively pure catalogs of black holes can be identified from photometry alone. We found 390 microlensing events in OGLE-III and -IV that PopSyCLE simulations suggest are ~50% black hole lensing events. While this benefits from the full rise and fall of the microlensing event, if parallax can be constrained below such a cut on the rising portion of an event, this technique could be used to identify the best targets for astrometric follow-up.

7. We recommend not applying long t_E selection limits in future population studies of microlensing events, as they greatly impact the ability to conduct black hole studies.

8. At a minimum, when working with OGLE light curves, the baseline flux must be checked for quasiperiodic and low-order polynomial trends, especially for long-time-scale events. We demonstrated catastrophic failures in an MCMC analysis by not modeling these parameters. We recommend the use of additional model parameters to specifically handle these systematics where present.

Appendix
The t_E Distribution with a Uniform Prior

As noted in Section 6.4, we observed a decrement of $t_E > 100$ day events relative to the distributions of Wyrzykowski et al. (2015), Mróz et al. (2019), see Figures 11 and 12. To investigate whether our prior is responsible, we importance sampled the individual light-curve samples, replacing the t_E prior of Table 1 with a uniform prior in $\log(t_E)$ space from 1 day $< t_E < 1000$ days. As can be seen from the results of this exercise in Figures 19 and 20, the decrement of high-t_E events relative to the distributions of Wyrzykowski et al. (2015), Mróz et al. (2019) is not due to our prior. This result is as expected since the prior of Table 1 is much broader than the typical
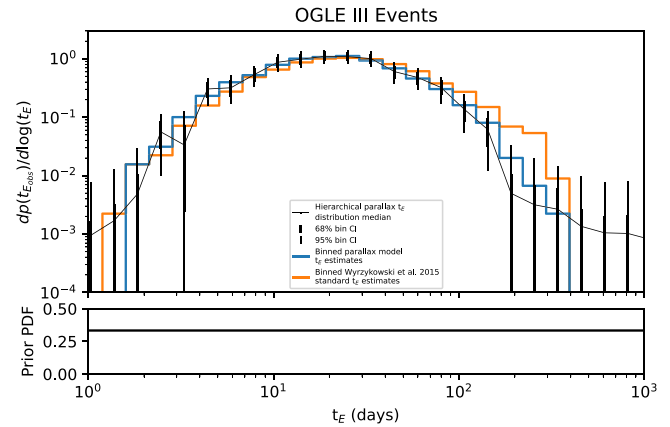
Figure 19. The same as Figure 11 except that the hierarchical parallax t_E distribution (black) in the top panel corresponds to events fit with a uniform prior in $\log(t_E)$ space from 1 day $< t_E < 1000$ days (see bottom panel) instead of the prior noted in Table 1 and shown in Figure 11. We observe the same significant decrement of $t_E > 100$ day events relative to the distribution from Wyrzykowski et al. (2015; orange), suggesting that the prior is not responsible for the observed decrement.
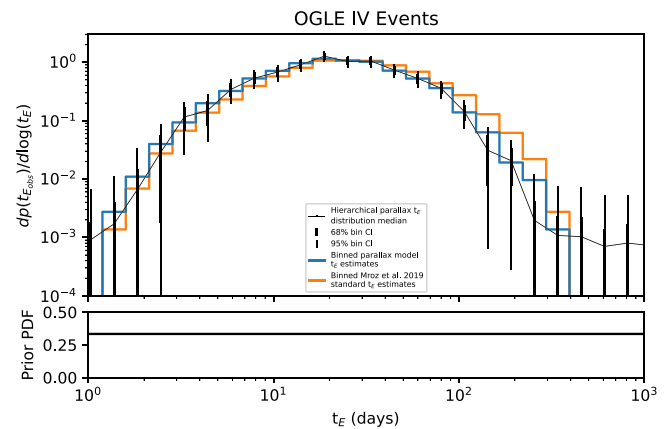

Figure 20. The same as Figure 12 except that the hierarchical parallax t_E distribution (black) in the top panel corresponds to events fit with a uniform prior in $\log(t_E)$ space from 1 day $< t_E < 1000$ days (see bottom panel) instead of the prior noted in Table 1 and shown in Figure 11. We observe the same significant decrement of $t_E > 100$ day events relative to the distribution from Mróz et al. (2019; orange), suggesting that the prior is not responsible for the observed decrement.

posterior distribution of any given light-curve fit, thus further broadening the prior is unlikely to cause a significant change.

ORCID iDs

Nathan Golovich ● https://orcid.org/0000-0003-2632-572X
William Dawson ● https://orcid.org/0000-0003-0248-6123
Fran Bartolić ● https://orcid.org/0000-0001-8630-9794
Casey Y. Lam ● https://orcid.org/0000-0002-6406-1924
Jessica R. Lu ● https://orcid.org/0000-0001-9611-0009
Michael S. Medford ● https://orcid.org/0000-0002-7226-0659
Michael D. Schneider ● https://orcid.org/0000-0002-8505-7094
Edward F. Schlafly ● https://orcid.org/0000-0002-3569-7421
Alex Drlica-Wagner ● https://orcid.org/0000-0001-8251-933X
Kerianne Pruett ● https://orcid.org/0000-0002-2911-8657

References

Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, Asteroseismology (Dordrecht: Springer)
Afonso, C., Albert, J. N., Andersen, J., et al. 2003, A&A, 400, 951
Agol, E., & Kamionkowski, M. 2002, MNRAS, 334, 553
Agol, E., Kamionkowski, M., Koopmans, L. V. E., & Blandford, R. D. 2002, ApJL, 576, L131
Agol, E., Luger, R., & Foreman-Mackey, D. 2020, AJ, 159, 123
Alcock, C., Allsman, R. A., Alves, D., et al. 1995, ApJL, 454, L125
Alcock, C., Allsman, R. A., Alves, D., et al. 1996, ApJ, 471, 774
Alcock, C., Allsman, R. A., Alves, D., et al. 1998, ApJL, 499, L9
Alcock, C., Allsman, R. A., Alves, D. R., et al. 2001, ApJL, 550, L169
Alcock, C., Allsman, R. A., Axelrod, T. S., et al. 1993, in ASP Conf. Ser. 43, Sky Surveys. Protostars to Protogalaxies, ed. B. T. Soifer (San Francisco, CA: ASP), 291
Arnaud, M., Aubourg, E., Bareyre, P., et al. 1994, ExA, 4, 265
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bachelet, E., Street, R., & Bozza, V. 2018, AAS Meeting, 231, 228.05
Bailer-Jones, C. A. L. 2015, PASP, 127, 994
Bennett, D. P., Becker, A. C., Quinn, J. L., et al. 2002, ApJ, 579, 639
Betancourt, M. 2017, arXiv:1701.02434
Bird, S., Cholis, I., Muñoz, J. B., et al. 2016, PhRvL, 116, 201301
Brewer, B. J., & Stello, D. 2009, MNRAS, 395, 2226
Byrd, R. H., Nocedal, J., & Yuan, Y.-X. 1987, SJNA, 24, 1171
Cranmer, S. R., Bastien, F. A., Stassun, K. G., & Saar, S. H. 2014, ApJ, 781, 124
Czekaj, M. A., Robin, A. C., Figueras, F., Luri, X., & Haywood, M. 2014, A&A, 564, A102
Dominik, M. 1998, A&A, 329, 361
Dominik, M. 2009, MNRAS, 393, 816
Durrande, N., Ginsbourger, D., & Roustant, O. 2011, arXiv:1103.4023
Duvenaud, D. 2014, PhD thesis, Univ. of Cambridge, https://www.repository.cam.ac.uk/handle/1810/247281
Einstein, A. 1936, Sci, 84, 506
Foreman-Mackey, D. 2016, JOSS, 1, 24
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Foreman-Mackey, D., Czekala, I., Luger, R., et al. 2019, exoplanet-dev/exoplanet v0.2.4, Zenodo, doi:10.5281/zenodo.1998447
Foreman-Mackey, D., Czekala, I., Luger, R., et al. 2020, exoplanet-dev/exoplanet v0.3.0, Zenodo, doi:10.5281/zenodo.1998447

Gaudi, B. S. 2012, ARA&A, 50, 411
Gould, A. 1992, ApJ, 392, 442
Gould, A. 2000, ApJ, 542, 785
Gould, A. 2004, ApJ, 606, 319
Gould, A. 2013, ApJL, 763, L35
Harvey, J. 1985, in ESA Special Publication 235, Future Missions in Solar, Heliospheric & Space Plasma Physics, ed. E. Rolfe & B. Battrick (Noordwijk: ESA), 199
Hoffman, M. D., & Gelman, A. 2011, arXiv:1111.4246
Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166
Hunter, J. D. 2007, CSE, 9, 90
Kallinger, T., De Ridder, J., Hekker, S., et al. 2014, A&A, 570, A41
Kipping, D. M. 2013, MNRAS, 435, 2152
Lam, C. Y., Lu, J. R., Hosek, M. W. J., Dawson, W. A., & Golovich, N. R. 2020, ApJ, 889, 31
Li, S. S., Zang, W., Udalski, A., et al. 2019, MNRAS, 488, 3308
Littenberg, T. B., Farr, B., Coughlin, S., Kalogera, V., & Holz, D. E. 2015, ApJL, 807, L24
Lu, J. R., Lam, C. Y., Medford, M., Dawson, W., & Golovich, N. 2019, RNAAS, 3, 58
Lu, J. R., Sinukoff, E., Ofek, E. O., Udalski, A., & Kozlowski, S. 2016, ApJ, 830, 41
Luger, R., Agol, E., Foreman-Mackey, D., et al. 2019, AJ, 157, 64
Mao, S., & Paczynski, B. 1996, ApJ, 473, 57
Mao, S., Smith, M. C., Woźniak, P., et al. 2002, MNRAS, 329, 349
Medford, M. S., Lu, J. R., Dawson, W. A., et al. 2020, ApJ, 897, 144
Michel, E., Samadi, R., Baudin, F., et al. 2009, A&A, 495, 979
Mróz, P., Udalski, A., Skowron, J., et al. 2017, Natur, 548, 183
Mróz, P., Udalski, A., Skowron, J., et al. 2019, ApJS, 244, 29
Mróz, P., Udalski, A., Szymański, M. K., et al. 2020, ApJS, 249, 16
Muraki, Y., Sumi, T., Abe, F., et al. 1999, PThPS, 133, 233
Paczynski, B. 1986, ApJ, 304, 1
Paczynski, B. 1996, ARA&A, 34, 419
Park, B. G., DePoy, D. L., Gaudi, B. S., et al. 2004, ApJ, 609, 166
Penny, M. T., Gaudi, B. S., Kerins, E., et al. 2019, ApJS, 241, 3
Poindexter, S., Afonso, C., Bennett, D. P., et al. 2005, ApJ, 633, 914
Poleski, R., & Yee, J. C. 2019, A&C, 26, 35
Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning) (Cambridge, MA: The MIT Press)
Remillard, R. A., & McClintock, J. E. 2006, ARA&A, 44, 49
Renault, C., Afonso, C., Aubourg, E., et al. 1997, A&A, 324, L69
Saha, P. 2000, AJ, 120, 1654
Sajadian, S., & Poleski, R. 2019, ApJ, 871, 205
Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Comput. Sci., 2, e55
Smith, M. C., Belokurov, V., Evans, N. W., Mao, S., & An, J. H. 2005, MNRAS, 361, 128
Smith, M. C., Mao, S., & Paczyński, B. 2003, MNRAS, 339, 925
Smith, M. C., Mao, S., Woźniak, P., et al. 2002, MNRAS, 336, 670
Smith, M. C., Woźniak, P., Mao, S., & Sumi, T. 2007, MNRAS, 380, 805
Sumi, T., Kamiya, K., Bennett, D. P., et al. 2011, Natur, 473, 349
Theano Development Team 2016, arXiv:1605.02688
Tisserand, P., Le Guillou, L., Afonso, C., et al. 2007, A&A, 469, 387
Udalski, A. 2003, AcA, 53, 291
Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M., & Mateo, M. 1992, AcA, 42, 253
Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Wegg, C., Gerhard, O., & Portail, M. 2016, MNRAS, 463, 557
Wegg, C., Gerhard, O., & Portail, M. 2017, ApJL, 843, L5
Woźniak, P., & Paczyński, B. 1997, ApJ, 487, 55
Wyrzykowski, Ł., Kostrzewa-Rutkowska, Z., Skowron, J., et al. 2016, MNRAS, 458, 3012
Wyrzykowski, Ł., & Mandel, I. 2020, A&A, 636, A20
Wyrzykowski, Ł., Rynkiewicz, A. E., Skowron, J., et al. 2015, ApJS, 216, 12
Wyrzykowski, L., Udalski, A., Mao, S., et al. 2006, AcA, 56, 145
Yee, J. C., Shvartzvald, Y., Gal-Yam, A., et al. 2012, ApJ, 755, 102
Yee, J. C., Udalski, A., Calchi Novati, S., et al. 2015, ApJ, 802, 76