Spatially resolved microlensing timescale distributions across the Galactic bulge with the VVV survey

Zofia Kaczmarek1,2⋆, Peter McGill3, N. Wyn Evans1, Leigh C. Smith1, Nathan Golovich3, Eamonn Kerins4, David Specht4, William A. Dawson3

1Institute of Astronomy, University of Cambridge, Madingley Rd, Cambridge CB3 0HA, UK
2Zentrum für Astronomie der Universität Heidelberg, Astronomisches Rechen-Institut, Mönchhofstr. 12-14, 69120 Heidelberg, Germany
3Space Science Institute, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA
4Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Accepted XXX. Received YYY; in original form ZZZ.

ABSTRACT

We analyze 1602 microlensing events found in the VISTA Variables in the Via Lactea (VVV) near-infrared (NIR) survey data. We obtain spatially-resolved, efficiency-corrected timescale distributions across the Galactic bulge (|ℓ| < 10°, |b| < 5°), using a Bayesian hierarchical model. Spatially-resolved peaks and means of the timescale distributions, along with their marginal distributions in strips of longitude and latitude, are in agreement at a 1σ level with predictions based on the Besançon model of the Galaxy. We find that the event timescales in the central bulge fields (|ℓ| < 5°) are on average shorter than the non-central (|ℓ| > 5°) fields, with the average peak of the lognormal timescale distribution at 23.6 ± 1.9 days for the central fields and 29.0 ± 3.0 days for the non-central fields. Our ability to probe the structure of the Bulge with this sample of NIR microlensing events is limited by the VVV survey’s sparse cadence and relatively small number of detected microlensing events compared to dedicated optical surveys. Looking forward to future surveys, we investigate the capability of the Roman telescope to detect spatially-resolved asymmetries in the timescale distributions. We propose two pairs of Roman fields, centred on (ℓ = ±9.5°, b = −0.125°) and (ℓ = ±5°, b = ±1.375°) as good targets to measure the asymmetry in longitude and latitude, respectively.

Key words: gravitational lensing: micro – Galaxy: bulge – Galaxy: structure

1 INTRODUCTION

Griest et al. (1991) and Paczynski (1991) first proposed monitoring the Milky Way bulge for microlensing events, partly as a reliability check on the results towards the Magellanic Clouds (Alard et al. 1995; Evans 2003). The early measurements of the photometric microlensing optical depth (or the probability that a microlensing event occurs) towards the bulge turned out to be higher than expected (Udalski et al. 1994; Alcock et al. 1995). Exploiting this high rate, modern microlensing surveys of the Milky Way bulge provide the largest source of microlensing events. They detect about 2000 photometric events per year (Udalski et al. 2015; Kim et al. 2016), which have proved useful for a range of different astrophysical applications (Gould 2001; Mao 2012). Most notably, short timescale deviations, caused by planets around the lens, have been detected for (e.g., Bond et al. 2004; Udalski et al. 2005; Beaulieu et al. 2006). These events have provided a probe of low-mass planets (down to Earth-mass) at intermediate separations from their hosts (≈ 1 au) which are typically off-limits to other planet detection techniques (Tsapras 2018).

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

\[ t_E = \frac{\theta_E}{\mu_{rel}} = \frac{1}{v_{rel}} \sqrt{\frac{4GM_L D_L (D_S - D_L)}{c^2 D_S}}. \]

This is the time taken for the source to traverse the angular Einstein radius θE. Here, μrel and vrel are the lens-source relative proper motion and velocity, while D_L, D_S and M_L are the lens distance, source distance and lens mass, respectively. While t_E contains information about the lens-mass and lens-source relative distance and velocities, it is completely degenerate. This means that for a single event, t_E contains limited information about the lens. However, measurements of t_E for large samples of events can tell us about the characteristics of the population of lenses causing them and hence probe the structure of the Galaxy (e.g., Kiraga & Paczynski 1994; Evans & Belokurov 2002; Han & Gould 2003; Kerins et al. 2009; Perkins et al. 2024).

While current monitoring surveys of the Galactic bulge like OGLE-IV (Udalski et al. 2015) and KMTNet (Kim et al. 2016) have yielded thousands of photometric microlensing events, they are restricted to optical wavelengths. This constrains them to high galactic latitudes (|b| > 2°) around the bulge in windows where extinction from interstellar dust is low. Of course, this does not mean there are no microlensing events in the inner bulge regions. In fact, event rates in the innermost regions of the bulge are expected to be high (e.g.,...
Evans & Belokurov 2002; Specht et al. 2020), they are just not bright enough to be detected at optical wavelengths due to extinction.

Gould (1994) noted that this issue could be overcome by monitoring the inner bulge regions in the Near-Infrared (NIR) wavelengths. NIR photometry is less affected by extinction and provides a means to look through the interstellar dust. The United Kingdom Infrared Telescope (UKIRT) microlensing survey (Shvartzvald et al. 2017) was the first to survey the inner regions of the bulge, examining \( \approx 10 \) square degree patch between J2015-J2016 close to the Galactic plane (\( b \sim 0 \)). In addition to supporting space-based microlensing campaigns by both the Kepler and Spitzer space telescopes (e.g. Henderson et al. 2016), the UKIRT microlensing survey reported five highly extinct events missed by optical surveys demonstrating the advantage of observing in the NIR. Recent analysis of the UKIRT microlensing survey revealed 522 clear events and a lower bound on the event rate increasing towards the bulge (Wen et al. 2023). Subsequently, efforts to search regions of the inner bulge in the NIR with the Vista Variables in the Via Lactea (VVV) survey (Minniti et al. 2010) were undertaken (e.g., Navarro et al. 2017, 2018, 2020a,b).

The VVV survey has been monitoring \( \approx 560 \) square degrees of sky centred on the Galactic bulge and inner disc in the NIR \( K_s \)-band for nearly a decade. While the VVV may seem ideal for finding microlensing events covering the region within \( 0 < \ell < 182 \) and \( 0 < \delta < -46^\circ \), it may be less suitable for finding microlensing events towards the inner region of the bulge in terms of its wavelength and spatial coverage, as it is not a dedicated microlensing survey. Consequently, the adopted sparse and spatially varying cadences makes detecting and characterizing microlensing events difficult.

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

Building on this work, Husseiniova et al. (2021) built a scalable machine learning classifier to extract microlensing events from the sparse VVV photometry (Smith et al. 2018). They extracted 1959 microlensing events across the entire 560 deg\(^2\) VVV survey footprint. Husseiniova et al. (2021) were also able to overcome the problem of the 'forsaken' events by using nested sampling (Skilling et al. 2006; Higson et al. 2019; Speagle 2020) to fully characterise the event’s multi-modal posterior distributions. Husseiniova et al. (2021) also characterized the spatial dependence on recovery efficiency of their classifier. So, they also were able to compute spatially resolved completeness maps as a function of Einstein crossing time over the entire VVV footprint. Kaczmarek et al. (2022) subsequently extended this work by incorporating parallax effects into the microlensing models and studied events likely to be caused by dark remnant lenses. In addition to finding photometric microlensing events, there have also been efforts to predict astrometric microlensing events (e.g., Zurlo et al. 2018; Bramich 2018; McGill et al. 2020, 2023) using NIR data for brown dwarfs lenses (Nielsen & Bramich 2018; Luberto et al. 2022) and for sources in the Galactic Bulge (McGill et al. 2019).

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

All current tests of the MaBILS-2 have been in the optical wavelengths (Sumi & Penny 2016; Mróz et al. 2019) and therefore restricted to low-extinction windows at high galactic latitudes (\( 1^\circ < |b| < 6^\circ \), Specht et al. 2020). Even though Navarro et al. (2018) and Navarro et al. (2020a) extracted microlensing events along strips of longitude and latitude in the NIR and covered regions of \( |b| < 1^\circ \), detailed comparisons with Galactic models were not carried out. Regarding the latitudinal strip, Navarro et al. (2020a) reported an observed decrease in events with increasing latitude and a shorter mean timescale for events found closer to the Galactic plane. For the longitudinal strip, Navarro et al. (2018) noted a smooth increase in the number of events towards the Galactic centre, with a slight asymmetry towards more events at negative latitudes. While the work of Navarro et al. (2018, 2020a) has demonstrated the power of the VVV for microlensing, work utilizing the full survey is the clear next step. In this paper, we use the work of Husseiniova et al. (2021) and Kaczmarek et al. (2022) to examine the spatial dependence of the microlensing timescale distribution in the NIR across the Galactic Bulge.

2 METHODS

2.1 Data

We extract the subsample of 1602 microlensing events limited to the Galactic bulge (\( |l| < 10^\circ \), \( |b| < 5^\circ \)) from the 1959 total events found by Husseiniova et al. (2021) in version 2 of the VVV infrared astrometric catalog (VirAC; see Smith et al. 2018, for version 1). This catalog provides \( K_s \)-band time series photometry for \( < 10^3 \) sources covering \( \approx 560 \text{ deg}^2 \) of the Galactic bulge and plane. The VirAC version 2 catalog was processed with a modified version of DoPHOT (Schechter et al. 1993; Alonso-García et al. 2012). The full details of the photometric reduction are described in Smith et al. (in prep). The VVV lightcurves span observations between 2010 and 2019 and have between \( 50 - 500 \) data points.
2.2 Einstein timescale posteriors

We used the posterior distributions obtained by the methods presented in Kaczmarek et al. (2022). Specifically, we used nested sampling (Higson et al. 2019) implemented via dynesty (Speagle 2020), to obtain posterior samples for an annual parallax microlensing model parameterized by the Einstein timescale, the normalized lens-source impact parameter, the source blending parameter, the time of lens-source closest approach, $K_e$-band baseline magnitude, and the microlensing parallax vector components in the north and east directions (Gould 2004), or \( \{t_E, a_0, f_s, l_0, m_0, \pi_{EN}, \pi_{EE} \} \), respectively.

For all of the parameters, flat priors were used which are detailed in Kaczmarek et al. (2022). The dynesty nested sampling settings were random walk sampling (Skilling et al. 2006) with multiple bounding ellipsoids and 1000 initial live points, a stopping criterion in the remaining fractional evidence of 0.01, and 100 percent of weight allocated on computing the posterior distributions. The nested sampling method was chosen due to its ability to sample multi-modal and degenerate posterior distributions, which is important for VVV microlensing events with sparse cadences (Russhenhoorn et al. 2021; Kaczmarek et al. 2022).

2.3 Hierarchical timescale inference

To infer the \( t_E \) distribution for a sample of events, we follow Golovich et al. (2022) and fit flexible step function to the distribution, \( f_\alpha \), within a hierarchical model of the form,

\[
f_\alpha(\log_{10} t_E) = \sum_{m=1}^{M} \alpha_m \frac{ (t_E / M)_{m-1} - (t_E / M)_m }{ t_E / M - (t_E / M)_m } ,
\]

where \( \alpha \) is a vector of length \( M \), with the elements \( \alpha_m \) and,

\[
s(x, L, H) = \begin{cases} 
0 & \text{for } x < L, \\
(H - L)^{-1} & \text{for } L \leq x \leq H, \\
0 & \text{for } x > H
\end{cases}
\]

where we divide the allowed range of event timescales into \( M = 30 \) log-space bins between \( t_E = 1 \) and \( t_E = 10^3 \) days, and specify that,

\[
\sum_{m=1}^{M} \alpha_m = 1,
\]

to ensure that \( f_\alpha \) is a normalized probability distribution. Under this model we can compute the likelihood efficiently and approximately over a sample of microlensing events for a value of \( \alpha \) by using importance sampling and re-weighting the posterior samples for each event obtained by the methods described in Section 2.2. For a sample of \( N_e \) events where \( t_{E,n,k} \) is the \( k \)th posterior sample of the \( n \)th event, the likelihood is (Hogg et al. 2010),

\[
\mathcal{L}(\{d_i\}|\alpha) \approx \prod_{n=1}^{N} \sum_{k=1}^{K} \frac{f_\alpha(t_{E,n,k})}{p_0(\alpha)}.
\]

Here, \( \{d_i\} \) is the set of \( N_e \) microlensing lightcurves and \( p_0 \) is the prior density from Table 1. in Kaczmarek et al. (2022). We place a Dirichlet prior on \( \alpha \) to enforce the normalization condition in Eq. (4),

\[
p(\alpha) \sim \text{Dir}(\alpha|\alpha), \text{ where: } \alpha = (w - \min(w))b + c
\]

Here, following Golovich et al. (2022), we set the concentration hyper-parameters, \( \alpha_u \), using the weights of the maximum a posteriori (MAP) \( t_E \) histogram \( w \) and we fix \( b = 20, c = 0.2 \). We then sample the posterior distribution of \( \alpha \) using the No U-Turn Sampler (NUTS) algorithm (Hoffman & Gelman 2011) implemented in PyMC3 (Salvatier et al. 2015) using 1600 burn-in steps and then a total of 32000 steps for inference.
2.4 Choice of metric

Throughout this work we operate in the log $t_E$ scale. Following Wyrzykowski et al. (2015), we fit the event timescale distribution with a lognormal model; we use the mean of this lognormal distribution (which we denote as $\langle t_E \rangle$) as a metric representing a typical event timescale for a given bin (e.g. Fig 1). This metric is more robust than the mean timescale $\langle t_E \rangle$, as the log $t_E$ distribution is roughly symmetric; the peak of this distribution should be minimally sensitive to outliers. This is especially important for the part of this work concerning VVV data, where outer wings of the $t_E$ distribution are dominated by noise. We describe the fitting procedure in more detail in Section 3.2.

2.5 The Manchester-Besançon Microlensing Simulator (MaBµs-2)

To compare our microlensing timescale distributions to models of the Galaxy we use the second version of the Manchester-Besançon Microlensing Simulator (MaBµs-2; Specht et al. 2020), which is the current state-of-the-art tool for creating maps of microlensing event population observables – optical depth, event rate and timescale distributions – along different lines of sight towards the Galactic Bulge. MaBµs-2 is publicly available and covers a $20^\circ \times 20^\circ$ region around the Galactic Centre and provides simulations for several optical and NIR bands. Based on the Besançon stellar population synthesis model (Robin et al. 2003, 2012), MaBµs-2 simulations include thin and thick disk, bar and halo components which all have individually defined density profiles, initial mass functions, star formation rates, kinematics and stellar ages (see Section 2 of Specht et al. 2020). Additionally, interstellar dust is accounted for in the simulations using a 3D Galactic extinction model (Marshall et al. 2006).

MaBµs and the underlying Besançon Galactic model have been iteratively improved when reconciled with data. For example, analysis of microlensing events from MOA-II (Sumi et al. 2013) led to additional populations of faint, low mass M dwarfs and brown dwarfs being added to the simulations (Awiphan et al. 2016; Sumi & Penny 2016). Most recently, motivated by upcoming large scale time-domain surveys such as the Roman Space Telescope (Spiegel et al. 2015) and the Vera C. Rubin Observatory (Ivezic et al. 2019), MaBµs-2 includes the simulation of higher-order microlensing effects such as finite sources, unresolved stellar backgrounds and is in good agreement with the largest samples of $\sim 10^3$ optical microlensing events from OGLE-IV (Udalski et al. 2015; Specht et al. 2020) and stellar kinematics observed by the Hubble Space Telescope towards the bulge.

2.6 Spatial mapping

We cover the region of interest around the Galactic Centre ($|l| < 10^\circ$, $|b| < 5^\circ$) with 12 rectangular bins. Those include 4 narrow bins around the Galactic plane ($-1^\circ < b < 1^\circ$, uniform $5^\circ$ steps in longitude between $-10^\circ$ and $10^\circ$), and 8 wider bins on either side ($-5^\circ < b < -1^\circ$ and $1^\circ < b < 5^\circ$, uniform $5^\circ$ steps in longitude between $-10^\circ$ and $10^\circ$). This choice of binning is motivated by the spatial variation of event timescales expected from Galactic models, as illustrated in Fig 1 and Table 1, respectively. In selecting our bins, we aim to both capture the most dramatic features of the event timescale map predicted with MaBµs-2 (e.g. ridge at $b \approx 0^\circ$, sharp rise in timescales around $l \approx \pm 7.5^\circ$), and test whether we can detect any asymmetry in the Bulge between positive and negative longitudes. The bin selection was also optimised to contain a sufficient number of events in every subsample.

2.7 Efficiency correction

The efficiency of recovering events with various timescales is imposed by the VVV observing pattern and cadence and is highly heterogeneous across the survey area (see Figure 8 in Husseiniova et al. 2021). After obtaining hierarchical $t_E$ distributions for each of the bins, we need to correct them for this efficiency; neglecting this step would introduce additional bias being solely the consequence of the particular VVV observing program used.

To generate efficiency correction maps, we adopt an approach similar to the one described in Section 4.3 of Husseiniova et al. (2021). The area around the Galactic Bulge is divided into rectangular bins, and microlensing events are simulated in each bin following a point source – point lens model assuming linear relative motion (no parallax effect). A source ‘seed’ – effectively, the position of the event, its observation times and source star baseline magnitude – is selected randomly using real light curves from within the area. The timescale $t_E$, for which we are determining the recovery efficiency, is fixed. The remaining model parameters $u_0$, $f_s$, $t_0$ are selected randomly from distributions described in Table 2. The simulated light curves are then passed through the automatic classifier developed in Husseiniova et al. (2021), using a threshold probability of 0.825 required for event recovery, until the recovery efficiency reaches the signal-to-noise ratio of 10.

The main difference is that our efficiency map generation adopts a higher resolution in $t_E$, but a lower spatial resolution. Adopting a coarse grid ($2^\circ \times 5^\circ$ to $4^\circ \times 5^\circ$ regions) matching the spatial bins makes the process significantly less computationally expensive; within each of the bins, the simulation was run for each of the 30 midpoints of $t_E$ bins used in the hierarchical $t_E$ distribution inference. This allows for straightforward correction of the height of each $t_E$ bin by its corresponding recovery efficiency in the spatial bin.

We find that the efficiency curves between our spatial bins are significantly different. This is mostly the effect of the specific VVV observing pattern (see Figure 7 in Husseiniova et al. 2021). The differences in efficiency curves mostly reflect two features of this pattern: longer survey coverage (higher standard deviation in observation date) at low latitudes, and a small window around $l = 5^\circ$, $b = 2, 5^\circ$ having a distinctly higher cadence. We show the spatial efficiency variations for two sample timescales in the middle and right panels of Fig 1. A small, though notable, asymmetry in latitude is also reflected in the efficiency curves, especially in the wing corresponding to longer events, as illustrated in Fig 2.

<table>
<thead>
<tr>
<th>parameter</th>
<th>PDF</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_0$</td>
<td>uniform($10^{-3}$, 1.5)</td>
<td>$t_E$</td>
</tr>
<tr>
<td>$f_s$</td>
<td>uniform(0, 1)</td>
<td>-</td>
</tr>
<tr>
<td>$t_0$</td>
<td>uniform(2019, 2021)</td>
<td>Julian years</td>
</tr>
</tbody>
</table>

Table 2. Model parameter probability distributions used for generating random events in the process of simulating microlensing event recovery efficiency in the VVV.
2.8 Limitations of the efficiency correction

Our efficiency correction is subject to several limitations:

(i) At the stage of creating mock events for the efficiency correction, we assume no parallax signal ($\pi_E = 0$); whereas for the hierarchical inference, to obtain maximally precise timescale measurements, we use samples from modelling that includes parallax effect. Kaczmarek et al. (2022) note that 176 out of 1959 modelled events (9.0%) had strong evidence for parallax signal (at the level of Bayes factor $K > 100$). The majority of the events with strong parallax signal have very high inferred event timescales ($t_E > 400$ days), but much shorter inferred timescales from the no-parallax modelling ($t_E < 400$ days). For such an event, its detection efficiency would be underestimated (to be rightward of the black dashed line in Fig. 2) and hence, the corresponding bins disproportionally upweighted. As it is not feasible to disentangle contributions of single events to the hierarchical inference, we also cannot amend the efficiency correction on a per-event basis. We consider this the main shortcoming of the efficiency correction model.

After analysing this subset of 176 events in more detail, we typically find very high uncertainties in timescale and parallax fits, as well as lightcurves with sparse cadence and only a single slope of the event covered. We conclude a significant fraction of those objects could be contaminants, mainly long-period variables whose other peaks are missed due to sparse cadence; however their nature (microlensing or intrinsic variable) cannot be unambiguously determined with the available data. This also motivates more strict masking of the long-timescale wing of the $t_E$ distribution (see Subsection 3.2). The contamination problem should be, to a large degree, alleviated with next-generation high-cadence NIR observations.

(ii) At the stage of creating mock events, we assume a uniform distribution of the blending parameter $f_\delta$ (following Husseiniova et al. 2021). The $f_\delta$ distribution in the $K_s$ band is not known as of today due to insufficient data; we expect this obstacle to be entirely eliminated in the near future with data from JWST and Roman Space Telescope. We note the possibility for post-correction of our hierarchical inference results once high-resolution images in the NIR are available for determining the $f_\delta$ distribution.

(iii) The final catalogue of 1959 events from Husseiniova et al. (2021) has been filtered with human inspection, which we do not account for in the efficiency correction model. However, Husseiniova et al. (2021) do not note a correlation of completeness of the visual inspection with event timescale (they note a correlation with baseline magnitude and impact parameter). We thus conclude this should not introduce a significant bias.

3 RESULTS

3.1 Obtaining the $t_E$ distributions

For all 12 spatial bins, we obtain the $t_E$ via hierarchical inference (Subsection 2.3), simulate the detection efficiency $\epsilon$ (Subsection 2.7) and correct for it by multiplying every $t_E$ distribution bin height by $1/\epsilon$, and then compare the results to the predictions from MaBulS-2 (Subsection 2.5). Those predictions were obtained using a custom run of the simulator modified to output full $t_E$ distributions – all input and settings are the same as in Specht et al. (2020) and the online simulator tool\(^4\). We interpolate $t_E$ distributions from the simulation in each MaBulS-2 spatial bin (0.25° x 0.25°) to match our $t_E$ binning (30 logspace bins between $t_E = 1$ and $t_E = 10^3$ days), and merge those bins (weighted by event rate) to match our spatial binning. The efficiency-corrected timescale distributions from the hierarchical inference, along with their respective MaBulS-2 predictions, are shown in Fig 3 for all the fine bins.

For clarity, we do not include the hierarchical inference results before the efficiency correction. To give an idea of the corrections for bins with different typical efficiencies, we present the non-corrected distributions from bins 1, 5 and 9 (leftmost column) in Fig A1 and discuss them in Appendix A.

The errorbars in Fig A1 and Fig 3 indicate the 68% confidence intervals on the $t_E$ bin heights. The non-corrected errorbars are taken directly from the samples of $t_E$ bin heights from the hierarchical inference code, while the efficiency-corrected distribution errorbars also include the error on the recovery efficiency value. The extents of those errorbars were calculated with a Monte Carlo simulation, drawing 100,000 efficiency-corrected bin height values per each $t_E$ bin with both a random choice of the bin height from the samples and a random choice of the efficiency correction from a normal distribution $N(\epsilon, \sigma_\epsilon)$. ($\sigma_\epsilon$ is included in the output of the detection efficiency simulator and is determined by the target signal-to-noise ratio criterion: $\sigma_\epsilon \lesssim 0.1 \epsilon$.)

3.2 Lognormal distribution fitting

As discussed in Section 2.4, we fit a lognormal distribution of $t_E$ and use the mean of this distribution, $\overline{t_E}$, as a metric.

At the preliminary stages of data analysis, we find very high noise levels in $t_E$ determination (see Appendix A) and we note that the outer wings of the distribution contain unrealistically large best-estimate bin heights, albeit with very large errorbars. This shape of the $t_E$ distribution is in disagreement with Mao & Paczynski (1996), who argue that the wings of the event timescale distribution should be power laws. Mao & Paczynski (1996) explain that long timescale events are produced by lenses with small transverse velocity (if the

\[ \text{www.mabuls.net} \]
Figure 3. Event timescale distributions for all 12 spatial bins. Purple, solid line/filled circles: efficiency-corrected $t_E$ bin heights from VVV events obtained using hierarchical inference. Errorbars indicate 68% confidence intervals. Blue, solid line/filled circles: $t_E$ bin heights predicted using the MaBuilS-2 simulation. Dark purple: lognormal distribution fit to datapoints obtained using hierarchical inference. Only the datapoints outside of the mask (light blue, filled) are included in the fit. The dashed line is the fitted curve and the vertical line is the mean ($\hat{t}_E$) of the distribution. Dark blue: as dark magenta, but fitted to MaBuilS-2 datapoints. Orange: median (dotted line) and 68% confidence intervals (light orange band) for the means ($\hat{t}_{E,i}$) of the lognormal distributions fitted to each histogram $y_i(\log t_E)$ from 10000 MC samples.
Figure 4. Timescales $\hat{t}_E$ corresponding to the means of the fitted lognormal timescale distribution for each bin. Top left: fit to MaBμlS-2 simulation predictions. Top right: fit to hierarchical inference results. Bottom left: Difference between predicted and inferred $\hat{t}_E$. Bottom right: Half width of the 1σ confidence interval for determining $\hat{t}_E$ from hierarchical inference results (see Fig 3). The dimensions match those of Fig 1; only regions covered by the bins included.

Figure 5. Event timescale distributions. As Fig 3, plotted per each longitude spatial bin; legend as in Fig 3.

transverse component is significantly lower than the velocity dispersion for the three-dimensional Gaussian velocity distribution, they derive the number of events to be $N \sim t^{-3}$). On the other slope, the short events are produced by the lenses close to the source or the observer; by expressing the event timescale as a function of lens distance, they arrive at $N \sim t^3$.

We find several possible causes for this effect. Firstly, the outer wings of the distribution from the hierarchical inference typically flatten out smoothly, yielding larger bin height values than histograms from raw samples or simulation predictions (compare Figures 11, 12 in Golovich et al. 2022). Secondly, the detection efficiency simulated for the VVV observing strategy is very low for the lowest and highest $t_E$ values, resulting in significant upweighing during efficiency correction. This means any overestimation in the outer wings of $t_E$ distribution will be highly amplified.

This effect is significantly stronger, and the slope steeper, for the long $t_E$ end, despite the efficiency detection being higher there. With visual inspection, we find that a large fraction of very long events ($t_E > 400$ days) appear to be ambiguous and could be consistent with both microlensing and intrinsic variables (see also Section 2.8).

In order to estimate the degree of ambiguity with minimal bias, we extract quality grades given by the authors in Husseiniova et al.
We fit a lognormal distribution using scipy.optimize.curve_fit (Virtanen et al. 2020) to the unmasked inference results from VVV data and the MaBµlS-2 predictions. We present the fits alongside individual distributions in Fig 3, and we compare the $t_E$ values per bin on a colour map in Fig 4. The systematic differences between predicted and inferred $t_E$ seen in those results are discussed in detail in Subsection 3.4.

3.3 Confidence intervals

To evaluate the precision of our $t_E$ estimation from data, we run a Monte Carlo simulation. For each spatial bin we apply the following procedure 10000 times. We draw a set of efficiency-corrected $t_E$ bin heights randomly out of the Monte Carlo samples obtained in Subsection 3.1, generating a single realisation (with index $i$) of a $t_E$ histogram: $y_i(\log t_E)$. For this realisation, we fit a lognormal distribution and save its mean $\log \hat{t}_E_i$.

In Fig 3, we plot the median values of $\hat{t}_E_i$ and their 68% confidence intervals, along with 1000 randomly chosen MC samples $y_i$, for each bin. We also plot the $t_E$ values for both our inference and MaBµlS-2 with vertical lines. We conclude that our inference results are consistent with MaBµlS-2 predictions within 1σ, except for bin 2 (21.9 days) which is just outside the 1 sigma band (22.1 - 34.2 days). We plot the half width of the 1σ confidence interval as part of Fig 4 (bottom right). This width is very highly correlated with event count per bin (see Table 1).

3.4 Trends in longitude and latitude

In addition to the fine binning, we then also repeat all steps outlined in Subsection 3.1 for the merged longitude-only binning ($5° > b > −5°$ and, respectively: $10° > ℓ > 5°$, $5° > ℓ > 0°$, $0° > ℓ > −5°$, $−5° > ℓ > −10°$) and latitude-only binning ($10° > ℓ > −10°$ and, respectively: $5° > b > 1°$, $|b| < 1°$, $−5° < b < −1°$) in order to maximise the signal-to-noise ratio and specifically search for asymmetry. The timescale distributions, analogous to Fig 3, are shown in Figs 5 and 6 for the longitude and latitude bins, respectively.

We note that the errorbars on $t_E$ shrink slightly when adapting the coarser binning, with the mean difference between the upper and lower bounds of the 68-percentile extent of all MC samples equal to 12.4 days for the fine binning, 11.1 days for the longitude-only binning and 8.9 days for the latitude-only binning. The differences between bins are still not statistically significant with the coarser binning.

To analyse the trends reflected in our estimates and biases influencing them, we plot the estimated peaks of the timescale distribution $\hat{t}_E$ as a function of longitude only and latitude only in Fig 7. As compared to MaBµlS-2 predictions, we reproduce the trend of lower timescales in central longitudes ($|ℓ| < 5°$) and higher timescales in non-central longitudes ($|ℓ| > 5°$), but we do not reproduce the trend of timescales increasing with decreasing latitude. Our estimates are slightly biased towards longer timescales – averaging the differences weighted by $1/\sigma^2$, we obtain a bias of +0.7 days for the fine binning (+1.9 days for longitude binning and +0.4 days for latitude binning, respectively). The differences are higher for higher latitudes and higher longitudes.

With inverse variance weighing over the fine binning, we obtain heights. After applying this normalisation, almost all the non-masked bin heights agree within 1σ with the MaBµlS-2 predictions, though there are a few outliers.
an average peak of the lognormal timescale distribution at $23.6 \pm 1.9$ days for the central bins and $29.0 \pm 3.0$ days for the non-central bins. This is in good agreement with the MaBulS-2 predictions of 22.4 days and 29.0 days (averaged with equal weights per bin), respectively.

### 4 DISCUSSION AND CONCLUSIONS

We analyzed a sample of 1602 microlensing events found in the VVV data and obtained spatially-resolved timescale distributions across the Galactic bulge, using hierarchical inference with all nested-sampling posterior samples from modelling the events as input. We then compared those timescale distributions with MaBulS-2 predictions. Therefore, we provide a new test of the MaBulS-2 simulation and, independent of previous studies. We find very good agreement with MaBulS-2 predictions. From the experiments visualised in Figs 3-7, we conclude that the results of our hierarchical inference almost always agree with MaBulS-2 predictions at a $1\sigma$ level - both for individual $t_E$ bin heights and for the estimated mean log $t_E$ values.

At the same time, we note that the $1\sigma$ confidence intervals for this estimation are large compared to differences between pairs of spatial bins, and hence these differences are not statistically significant within our inference (for neither individual $t_E$ bin heights nor the estimated mean log $t_E$ values). We note the very strong dependence of $1\sigma$ interval widths on event count. We conclude that more observational data, both in terms of a higher event count and higher number of datapoints per event, is needed in order to increase the precision of $t_E$ distribution inference.

We now use the spatially resolved timescale distributions from the MaBulS-2 simulation, created for this work, to analyse differences between fields in the central Milky Way regions and propose optimal observing strategies to capture their asymmetry.

We first identify the optimal fields to capture significant difference between $t_E$ distributions. In Fig 8, we compare the MaBulS-2 predictions for all spatial bins. We note that the most striking feature is the splitting of the timescale distributions into two groups: the central ($|\ell| < 5^\circ$) and non-central ($|\ell| > 5^\circ$) longitudes, with the latter tending towards larger timescales. While all distributions for the central bins are very similar, careful analysis of the lower $t_E$ edge would allow to identify difference between the low $|\ell|$ dotted in Fig 8 and high $|\ell|$ non-central regions (dotted in Fig 8) – that would require a very high-cadence (<1 day) survey. We note that among pairs of neighbouring bins, the highest difference in $t_E$ is between bins 7 and 8 ($\Delta t_E = 9.1$ days). We recommend investigating the region covered by those bins ($|b| < 1^\circ$, $0^\circ > \ell > -10^\circ$) to capture the spatial gradients in the $t_E$ distribution confidently. Among pairs of non-neighbouring bins, the highest differences in $t_E$ are between bins 3 and 8 ($\Delta t_E = 9.5$ days) and bins 2 and 8 ($\Delta t_E = 9.4$ days).

We then identify the most optimal fields to capture asymmetry. This is useful in constraining the properties of asymmetric components of the Milky Way, especially the Galactic bar. While direct observations of stars get less complete and more confined to the highest absolute magnitudes with increasing distance, microlensing remains unconstrained by the brightness of the lens and is an ideal tool to study diverse stellar populations and structure in the central Galactic regions (e.g., Kiraga & Paczynski 1994; Häfner et al. 2000; Bissantz & Gerhard 2002; Rattenbury et al. 2007).

We split the $t_E$ distribution from Fig 1 into a symmetric and asymmetric (in the respective coordinate) component:

$$(t_E)_{\text{sym, } b}(\ell, b) = \frac{t_E(\ell, b) - (t_E)_{\text{asy, } b}(\ell, b)}{2}$$

and analogously,

$$(t_E)_{\text{asy, } b}(\ell, b) = \frac{t_E(\ell, b) + (t_E)_{\text{sym, } b}(\ell, b)}{2}.$$

We plot the asymmetric components in Fig 9. The $t_E$ asymmetry in latitude and longitude is at similar levels, with maximum absolute
values of 2.4 and 2.8 days and median absolute values of 0.5 and 0.3 days, respectively.

We propose that with a high-cadence near-infrared survey across fields with high $t_E$ differences, asymmetry could be captured at a statistically significant level. A suite of simulations similar to MaBulS-2 could be performed with varying Galactic bar properties (e.g. bar strength, bar angle), and those properties could be constrained by testing consistency with observed $t_E$ distributions. This experiment is not yet possible with VVV data, but near-future near-infrared missions might provide sufficiently large and complete microlensing event databases.

The Nancy Grace Roman Space Telescope is expected to be launched by May 2027. Its multi-band camera, WFI (Wide-Field Instrument), will observe in the near infrared, covering wavelengths up to 2300 nm — similarly to VVV’s $K_s$ band. It is scheduled to carry out several surveys, of which the Galactic Bulge Time Domain Survey is of particular interest for microlensing. The survey will provide unprecedented cadence (of 15 minutes within 72-day seasons) and precision (astrometric: $\sim$ 1 mas, photometric: $\sim$ 10 milli-mag for a single observation – increased when combining observations) in surveying the Galactic Bulge (following Penny et al. 2019; WFIRST Astrometry Working Group et al. 2019; Gaudi et al. 2019).

Kaczmarek et al. (2022) have re-simulated one of the found VVV lensing events as seen by Roman, finding $\sim$0.1% precision in event timescale determination (compared to $\sim$10% precision from VVV data for the same event), as well as $\ll$1% precision in subsequent mass measurement. Sajadian & Sahu (2023) have also provided predictions of mass and timescales measurements of isolated stellar-mass black holes by Roman up to $\lesssim$1%. This level of precision will dramatically improve not only studies of single lensing objects, but also statistical analyses of the structure of the Galaxy.

As Roman will observe with a very high cadence and reach magnitudes as faint as $\sim$23 for a single exposure\(^5\), the event yield is also expected to increase very significantly. Particularly, Penny et al. (2019) simulated the microlensing event yield of $\sim$27,000 events\(^6\).

\(^6\) https://roman.gsfc.nasa.gov/science/WFI_technical.html
at impact parameter $|u| < 1$ and $|u| > 3$, distributed over 1.96 deg$^2$ of the field surveyed in Roman’s Galactic Bulge Time Domain Survey (compared to the microlensing event yield of 1959 for the VVV search of Husseiniova et al. (2021), and 1602 events distributed over 200 deg$^2$ used in this study). The high event yield, combined with high precision in $t_e$ determination, will allow for carrying out a study similar to this, but with much narrower confidence intervals for the spatially resolved $t_e$ distributions.

Efforts are underway to optimise the Roman observing strategy even further, with e.g. Sajadian & Sahu (2023) demonstrating the expected wide timescale distribution of detected events, assuming additional sparse observations in the mid-survey gap. Additionally, strategies are proposed for filling the gaps in Roman observations using other surveys, such as The Vera C. Rubin Observatory Legacy Survey of Space and Time (Gezari et al. 2022) or PRIME$^7$.

We overlay the proposed Roman observing fields of the Galactic Bulge Time Domain Survey (Penny et al. 2019)$^8$ on the asymmetry maps in Fig 9. We note that the asymmetry and overall $t_e$ variation is very small over the proposed fields. We set out to determine fields for a hypothetical community survey to capture the strongest asymmetry signal. We choose 2 pairs of Roman WFI fields, centred on ($\ell = +9.5^\circ$, $b = -0.125^\circ$) and ($\ell = -5^\circ$, $b = +1.375^\circ$) to capture asymmetry in longitude and latitude, respectively. To select those fields, we approximate the Roman fields by a $0.5^\circ$ x $0.75^\circ$ rectangle, i.e. 2x3 MaBJS-2 pixels. Per each simulated field, we bin the respective 6 MaBJS-2 pixels together (weighted by event count) and fit $t_e$ to the resulting distribution. We iterate over the extent covered by MaBJS-2 in 0.25$^\circ$ intervals and we choose fields yielding the largest asymmetry. We also overlay those fields on the maps in Fig 9. We plan to further explore an optimal observing strategy, taking into account the characteristics of the Roman Space Telescope and its surveys, in future work.

In summary, we have developed a rigorous framework to analyse spatially resolved microlensing timescale distributions. This is a particularly promising avenue in studies of Galactic structure. Although the VVV survey has insufficient cadence to constrain properties of the Galactic bar, we expect forthcoming opportunities with Roman to make this eminently feasible.

ACKNOWLEDGMENTS

ZK thanks EC H2020 ORP grant no 101004719 for travel support and the Cambridge Streams group for lively discussions. ZK is a Fellow of the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD). This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The document number is LLNL-JRNL-854902. This work was supported by the LLNL-LDRD Program under Project 22-ERD-037. We thank the anonymous referee for valuable comments that improved the quality of this work.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Gaudi B. S., et al., 2019, BAAS, 51, 211
Mao S., 2012, Research in Astronomy and Astrophysics, 12, 947
Minniti D., et al., 2010, New Astron., 15, 433

$^7$ http://www-ir.ess.sci.osaka-u.ac.jp/prime/science/science.html
$^8$ using the code available at https://github.com/mtpenny/wfirst-ml-figures
In this section, we discuss the excess bin heights occurring in the outer bins. We show how the convex shape of resulting timescale distributions originates at the stage of efficiency correction; we then show how those distributions dominated by the outer wings give unreliable $⟨t_E⟩$ estimates, which motivates our choice of fitting the peak of the timescale distribution instead.

In Fig A1 we present the timescale distributions for 3 example bins before and after the efficiency correction. We note the efficiency correction heavily upweights the outer wings of the $t_E$ distribution, making the distribution convex.

In Fig A2 we compare different metrics with their respective confidence intervals by drawing 10000 random samples representing the histogram $y_i(\log t_E)$. We show results for example bins 9 and 12; we note that while the bins are symmetrical and their timescale estimates should be similar, the $⟨t_E⟩$ and unmasked $t_E$ estimates are very poorly constrained and dominated by the outer wings of the distribution, motivating our final choice of metric.
Figure A1. Timescale distributions from the hierarchical $t_E$ inference with (purple) and without (black) the efficiency corrections. All $t_E$ distributions are normalised to 1. The errorbars indicate the 68% confidence intervals on the $t_E$ bin heights. Respective MaBuLS-2 predictions (blue) overplotted for comparison. Left: Timescale distributions for spatial fine bin 1. Centre: Timescale distributions for spatial fine bin 5. Right: Timescale distributions for spatial fine bin 9.

Figure A2. Comparison of different metrics to represent a typical event timescale and their sensitivity to noise for two example bins. Cyan: $\langle t_E \rangle$ – average over the full extent. Teal: mean of the lognormal distribution $\hat{t}_E$ fitted to full extent. Orange: mean of the fitted lognormal distribution $\hat{t}_E$ fitted to the unmasked extent. Event timescale distribution, mask and $\hat{t}_E$ values for the MaBuLS-2 predictions and hierarchical inference results as in Fig 3. Light purple represents 1000 randomly chosen samples $y_i (\log t_E)$. (Samples were re-drawn in case of runtime error (fit failed in 5000 iterations); this happened for 39.8% (bin 9) and 20.5% (bin 12) for the unmasked $\hat{t}_E$ fits, and for 1.2% and 1.9% for the masked $\hat{t}_E$ fits, respectively, and can result in underestimated errorbars especially for the unmasked $\hat{t}_E$ part.)