

# First semi-empirical test of the white dwarf mass-radius relationship using a single white dwarf via astrometric microlensing

Peter McGill<sup>1,2\*</sup>, Jay Anderson<sup>3</sup>, Stefano Casertano<sup>3</sup>, Kailash C. Sahu<sup>3</sup>, Pierre Bergeron<sup>4</sup>, Simon Blouin<sup>5</sup>, Patrick Dufour<sup>4</sup>, Leigh C. Smith<sup>1</sup>, N. Wyn Evans<sup>1</sup>, Vasily Belokurov<sup>1</sup>, Richard L. Smart<sup>6</sup>, Andrea Bellini<sup>3</sup>, Annalisa Calamida<sup>3</sup>, Martin Dominik<sup>7</sup>, Noé Kains<sup>3</sup>, Jonas Klüter<sup>8</sup>, Martin Bo Nielsen<sup>9,10,11</sup>, and Joachim Wambsganss<sup>12</sup>

<sup>1</sup>*Institute of Astronomy, University of Cambridge, Madingley Rd, Cambridge CB3 0HA, UK*

<sup>2</sup>*Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA*

<sup>3</sup>*Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA*

<sup>4</sup>*Département de Physique, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, Québec H3C 3J7, Canada*

<sup>5</sup>*Department of Physics and Astronomy, University of Victoria, Victoria, BC V8W 2Y2, Canada*

<sup>6</sup>*INAF – Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy*

<sup>7</sup>*University of St Andrews, Centre for Exoplanet Science, SUPA School of Physics & Astronomy, North Haugh, St Andrews, KY16 9SS, UK*

<sup>8</sup>*Department of Physics and Astronomy, Louisiana State University, 202 Nicholson Hall, Baton Rouge, LA 70803, USA*

<sup>9</sup>*School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK*

<sup>10</sup>*Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*

<sup>11</sup>*Center for Space Science, NYUAD Institute, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates*

<sup>12</sup>*Astronomisches Rechen-Institut (ARI), Zentrum fuer Astronomie der Universitaet Heidelberg (ZAH), 69120 Heidelberg, Germany*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

In November 2019 the nearby single, isolated DQ-type white dwarf LAWD 37 (WD 1142-645) aligned closely with a distant background source and caused an astrometric microlensing event. Leveraging astrometry from *Gaia* and followup data from the *Hubble Space Telescope* we measure the astrometric deflection of the background source and obtain a gravitational mass for LAWD 37. The main challenge of this analysis is in extracting the lensing signal of the faint background source whilst it is buried in the wings of LAWD 37's point spread function. Removal of LAWD 37's point spread function induces a significant amount of correlated noise which we find can mimic the astrometric lensing signal. We find a deflection model including correlated noise caused by the removal of LAWD 37's point spread function best explains the data and yields a mass for LAWD 37 of  $0.56 \pm 0.08 M_{\odot}$ . This mass is in agreement with the theoretical mass-radius relationship and cooling tracks expected for CO core white dwarfs. Furthermore, the mass is consistent with no or trace amounts of hydrogen that is expected for objects with helium-rich atmospheres like LAWD 37. We conclude that further astrometric followup data on the source is likely to improve the inference on LAWD 37's mass at the  $\approx 3$  percent level and definitively rule out purely correlated noise explanations of the data. This work provides the first semi-empirical test of the white dwarf mass-radius relationship using a single, isolated white dwarf and supports current model atmospheres of DQ white dwarfs and white dwarf evolutionary theory.

**Key words:** gravitational lensing: micro – astrometry – white dwarfs

## 1 INTRODUCTION

Carbon-Oxygen (CO) core white dwarfs are the final evolutionary stage for the vast majority of stars ( $\leq 8 M_{\odot}$ ). They are expected to consist mainly of an electron-degenerate core, surrounded by a thin envelope of non-degenerate hydrogen and helium (e.g., Tremblay & Bergeron 2008). Their mainly degenerate nature means they are expected to follow a mass-radius relationship (MRR) as they evolve and cool. The white-dwarf MRR is important to many areas of astrophysics. It is typically relied upon to calculate the mass of white dwarfs from photometric or spectroscopic measurements (e.g., Fal-

con et al. 2010) and the white dwarf upper mass limit ( $\approx 1.4 M_{\odot}$ ) underpins our understanding of the progenitors of type Ia supernovae. Moreover, it is vital when using cooling white dwarfs to date stellar populations in globular clusters (e.g., Hansen et al. 2002).

To a first-order approximation, in a white dwarf, the inward force of gravity is balanced by the outward pressure of the electron-degenerate gas, resulting in a white dwarf's radius being inversely proportional to the cube root of its mass (Chandrasekhar 1935). Today, detailed evolutionary cooling models including specific degenerate core compositions, the mass of the non-degenerate interior hydrogen layers, and the effects of finite temperature are used to calculate theoretical MRRs (e.g., Bédard et al. 2020). Despite their sophistication, theoretical MRRs are forced to rely on assumptions about the interior

\* E-mail: pemcgill@ucsc.edu

structure of the white dwarf. This is because the masses of the gravitationally stratified non-degenerate interior hydrogen, helium, and CO layers are poorly constrained by observations of the white dwarf's surface. This is particularly important in the case of the mass fraction of the interior hydrogen layer ( $q_H$ ). Depending on temperature and whether a 'thick' or 'thin' hydrogen layer is assumed, theoretical MRRs can vary by 1-15 percent (Tremblay et al. 2017).

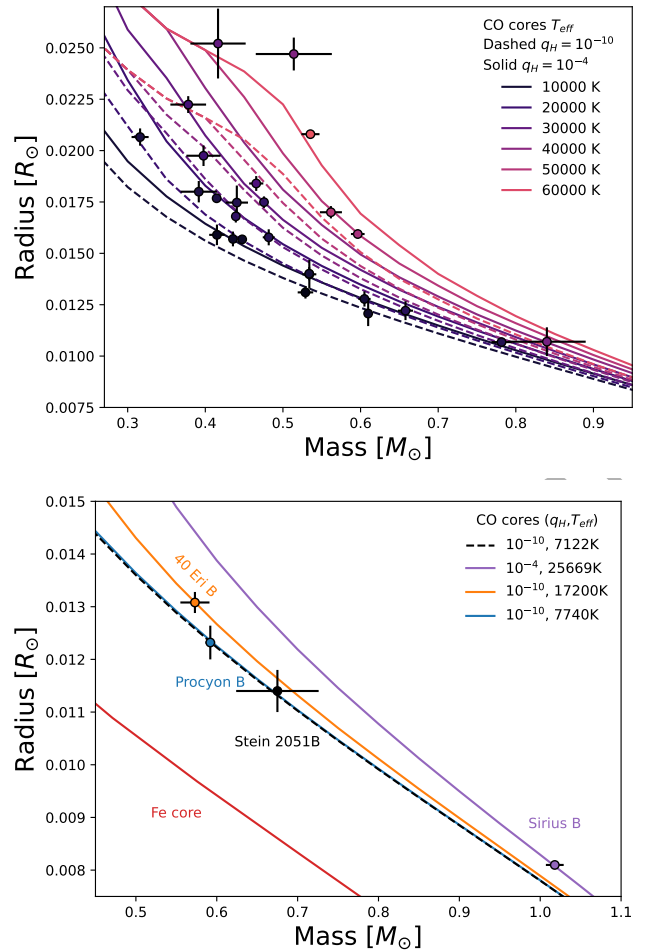
For hydrogen-rich white dwarfs (DA type), a thick hydrogen layer is assumed ( $q_H = 10^{-4}$ ) which is the estimated maximum hydrogen mass that post-asymptotic-giant-branch evolution models predict for a  $0.6M_\odot$  white dwarf after residual nuclear burning (Iben & Tutukov 1984). Helium-rich white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

Despite the white dwarf MRR's importance, observational tests of the relationship are challenging. The state-of-the-art in direct tests of the white dwarf MRR come from white dwarfs spanning a range of masses and radii that are in eclipsing binary systems (see e.g., Parsons et al. 2016, 2017, upper panel in Fig. 1). In these scenarios, both the white dwarf's mass and radius can be determined independently of atmospheric models.<sup>1</sup> Otherwise, for white dwarfs not in eclipsing binary systems, semi-empirical tests of the MRR which depend on white dwarf atmospheric models, are possible.

By fitting atmospheric models to broad-band photometry and spectroscopy (e.g., Giammichele et al. 2012), a white dwarf's atmospheric parameters ( $T_{\text{eff}}$ ,  $\log g$ ), and its solid angle can be measured. Combining the solid angle with distance information (from parallax), allows the photometric radius of the white dwarf to be measured (e.g., Kilic et al. 2020). The photometric radius can then be combined with  $\log g$  to infer the mass of the white dwarf, and the MRR can be tested (Schmidt 1996; Bédard et al. 2017; Bergeron et al. 2019). The problem with this approach, however, is that both the mass and the radius are entirely derived from the atmospheric models and it is often difficult to disentangle observed features of the MRR from systematic effects and degeneracies in the atmospheric models (Tremblay et al. 2017). For more robust semi-empirical tests of the MRR, the photometric radius needs to be combined with a mass determination independent of atmospheric models.

For white dwarfs, it is also possible to obtain MRR information from gravitational redshift measurements (Einstein 1916). The difficulty with this technique is that, for white dwarfs, the shift in the absorption lines due to gravitational redshift is of a similar size to, and degenerate with, the Doppler shift caused by radial motion. This means that the observed shift in spectral lines is a combination of both effects and the gravitational redshift signal can only be isolated if the radial velocity of the white dwarf is precisely known. Determining the radial velocity of a white dwarf is possible when it is in a binary system by taking measurements of its companion (e.g., Joyce et al. 2018). The gravitational redshift can also be measured for groups of white dwarfs that are co-moving (e.g., Pasquini et al. 2019), or by averaging over random radial motions (e.g., Falcon et al. 2010; Chandra et al. 2020). However, in all of these cases, a test of the MRR using gravitational redshifts measurements requires an atmospheric model-dependant photometric radius determination.

The most precise direct masses (and semi-empirical tests of the



**Figure 1.** **Top:** MRRs for 26 white dwarfs in eclipsing binary systems from Parsons et al. (2017) and references therein. Both the masses and radii of these objects were determined directly. Figure reproduced from Fig. 9 in Parsons et al. (2017). **Bottom:** MRRs for nearby white dwarfs in visual binary systems. The masses for 40 Eri B, Procyon B, and Sirius B were determined from astrometric measurement of their orbits (Bond et al. 2015, 2017a,b). The mass for Stein 2051B was obtained via astrometric microlensing (Sahu et al. 2017). For comparison, the red curve shows the theoretical MRR for zero-temperature white dwarfs with an iron (Fe) core (Hamada & Salpeter 1961). Figure reproduced from Fig. 4 in Bond et al. (2017b). In both panels, the theoretical MRRs for CO core white dwarfs were obtained from the cooling models of Bédard et al. (2020).

MRR) for white dwarfs that are not post-common envelope (i.e., white dwarfs not in an eclipsing binary system), come from astrometric measurements of visual binary systems. In these particular semi-empirical MRR tests, the radius of the white dwarf is derived by fitting photometry and spectroscopy and so these tests remain model-dependant and are not entirely free from systematic effects in white dwarf atmospheric models (Tremblay et al. 2017). Fig. 1 shows the MRR for nearby white dwarfs in visual binaries with direct mass determinations; 40 Eri B (Bond et al. 2015), Procyon B (Bond et al. 2017a), and Sirius B (Bond et al. 2017b). Fig. 1 also shows Stein 2051 B, which also happens to be in a visual binary system but its mass has been determined by astrometric microlensing (Sahu et al. 2017). All of these objects have photometric radius determinations, and are in agreement with the theoretical MRRs (Bédard et al. 2020). A semi-

<sup>1</sup> These MRR tests still require limb darkening coefficients, but the adopted values of the limb darkening coefficients impact the final physical parameters typically below their statistical uncertainties (Parsons et al. 2017)

empirical test of the MRR for a single and isolated white dwarf has yet to be performed.

Astrometric microlensing events offer unique opportunities to measure the mass of isolated objects (e.g., [Miralda-Escude 1996](#); [Kains et al. 2017](#); [Rybcicki et al. 2018](#); [Dong et al. 2019](#); [Sahu et al. 2022](#); [Lam et al. 2022](#); [Kaczmarek et al. 2022](#)). When a foreground lens with mass  $M_L$  aligns closely enough with a more distant background source, the gravitational field of the lens deflects the light of the background source forming a major and minor image of the source. The major image is formed on the same side of the lens as the source and the minor image on the opposite side of the lens. The two images, lens, and unlensed source position always lie along the same line (see e.g., [Bramich 2018](#), for a review). As the lens intervenes between the source and observer, the images change position causing an apparent excursion of the source position (astrometric microlensing; [Høg et al. 1995](#); [Miyamoto & Yoshii 1995](#); [Walker 1995](#)). For the astrometric microlensing event considered in this work, we are in a regime where the major image is resolved from the lens for the duration of the event. In this case the astrometric shift due to microlensing from the unlensed source position is (e.g., [Sahu et al. 2017](#); [Bramich 2018](#)),

$$\delta_+(u) = \frac{1}{2} \left[ \sqrt{u^2 + 4} - u \right] \Theta_E \hat{u} \quad (1)$$

Here,  $\Theta_E = \sqrt{4Gc^{-2}M_L(D_L^{-1} - D_S^{-1})}$  is the angular Einstein radius of the lensing system ([Chwolson 1924](#); [Einstein 1936](#)), and  $G$ ,  $c$ ,  $D_L$ , and  $D_S$  are the gravitational constant, the speed of light, the distance to the lens and the distance to the source, respectively.  $\mathbf{u}$  is the angular lens-source separation vector pointing towards the source position in units of  $\Theta_E$  or  $\mathbf{u} = (\Phi_S - \Phi_L)/\Theta_E = \beta/\Theta_E$ , where  $\Phi_{L,S}$  denotes the angular position of the lens and source, respectively.  $u = |\mathbf{u}|$  and  $\hat{u}$  denotes the unit vector. Crucially, if  $\delta_+$  can be measured during the event and  $D_L$  and  $D_S$  are known, then  $M_L$  can be inferred.

For an astrometric microlensing event to be monitored it first must be found. [Refsdal \(1964\)](#) first noted that if the positions and proper motions of celestial objects were known with sufficient accuracy, then close alignments between them, and hence microlensing events, could be predicted ahead of time. This is in contrast to the currently dominant channel of finding photometric microlensing events, in which hundreds of millions of stars are monitored in the Galactic bulge and plane to catch event as they unfold (e.g., [Mróz et al. 2019](#); [Kim et al. 2016](#); [Husseiniova et al. 2021](#)). Following the suggestion of [Refsdal \(1964\)](#) and [Paczynski \(1995, 1996\)](#) many attempts to predict microlensing events followed ([Feibelman 1966, 1986](#); [Sahu et al. 1998](#); [Salim & Gould 2000](#); [Proft et al. 2011](#); [Sahu et al. 2014](#); [Proft 2016](#); [Lépine & DiStefano 2012](#); [Harding et al. 2018](#)). Follow-up of these predictions has proven difficult because of imprecise astrometry which lead to low confidence event predictions. Interest in predicting events was reignited with the advent of astrometry from the *Gaia* satellite ([Gaia Collaboration et al. 2016a](#)), orders of magnitude more precise and numerous than any of its predecessors. First, using *Gaia* Data Release 1 ([Gaia Collaboration et al. 2016b](#)), [McGill et al. \(2018\)](#) made one high confidence prediction of an astrometric microlensing event by a nearby white dwarf. Then, following *Gaia* Data Release 2 ([Gaia Collaboration et al. 2018](#)) many independent studies searched for both astrometric and photometric events resulting in  $\sim 5000$  predictions ([Klüter et al. 2018a,b](#); [Bramich 2018](#); [Bramich & Nielsen 2018](#); [Mustill et al. 2018](#); [Nielsen & Bramich 2018a](#); [Ofek 2018](#); [McGill et al. 2019a,b, 2020](#)). Most recently [Klüter et al. \(2021\)](#) searched for predicted microlensing events using *Gaia* Early Data Release 3 (GEDR3; [Gaia Collaboration et al. 2021](#)) finding 1758

new events, and [Luberto et al. \(2022\)](#) searched for events with brown dwarf lenses.

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

In this paper we present analysis of the follow-up of the astrometric microlensing event caused by the nearby DQ-type white dwarf LAWD 37 (WD 1142-645), originally predicted by [McGill et al. \(2018\)](#). This event peaked in November 2019, with a predicted astrometric shift of  $\delta_+ \approx 2.8$  mas. First, we describe the two sources of data used in this analysis which are the multi-epoch astrometry from *HST*, and the astrometric solutions for the source and lens from GEDR3. Next, in contrast to the analysis of the two previous predicted astrometric microlensing in [Sahu et al. \(2017\)](#) and [Zurlo et al. \(2018\)](#), we detail the combination of these sources of data within a fully Bayesian framework to infer the mass of LAWD 37. Then we describe the concept of leave-one-out cross validation and use it to compare different noise models of the data. We then use the inferred gravitational mass from the astrometric microlensing event as a semi-empirical test of the white dwarf MRR. Finally, we explore the implications of this work on the follow-up of future predicted microlensing events.

## 2 DATA

There are two sources of data used in this study. First, we have multi-epoch *HST* observations of the lensed source position before, during the predicted maximum of the event, and after. And, second, we have the astrometric solution of the source and lens from GEDR3. These astrometric solutions provide information on the unlensed source and lens trajectories. The combination of these two sources of data allows the astrometric shift due to microlensing to be measured, and consequently, LAWD 37's mass to be determined.

### 2.1 *HST* astrometric measurements

We have nine successful epochs of WFC3/UVIS *HST* observations spanning over one year. A summary of these observation used to determine the position of the source<sup>2</sup> from *HST* programs GO- 16251, 15961, and 15705 (PI Kailash C. Sahu) is given in Table 1. All *HST*

<sup>2</sup> Additionally, a series of WFC3/UVIS F814W short exposures (1.5s) were taken at each epoch in an attempt to constrain the position of the lens which was saturated in the long exposures. Unfortunately there were too few unsaturated, high-signal-to-noise stars common to both exposures needed to constrain the *HST* focus drift between the two exposure pointings. Therefore, we did not use any *HST* data to constrain the lens position and instead relied



Epoch	Observation date	Filter	Subarray size (pixels)	Exposure time (s)	Number of exposures
1	1 May 2019	F814W	2048x2048	95	5
2	18 Sep 2019	F814W	2048x2048	82	5
3	25 Oct 2019	F814W	2048x2048	89	5
4	10 Nov 2019	F814W	2048x2048	95	5
5	26 Nov 2019	F814W	1024x1024	85	13
6	5 Dec 2019	F814W	1024x1024	85	13
7	3 Jan 2020	F814W	1024x1024	85	13
8	5 May 2020	F814W	1024x1024	85	13
9	16 Sep 2020	F814W	1024x1024	85	14

**Table 1.** Summary of the *HST* WFC3/UVIS observations used to constrain the position of the source. The subarray is the pixel subarray of the UVIS2 detector which were chosen to minimize Charge Transfer Efficiency (CTE) effects. The data are from *HST* programs GO- 16251, 15961, and 15705 (PI Kailash C. Sahu).

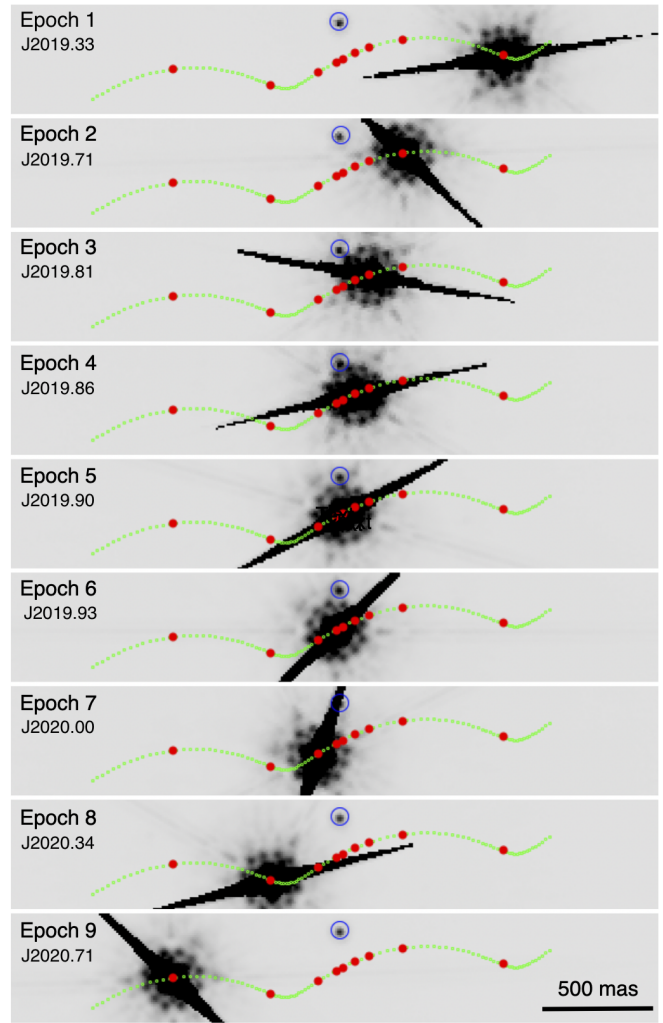
epochs were taken with moderate-sized dithers ( $\pm 100$  pixels) among the pointings. The dither throw was small enough to allow all exposures to include about 20 stars that could serve as astrometric references, but it was large enough to provide some control on Charge Transfer Efficiency (CTE) and distortion residuals.

Figure 2 shows the path of LAWD 37 relative to the source star. In epoch 7, the image of the source is largely occulted by LAWD 37’s bleed column. In epochs 3 through 6, the bump-like features of the PSF introduce considerable complications into our measurement of the source’s location. For this reason, we have developed a model of the extended LAWD 37 PSF in the F814W images.

We combined together all the F814W images of LAWD 37 from each of the 9 epochs and constructed a 4 $\times$ -supersampled version of the LAWD 37 image for that epoch. The top row in Fig. 3 shows the 9 stacked images (in the raw-pixel frame, so that the PSF features will all have the same orientation). In order to subtract the PSF of the WD at each epoch, we used an average of the PSFs constructed from the other 8 epochs (since subtracting the PSF constructed from the same-epoch images would remove the source, particularly at close separations). As can be seen from Fig. 2, the source positions are well separated from each other in different epochs, so that the position measurements of the source are not adversely affected by the presence of the source at other epochs. After this subtraction, we noted that most of the subtracted images were quite clean, but the fourth epoch has considerably larger subtraction residuals than the others, presumably due to the telescope experiencing an unusual focus level due to breathing (Anderson 2018). Unfortunately, this fourth epoch also has the source very close to the lens, which means that the position we measure for the source will be impacted by considerable PSF-subtraction errors. Additionally, the observation in this epoch also suffered from large focus variations. As a result, the PSF subtraction at the position of the source showed large and variable residuals, making the results unreliable. For these reasons, we rejected the fourth epoch from further analysis.

In most of the exposures of epoch 7, the source’s brightest pixels were just offset from the bleed-column and bleed-column-adjacent pixels (see Fig. 2). Hysteresis in the readout amplifier causes some slight correlation from column to column; this is negligible for pixels of similar brightness, but it can be appreciable for low-value pixels side-adjacent to high-value pixels, such as one finds in bleed columns.

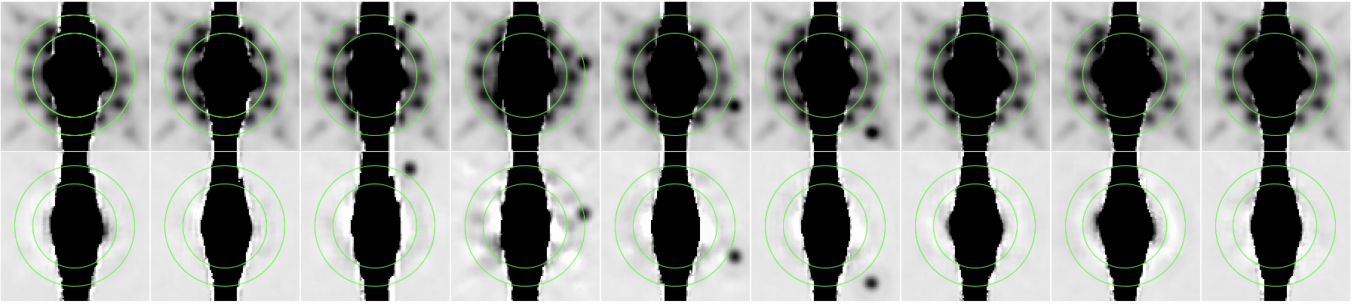
on the projected GEDR3 position of the lens which turned out to be more than adequate for the analysis (see Sections 2.2 and 5.3).



**Figure 2.** *HST* F814W-band image cutouts (co-added stacks by epoch) for each epoch of data during the LAWD 37 event. North is directly upwards in each epoch. The source star is marked with a blue circle. LAWD 37 is the saturated source moving from right to left. The images were made by stacking all images in a given epoch. Dates of each epoch are indicated in Julian years.

When the brightest pixel of a source is available, though, it is possible to measure a position and flux. Normally when a position and flux is measured, a 5 $\times$ 5 raster of pixels centered on the star’s brightest pixel is used. For epoch 7, many of these pixels were corrupted by the bleed column, so the fit-box was modified to only include legitimate pixels.

The source can be seen close to LAWD 37 in the third through the sixth epochs. We used the average PSFs from the eight good-focus epochs (with the source and its vicinity masked out in the relevant images) to construct an average F814W PSF for LAWD 37. Next, this average PSF was subtracted from the LAWD 37 image in each of the individual `_flc` images, which were corrected for CTE using the pixel-based correction (Anderson 2021). We then fitted the source and the other reasonably bright stars (unsaturated and a signal-to-noise ratio  $\geq 50$ ) in each of these images using the F814W “library” PSF described in Sabbi & Bellini (2013). The PSF fits were



**Figure 3.** Top row: Detector-frame cutout images (co-added stacks by epoch) of the inner  $31 \times 31$  pixels ( $1240 \times 1240$  mas) of LAWD 37 for each of the epochs (Epochs 1 to 9 are left to right). The "bumps" in PSF in the annular region between a radius of 7 and 10 pixels (marked in green on the plot) are clear. Bottom row: The same images, but with the average LAWD 37 PSF subtracted (Epoch 4 was excluded from the average). The source can be seen close to LAWD 37 in the third, fourth, fifth, and sixth epochs. The large vertical black region at the center of each of these images corresponds to the saturated pixels at the center of each deep exposure of LAWD 37.

done by a chi-squared-minimization fit of the PSF model<sup>3</sup> to the star's inner  $3 \times 3$  pixels, after subtraction of a modal sky taken from an annulus with radii 8 to 12 pixels (Anderson 2016). This small fitting aperture was used in an effort to minimize any influence of the LAWD-PSF-subtraction residuals on the fit. The fitting solved for three parameters: a flux and a  $(x, y)$  position for each star in the raw frames. These raw positions were then corrected for distortion (Bellini et al. 2011).

The `_f1c` images were corrected for CTE using the most recent pixel-based correction (v2.0, which is currently in the WFC3/UVIS pipeline). Even with this correction, there are small residual CTE-related trends present. Since the various epochs were taken at different telescope orientations, we can inter-compare the measured positions to examine any residual trends. We used linear transformations to map the distortion-corrected positions measured in each exposure in the various epochs into the distortion-corrected frame of the first exposure, using the positions of the unsaturated stars with S/N greater than 100 in the reference exposure and the individual exposures. We then fitted each star with an average position and proper motion. We then transformed these modeled positions back into the individual exposures to examine position residuals as a function of raw  $y$  position and instrumental magnitude ( $m = -2.5 \times \log_{10} F_e$ , where  $F_e$  is the flux in electrons). This allowed us to determine a simple table  $T$  of relative corrections that removed the residual CTE trends as a function of instrumental magnitude. This correction had the form  $T[m](y_{\text{raw}}/2048)$ , and its typical amplitude was  $\sim 0.01$  pixel.

The above analysis allowed us to construct generalized corrections for astrometry in the individual exposures. In order to measure the best possible reference-frame position for the source star over time, we based the final transformations on stars within  $\pm 1$  F814W magnitude of the source star's brightness and within 625 pixels ( $25''$ ) of the source star's position. There were 22 such stars. We used the *Gaia* catalog to predict a reference-frame location for each of these stars at the epoch of each exposure. We removed one star that had a predicted parallax greater than 1 mas, just to be sure we are dealing with distant stars.

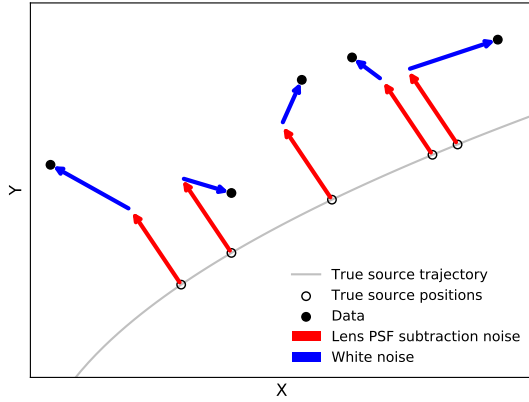
Using the *Gaia*-predicted positions in the reference frame and the observed, distortion and CTE-corrected positions of the stars, we solved for a 6-parameter linear transformation from each observed

image frame into the reference frame. We then examined the residuals of this transformation and tweaked the positions and the proper motions of the individual stars to improve the fits. Note that *HST* astrometric measurements are more precise than *Gaia* measurements, particularly for stars at  $V \sim 18$  or fainter, so it makes sense that the reference-frame positions can be improved via this iteration, while maintaining the average absolute properties of the reference frame (zeropoint, scale, bulk motion, etc) that only *Gaia* can provide. At this stage, one star was found to have larger than expected residuals, likely due to an unresolved companion. This star was rejected, which left us with 20 stars to use as the basis for the transformations into the reference frame. These reference stars have a nominal position precision of 0.02 pixel (0.8 mas) in each exposure, due to shot-noise, read-noise, and small errors in the PSF model, all of which are random sources of noise (see Fig. 2 in Bellini et al. 2014). Thus, we have a distortion- and CTE-corrected position for each of 21 stars (the source plus 20 good reference stars) in each of 87 exposure frames. These positions are then used in the analysis that follows to examine how the source position changes over time.

Since the PSF is known to vary with time, primarily as a consequence of *HST* focus drift, there is a variation in the uncertainty in the quality of the PSF for each epoch. It is important to mention that the *HST* focus does not change gradually and predictably with time. There is an overall trend, but variations during an orbit are larger than the secular variations, and can only be determined after the fact (and with some uncertainty). This means that there are significant PSF variations within an orbit ( $\sim 6\%$  peak-to-peak maximum variations with respect to the library PSF; Bellini et al. 2017, Fig. 5). Critically, this results in a correlated error in the target source position within an epoch. Fig. 4 shows an example realization of noise believed to corrupt the astrometric measurements of the target source within an epoch. The lens PSF subtraction introduces a correlated, within-epoch scatter in addition to the white instrument noise scatter. In order to estimate the size of this correlated error for each epoch, we repeat the PSF subtraction for each epoch with the PSF obtained for each of the other epochs in succession. We use the distribution of residuals as a proxy for the size of the within-epoch correlated error in the target position. The mean of the distribution of the residuals for each epoch ( $m \sigma_{e, \text{corr}}$ ) are shown in Table 2.

In the analysis that follows, all astrometric data are on the tangent plane projected at reference position right ascension  $\alpha_{\text{ref}} = 176.46045340^\circ$ , and declination  $\delta_{\text{ref}} = -64.84488414^\circ$ , on the ICRS. Coordinate  $(\alpha_{\text{ref}}, \delta_{\text{ref}})$  corresponds to (1000, 1000) pixels in

<sup>3</sup> <https://www.stsci.edu/hst/instrumentation/wfc3/data-analysis/psf>



**Figure 4.** Schematic of how the source position data are believed to be generated within an epoch. The data receive a correlated error from lens PSF subtraction, which is the same for all data within an epoch (red). The data are also scattered with white noise (blue).

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

**Table 2.** Estimated sizes of the correlated noise standard deviation due to the PSF subtraction in each epoch of data. These values were obtained via simulated PSF subtraction. Note that the size of the correlated noise tends to increase when the lens and source are closest. Epoch 4 is omitted due to the reasons outlined in Section 2.1. The large scatter on epoch 7 is attributed to the specialized measuring routine used to fit only the uncorrupted pixels due to the source’s proximity to a charge bleed column (see Section 2.1)

the reference-frame tangent plane with the first coordinate,  $X$ , having the direction of the local west unit vector and, the second coordinate,  $Y$ , having the direction of the local north unit vector. Units in both of the coordinates are then scaled to be 1 mas (1pixel=40 mas). In what follows we denote a position in this tangent plane as,  $\zeta = [X, Y]$ .

## 2.2 *Gaia* astrometric solution

GEDR3 provides reference positions, proper motions, and parallax values for both the source and lens. Additionally, each parameter’s standard errors and correlations, which are derived from a linear least squares fit of single-epoch astrometric measurements, are provided (Lindgren et al. 2021). For GEDR3, the astrometric solutions are based on measurements taken between July 2014 and May 2017<sup>4</sup>.

The mean and covariance matrix for the source (GEDR3 source ID 5332606350796955904) astrometric parameters in the tangent plane

projection defined in Section 2.1 are,

$$m_S^G = \begin{bmatrix} 35186.9625 \\ 45709.7257 \\ 8.89044 \\ 0.03676 \\ 0.19940 \end{bmatrix}, \quad (2)$$

$$\Sigma_S^G = 10^{-1} \begin{bmatrix} 0.2 & -0.01 & 0.04 & -0.06 & -0.04 \\ -0.01 & 0.2 & -0.05 & 0.003 & 0.05 \\ 0.04 & -0.05 & 0.3 & -0.02 & -0.02 \\ -0.06 & 0.003 & -0.02 & 0.3 & 0.07 \\ -0.04 & 0.05 & -0.02 & 0.07 & 0.3 \end{bmatrix}. \quad (3)$$

Here, the mean and covariance matrices have the parameter order,  $X_{0,S}, Y_{0,S}$  source reference positions (mas),  $\mu_{X,S}, \mu_{Y,S}$  proper motions in the  $X$  and  $Y$  directions (mas/year), and  $\pi_S$  parallax amplitude (mas). Similarly, for the lens (GEDR3 source ID 5332606522595645952),

$$m_L^G = \begin{bmatrix} 45825.73799 \\ 46592.48286 \\ -2661.63959 \\ -344.93250 \\ 215.67527 \end{bmatrix}, \quad (4)$$

$$\Sigma_L^G = 10^{-3} \begin{bmatrix} 0.2 & -0.03 & -0.02 & -0.04 & -0.03 \\ -0.03 & 0.2 & -0.03 & 0.01 & 0.07 \\ -0.02 & -0.03 & 0.3 & -0.05 & 0.02 \\ -0.04 & 0.01 & -0.05 & 0.4 & 0.06 \\ -0.03 & 0.07 & 0.02 & 0.06 & 0.3 \end{bmatrix}. \quad (5)$$

This astrometric information from GEDR3 for both the source and lens will be used in Gaussian priors for the models described in Section 3. Fig. 5 shows the GEDR3 predicted unlensed source trajectory, and the *HST* astrometric measurements of the source. We can see that there are clear offsets from the GEDR3 unlensed source trajectory in the expected direction of predicted lensing signal, although the data are clearly noisy.

## 3 MODELS

Fig. 5 shows that we are in a low signal-to-noise regime, as the offset in the data relative to the projected GEDR3 unlensed position of the source is comparable to the size of the scatter at each epoch. It is therefore important that we investigate a range of models. In this section we describe the different models that were fitted to the data, and compared to one another. We fit models with and without the astrometric lensing signal and with and without a correlated noise component. We first address the choice of parameterisation of the astrometric lensing signal. Next, we outline the different likelihoods used in each of the models. Finally, we detail the prior distributions used in each model and the methods used to sample the posterior distributions in all models.

### 3.1 Parameterisation of the microlensing signal

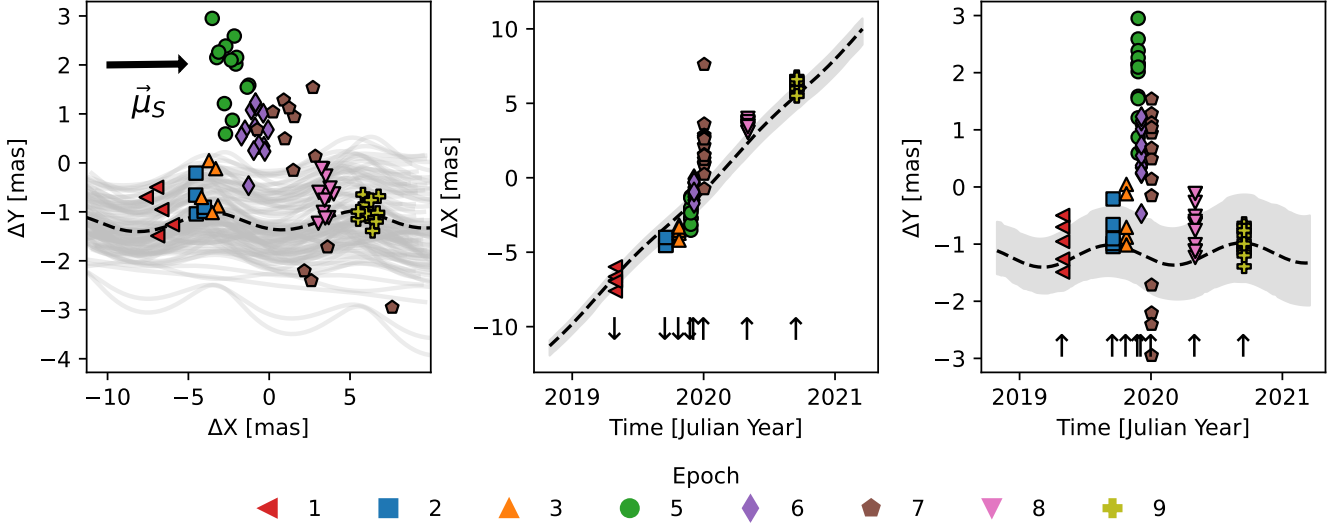
Using the expression for  $\delta_+$  (Eq. 1), and calculating  $\beta$ , using the lens and source trajectories, the lensed source position can be seen to be dependent on 11 parameters,

$$[X_{0,S}, Y_{0,S}, \mu_{X,S}, \mu_{Y,S}, \pi_S, X_{0,L}, Y_{0,L}, \mu_{X,L}, \mu_{Y,L}, \pi_L, M_L]. \quad (6)$$

Specifically, the source ( $\theta_S^{\text{ast}} = [X_{0,S}, Y_{0,S}, \mu_{X,S}, \mu_{Y,S}, \pi_S]$ ) and lens ( $\theta_L^{\text{ast}} = [X_{0,L}, Y_{0,L}, \mu_{X,L}, \mu_{Y,L}, \pi_L]$ ) astrometric parameters

<sup>4</sup> <https://www.cosmos.esa.int/web/gaia/earlydr3>





**Figure 5.** *HST* astrometric follow-up data during the predicted microlensing event by LAWD 37. **Left:** Single *HST* astrometric measurements coloured by time are shown as circles. Measurements are clustered together in time within eight epochs of data. Also shown are 100 random samples of the source unlensed projected trajectory from the GEDR3 astrometric solution. Specifically, projected trajectories corresponding to samples  $\theta_S^{\text{ast}} \sim \mathcal{N}(\mathbf{m}_S^G, \Sigma_S^G)$ , are shown in grey and the black dashed line corresponds to  $\theta_S^{\text{ast}} = \mathbf{m}_S^G$ . **Middle and right:** Projections of the source data in the  $X$  and  $Y$  directions, respectively. Small arrows indicate the predicted direction of the astrometric deflection signal.

and the lens mass,  $M_L$ . However, modeling the signal with these parameters is not straightforward. This is due to the fact that parallax enters the model in two different ways. Firstly, the source and lens parallaxes control the lens and source unlensed trajectories and hence the lens-source angular separation. Secondly, the lens and source parallaxes enter as distance terms and control the size of  $\Theta_E$ . The problem arises due to the interpretation of negative source parallax values when trying to include the GEDR3 astrometric solution as priors in the model.

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

There are a number of ways to mitigate this problem. A simple solution is that  $\Theta_E$  could be fitted for instead of  $M_L$ . In this case the astrometric parameters of the model would be  $[\theta_S^{\text{ast}}, \theta_L^{\text{ast}}, \Theta_E]$ . Here the prior on  $\pi_S$  only enters the model as a trajectory term and hence negative values are permitted. Finally,  $M_L$  can then be extracted after inferring a value for  $\Theta_E$ .

### 3.2 Likelihoods

In order to model the data, we need to setup a likelihood function that encodes our beliefs as to how the data were generated. Specifically, for all models, we assume the process that generated the data at time  $t_i$ , for the source is of the form,

$$\zeta^{\text{obs}}(t_i; \theta, \mathcal{M}_{TN}) = \zeta^T(t_i; \theta^{\text{ast}}) + \epsilon^N(t_i; \theta^{\text{noise}}). \quad (7)$$

Here,  $\zeta^{\text{obs}}$  is the observed source position in the tangent plane given by model  $\mathcal{M}_{TN}$  with trajectory component  $T$  and noise compo-

nent  $N$ .  $\zeta^T$  is the unlensed source position predicted by trajectory component  $T$  with astrometric parameters  $\theta^{\text{ast}}$ .  $\epsilon^N$  is an additive zero-mean Gaussian noise component  $N$  with parameters  $\theta^{\text{noise}}$ . We consider two different trajectory models, with and without the astrometric lensing deflection term. The model without the deflection is,  $\zeta^T(t_i; \theta^{\text{ast}}) = \zeta^N(t_i; \theta^{\text{ast}}) \equiv \zeta(t_i; \theta_S^{\text{ast}})$ . Here,

$$\zeta(t_i; \theta_S^{\text{ast}}) = \begin{bmatrix} X_{0,S} \\ Y_{0,S} \end{bmatrix} + (t_i - t_{\text{ref}}) \begin{bmatrix} \mu_{X,S} \\ \mu_{Y,S} \end{bmatrix} + \pi_S \mathbf{J}^{-1} \mathbf{R}_{\oplus}(t_i) \quad (8)$$

$\mathbf{R}_{\oplus}(t)$  are Cartesian Barycentric solar system coordinates in au of the Earth at time  $t$ .  $\mathbf{R}_{\oplus}(t)$  was retrieved via the Astropy Python package (Astropy Collaboration et al. 2013, 2018) which uses values computed from NASA JPL's Horizons Ephemeris<sup>5</sup>.  $\mathbf{J}^{-1}$  is the inverse Jacobian matrix of the transformation from Cartesian to spherical coordinates (e.g., Urban & Seidelmann 2014, Section 7.2.2.3) projected at the reference position  $(\alpha_{\text{ref}}, \delta_{\text{ref}})$ , and  $t_{\text{ref}} = J2016.0$  is the GEDR3 reference epoch. Eq. (8) is just the standard motion of the source with proper motion and parallax. The model with the deflection is  $\zeta^T(t_i; \theta^{\text{ast}}) = \zeta^D(t_i; \theta^{\text{ast}}) \equiv \zeta(t_i; \theta_S^{\text{ast}}) + \delta_+(t_i; [\theta_S^{\text{ast}}, \theta_L^{\text{ast}}, \Theta_E])$  where  $\theta^{\text{ast}} \equiv [\theta_S^{\text{ast}}, \theta_L^{\text{ast}}, \Theta_E]$ .

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

$$p(D_e | \theta_e, \mathcal{M}_{TN}) = \mathcal{N}(X_e | X_e^T(t_e; \theta^{\text{ast}}), \Sigma_e^N(\theta_e^{\text{noise}})) \times \mathcal{N}(Y_e | Y_e^T(t_e; \theta^{\text{ast}}), \Sigma_e^N(\theta_e^{\text{noise}})) \quad (9)$$

<sup>5</sup> <https://ssd.jpl.nasa.gov/?horizons>

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

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

Here  $\mathcal{N}$  is the  $K_e$  dimensional multivariate Gaussian density,  $\theta_e \equiv [\theta^{\text{ast}}, \theta^{\text{noise}}]$ ,  $X_e^T(t_e; \theta^{\text{ast}})$  and  $Y_e^T(t_e; \theta^{\text{ast}})$  are vectors of  $X$  and  $Y$  source positions of length  $K_e$  obtained by evaluating the vector of times  $t_e$  for the trajectory model component  $T$ .  $\Sigma_e^N(\theta^{\text{noise}})$  is a  $K_e \times K_e$  covariance matrix. We consider two noise models for the covariance matrix. An uncorrelated white noise model with  $\Sigma_e^N(\theta^{\text{noise}}) = \Sigma_e^W(\theta^{\text{noise}}) \equiv \sigma_{\text{white}}^2 \mathbf{I}$ , where  $\mathbf{I}$  is the  $K_e \times K_e$  unit diagonal identity matrix, and  $\theta_e^{\text{noise}} \equiv [\sigma_{\text{white}}]$ . We also consider a correlated and white noise model with  $\Sigma_e^N(\theta^{\text{noise}}) = \Sigma_e^C(\theta^{\text{noise}}) \equiv \Sigma_e^W([\sigma_{\text{white}}]) + \sigma_{e,\text{corr}}^2 \mathbf{1}$ , where  $\mathbf{1}$  is the  $K_e \times K_e$  full ones matrix, and  $\theta_e^{\text{noise}} \equiv [\sigma_{\text{white}}, \sigma_{e,\text{corr}}]$ . The correlated noise model corresponds to the data generation process shown in Fig. 4.

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

$$p(\mathcal{D}|\theta, \mathcal{M}_{TN}) = \prod_{e=1}^{N_e} p(D_e|\theta_e, \mathcal{M}_{TN}). \quad (10)$$

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

### 3.3 Priors

For the source and lens astrometric parameters (10 parameters total), there is prior information from GEDR3. Specifically we assume multivariate normal distributions,  $p(\theta_S^{\text{ast}}) = \mathcal{N}(\mathbf{m}_S^G, \Sigma_S^G)$ , and  $p(\theta_L^{\text{ast}}) = \mathcal{N}(\mathbf{m}_L^G, \Sigma_L^G)$  using the values in Eqs. (3) and (5).

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

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

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

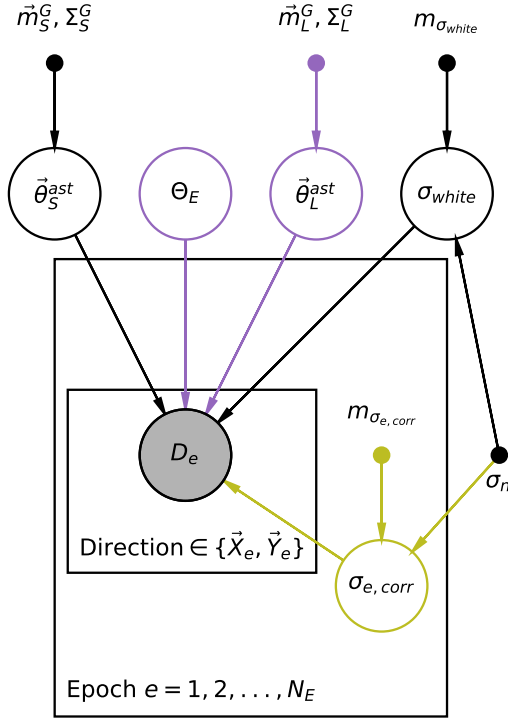
lens and source astrometric parameters would not be representative of the true unlensed lens and source trajectories.

Firstly, for the source, we have to check that when the GEDR3 data were taken, there was not a significant lensing signal present. As LAWD 37 is so bright *Gaia* downloads cut-outs and carries out gating so when the background object is within 500 mas it is downloaded as part of the cut out but with less exposure time, otherwise, standard one dimensional processing is used (Fabricius et al. 2016). GEDR3 is based on data collected between July 2014 and May 2017. In May 2017, the predicted deflection of the source is  $< 0.2$  mas and the lens source separation is  $\gg 500$ mas. This is below the along scan (AL) precision for a  $G \approx 18$  mag source ( $\sigma_{\text{AL}} \approx 0.8$  mas with the standard *Gaia* pipeline; Rybicki et al. 2018; Bramich 2018; Everall et al. 2021). We therefore conclude that the astrometric lensing signal was not detectable during the time GEDR3 data were collected and therefore did not significantly influence the GEDR3 astrometric solution of the source. Secondly, for the lens, we have to check if the shift due to the blending with the minor image was detectable during GEDR3 (see Eq. (15) in Bramich 2018). In May 2017 this shift was  $\approx 10^{-13}$  mas, we therefore safely conclude that the lens GEDR3 astrometric solution does not contain a significant astrometric lensing signal.

Finally, we have to consider if the GEDR3 astrometric solution of the  $G \approx 18$  mag source has likely been influenced by the presence of the comparatively bright  $G \approx 11$  mag lens. The current *Gaia* processing pipeline is able to resolve sources for separations in the most optimal cases down to  $\approx 200$  mas (Arenou et al. 2018). While the lens-source contrast ratio is far from optimal in our case, in May 2017, the lens and source had a predicted separation of  $\approx 6400$  mas. At this separation, the lens and source were unlikely to be close to each other on the *Gaia* focal plane during the GEDR3 time baseline. Therefore, it is safe to conclude that the source GEDR3 astrometric solution is unlikely to have been significantly affected by the presence of the lens. Overall we conclude, for both the lens and source, the GEDR3 solutions are safe to use as priors on the unlensed source and lens trajectories in the models. Finally, it is worth mentioning that the astrometric solution priors from GEDR3 will affect the precision at which we can measure the mass of the lens, we defer discussion of this point to Section 5.3.

For the white noise parameter present in all models, we set  $p(\sigma_{\text{white}}) = \mathcal{N}(\sigma_{\text{white}}|m_{\sigma_{\text{white}}}, \sigma_n^2)$  where  $m_{\sigma_{\text{white}}} = 0.8$  mas and  $\sigma_n = 0.1$  mas, reflecting the estimated instrument precision of WFC3/UVIS (Bellini et al. 2014). For the correlated noise parameters  $\sigma_{\text{corr}} = [\sigma_{1,\text{corr}}, \sigma_{2,\text{corr}}, \dots, \sigma_{N_e,\text{corr}}]$  we assume a Gaussian prior  $p(\sigma_{\text{corr}}) = \mathcal{N}(\sigma_{\text{corr}}|\mathbf{m}_{\text{corr}}, \sigma_n^2 \mathbf{I})$ , truncated at zero to avoid negative values. Here,  $\mathbf{m}_{\text{corr}} = [m_{\sigma_{1,\text{corr}}}, m_{\sigma_{2,\text{corr}}}, \dots, m_{\sigma_{N_e,\text{corr}}}]$  are the estimated size of the correlated noise components in Table 2 and we have assumed no correlation between epochs. Finally, we then assume a uniform prior on  $\Theta_E$ , as  $p(\Theta_E) = \mathcal{U}(\Theta_E|\text{lower} =$





**Figure 6.** Probabilistic graphical model showing the dependence structure of the models considered in this work. An arrow from one node to another indicates a conditional dependence. No arrow between two nodes means they are conditionally independent. Unfilled circles are latent random variables in the models or parameters that are fitted for. Filled small circles are fixed values in the model (parameters for the informative prior distributions). The shaded circles are the observed data. Parameters inside a plate are repeated for each epoch and then direction. Parameters outside the plates are global parameters. Black parts of the graph are common to all models considered. Purple parts are common to models with an astrometric deflection. Pale green parts of the graph are common to models with correlated noise.

20 mas, upper = 60 mas). This prior was chosen to be wide enough to be uninformative, yet narrow enough to constrain the model to reasonable areas of the parameter space which allowed fast model fitting. In all models, we build the full prior by taking the product over all the priors of the required parameters. Table 4 contains a summary of all parameter priors, and Fig. 6 shows the probabilistic graph illustrating all parameter dependencies and structure of the models.

### 3.4 Sampling the posterior

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

$$p(\theta|\mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\theta, \mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})} \propto p(\mathcal{D}|\theta, \mathcal{M})p(\theta|\mathcal{M}). \quad (11)$$

We obtain samples from the posterior distribution using an MCMC algorithm. Specifically, we use the No-U-Turn Sampler (NUTS) Hamiltonian Monte Carlo algorithm (Homan & Gelman 2014) implemented by the PyMC3 Python package (Salvatier et al. 2016). NUTS allows samples from the posterior distribution to be obtained faster than other traditional MCMC samplers (e.g., Foreman-Mackey

et al. 2013) because it uses first-order gradient information to efficiently step through the parameter space. We also take advantage of using the dense full mass matrix step implemented in the Exoplanet Python package for further performance gains (Foreman-Mackey et al. 2021).

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

## 4 MODEL COMPARISON AND CRITICISM

### 4.1 Leave-one-out cross-validation

We are fitting four models to the data.  $\mathcal{M}_{DC}$  - a model with the astrometric microlensing deflection and correlated noise,  $\mathcal{M}_{DW}$  - a model with the astrometric microlensing deflection and just white noise,  $\mathcal{M}_{NC}$  - a model with no astrometric microlensing deflection but with correlated noise, and finally,  $\mathcal{M}_{NW}$  - a model with no astrometric microlensing deflection and just white noise. We then need to assess which model best explains the data and critically examine the strengths and weakness of each of the models. To do this, we use the Bayesian Leave-One-Out cross-validation score (LOO). LOO is one method to estimate point-wise out-of-sample prediction accuracy of a given model (Vehtari et al. 2017). LOO is calculated for a given model by fitting the model to the data set where one of the data points has been left out. The posterior samples of that fit are then projected through the model likelihood to assess how well the left-out data point is predicted by the model. The procedure is then repeated so each data point is left out in turn. For a given model, this provides a per data point score which can be totalled over the data to give an indication of overall model performance, or compared data point-wise with a different model allowing an interpretable comparison between models.

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

$$\text{LOO}_{i, \mathcal{M}} = \log p(D_i|\mathcal{D}_{-i}, \mathcal{M}). \quad (12)$$

Here  $\mathcal{D}_{-i}$  is the full data set  $\mathcal{D}$  with the  $i$ th data point or  $D_i$  removed. This is the log of the LOO predictive density conditioned on the data set without the  $i$ th data point. It can be written in terms of expected value of the likelihood of the left-out data point over the posterior distribution obtained while fitting the model to  $\mathcal{D}_{-i}$ ,

$$p(D_i|\mathcal{D}_{-i}, \mathcal{M}) = \int p(D_i|\theta, \mathcal{D}_{-i}, \mathcal{M})p(\theta|\mathcal{D}_{-i}, \mathcal{M})d\theta. \quad (13)$$

Practically, if we have  $S$  samples from the posterior distribution of the model parameters  $\{\theta_{-i}^s\}_{s=1}^S$ , obtained by fitting model  $\mathcal{M}$  to the data set  $\mathcal{D}_{-i}$ , we can use these samples to compute Eq. (13) as,

$$p(D_i|\mathcal{D}_{-i}, \mathcal{M}) = \frac{1}{S} \sum_{s=1}^S p(D_i|\mathcal{D}_{-i}, \theta_{-i}^s, \mathcal{M}). \quad (14)$$

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

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

$$\text{LOO}_{\mathcal{M}} = \sum_{i=1}^N \text{LOO}_{i,\mathcal{M}}. \quad (15)$$

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

$$\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2} = \text{LOO}_{\mathcal{M}_1} - \text{LOO}_{\mathcal{M}_2}. \quad (16)$$

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

$$\text{se}(\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}) = \sqrt{N V_{i=1}^N (\text{LOO}_{i,\mathcal{M}_1} - \text{LOO}_{i,\mathcal{M}_2})}. \quad (17)$$

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

In the case of this analysis, it is also informative to sum over the point-wise LOO score over the data in a single epoch,  $e$ . This is because the data are tightly temporally clustered within an epoch, and in the correlated noise models ( $\mathcal{M}_{DC}, \mathcal{M}_{NC}$ ), data within an epoch share correlated noise properties. We define the difference in LOO predictive accuracy over an epoch  $e$ , as  $\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}^e$ . This quantity is analogous to  $\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}$  (defined in Eqs. 15 and 16), but instead of summing the point-wise score over all data points, the sum is taken only over the data in epoch,  $e$ . The standard error,  $\text{se}(\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}^e)$ , is also calculated analogously by instead calculating the variance of the point-wise difference of data within the epoch. In this case,  $N$  in Eq. 17 is the total number of data points within epoch  $e$ . Overall, computation of  $\Delta\text{LOO}_{\mathcal{M}_1,\mathcal{M}_2}^e$  will permit the comparison of models at epoch resolution.

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

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

## 4.2 The case for LOO over other comparison metrics

LOO is just one of many metrics that can be computed to assess the performance of a model. In this section, we briefly justify the decision to use LOO as a model criticism and comparison tool compared to

the two commonly used approaches in astronomy: a reduced  $\Delta\chi^2$  approach, and use of the Bayesian model evidence.

Typically in microlensing event analyses, albeit in analyses of photometric microlensing events, a reduced  $\Delta\chi^2$  approach is used to select between competing models (e.g., [Bond et al. 2004](#); [Smith et al. 2005](#); [Alcock et al. 2000](#); [Bennett et al. 2018](#)). There is a multitude of reasons why we do not use it here and choose LOO instead. Firstly, because some of the models considered in this work contain Gaussian-correlated noise components, reduced  $\Delta\chi^2$  is no longer fully descriptive of the likelihood of the model (reduced  $\Delta\chi^2$  is only related to the likelihood of an uncorrelated Gaussian likelihood with a diagonal covariance matrix). Secondly, reduced  $\Delta\chi^2$  is only valid for a model that is linear in its parameters; none of the models considered in this work are linear in all their parameters. Thirdly, reduced  $\Delta\chi^2$  fails to account for any posterior uncertainty on the parameters. [Andrae et al. \(2010\)](#) gives an extensive account of the pitfalls of using reduced  $\Delta\chi^2$  for comparison of non-linear models.

Comparison of models using the Bayesian evidence (the denominator in Eq. 11) is becoming popular in astrophysics due to nested sampling algorithms that readily allow its computation (e.g., [Skilling et al. 2006](#); [Higson et al. 2019](#); [Speagle 2020](#)). The Bayesian evidence has the appealing properties of fully capturing parameter uncertainty and naturally penalizes more complex models that do not significantly explain the data better. The critical downside, however, is that the evidence is sensitive to the choice of prior distribution (see e.g., [Fong & Holmes 2020](#)). This becomes a problem when comparing models possessing parameters with uninformative and somewhat arbitrarily set prior distributions (see e.g., Section 7.2 of [Gelman et al. 2013](#)). For example, for  $\Theta_E$  in this analysis, the arbitrary choice of a large width for its uninformative prior can arbitrarily change the model evidence without any resulting change of the posterior distribution.

Comparatively, LOO is typically less sensitive to the model priors. This is because each  $\text{LOO}_{i,\mathcal{M}}$  term is computed with the model conditioned on the rest of data (Eq. 13). This means in scenarios where the number of data points is large, the prior can be overwhelmed by the likelihood and has less of an effect on each  $\text{LOO}_{i,\mathcal{M}}$  term. In the analysis presented in this work, although the number of fitted data points is 81, it is difficult to assess if this constitutes a large data set due to the use of informative priors on the source and lens astrometry, and the noise components of the models. Finally, the Bayesian evidence only provides a single summary statistic for the whole model and data, shedding little light on precisely where a model fails, whereas LOO provides an interpretable per data point score.

## 5 RESULTS

### 5.1 Performance of the models

All models ( $\mathcal{M}_{DC}, \mathcal{M}_{DW}, \mathcal{M}_{NC}, \mathcal{M}_{NW}$ ) were fitted to the data. Posterior parameter summaries for all parameters can be found in Table 5, and posterior projections of the model over the data are shown in Fig. 7. Additionally, LOO scores were computed for all models as described in Section 4.1.

Table 6 shows the pairwise comparison between all model combinations and the total LOO scores.  $\mathcal{M}_{DC}$  (deflection and correlated noise model) has the best overall LOO score when compared to all other models.  $\mathcal{M}_{NW}$  is the least preferred model with all other models having a higher score. The model most competitive with  $\mathcal{M}_{DC}$ , is  $\mathcal{M}_{NC}$  with  $\Delta\text{LOO}_{\mathcal{M}_{DC},\mathcal{M}_{NC}} = 8.6$  and  $\text{sig}(\Delta\text{LOO}_{\mathcal{M}_{DC},\mathcal{M}_{NC}}) = 1.5$ . This means that the correlated component of the noise is an important feature of the model, since

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

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

the deflection model with just white noise ( $M_{DW}$ ) is comparably not competitive with  $M_{NC}$ . In fact, Table 6 shows that  $M_{NC}$  is preferred over  $M_{DW}$  with  $\Delta LOO_{M_{NC}, M_{DW}} = -20.9$  and  $\text{sig}(\Delta LOO_{M_{NC}, M_{DW}}) = 2.2$ . That is, the non-deflection correlated noise model is preferred over the model with the deflection and just white noise. While  $M_{DC}$  is the overall preferred model, it is informative to understand how exactly  $M_{NC}$  is able to explain the data with no deflection term and even beat the deflection model with just white noise,  $M_{DW}$ .

The starting point for understanding the good performance of  $M_{NC}$  is to examine the per-epoch LOO scores. Fig. 8 shows the per-epoch LOO scores for  $M_{DW}$  and  $M_{DC}$  compared to  $M_{NC}$ . For the comparison of the two correlated noise models ( $M_{DC}$  versus  $M_{NC}$ ), it is shown that  $M_{DC}$  marginally beats  $M_{NC}$  in every epoch with the exception of epoch 2 where  $M_{DC}$  is more clearly preferred and epoch 7 where  $M_{NC}$  beats  $M_{DC}$ . The reason why  $M_{NC}$  can explain the epoch with the largest deflection terms (epochs 5, 6, and 7), is that it inflates the correlated noise and alters the unlensed source trajectory.

Fig. 9 shows the priors and posterior distributions of the noise parameter in all of the models. For the correlated noise parameters,  $M_{NC}$  inflates the size of the noise parameters relative to  $M_{DC}$  and the prior for epochs 5, 6 and 7.  $M_{NC}$  does this to try and explain the deflection signal. For epoch 2, the correlated noise term in  $M_{NC}$  is slightly inflated compared to  $M_{DC}$  and the prior. At first glance, this seems counterintuitive because the deflection at epoch 2 is small. This begs the question as to what signal  $M_{NC}$  is trying to explain away with increased correlated noise. The reason for this is that, although the signal at epoch 2 is small, so is the estimated magnitude of the correlated noise, and so epoch 2 has one of the highest signal-

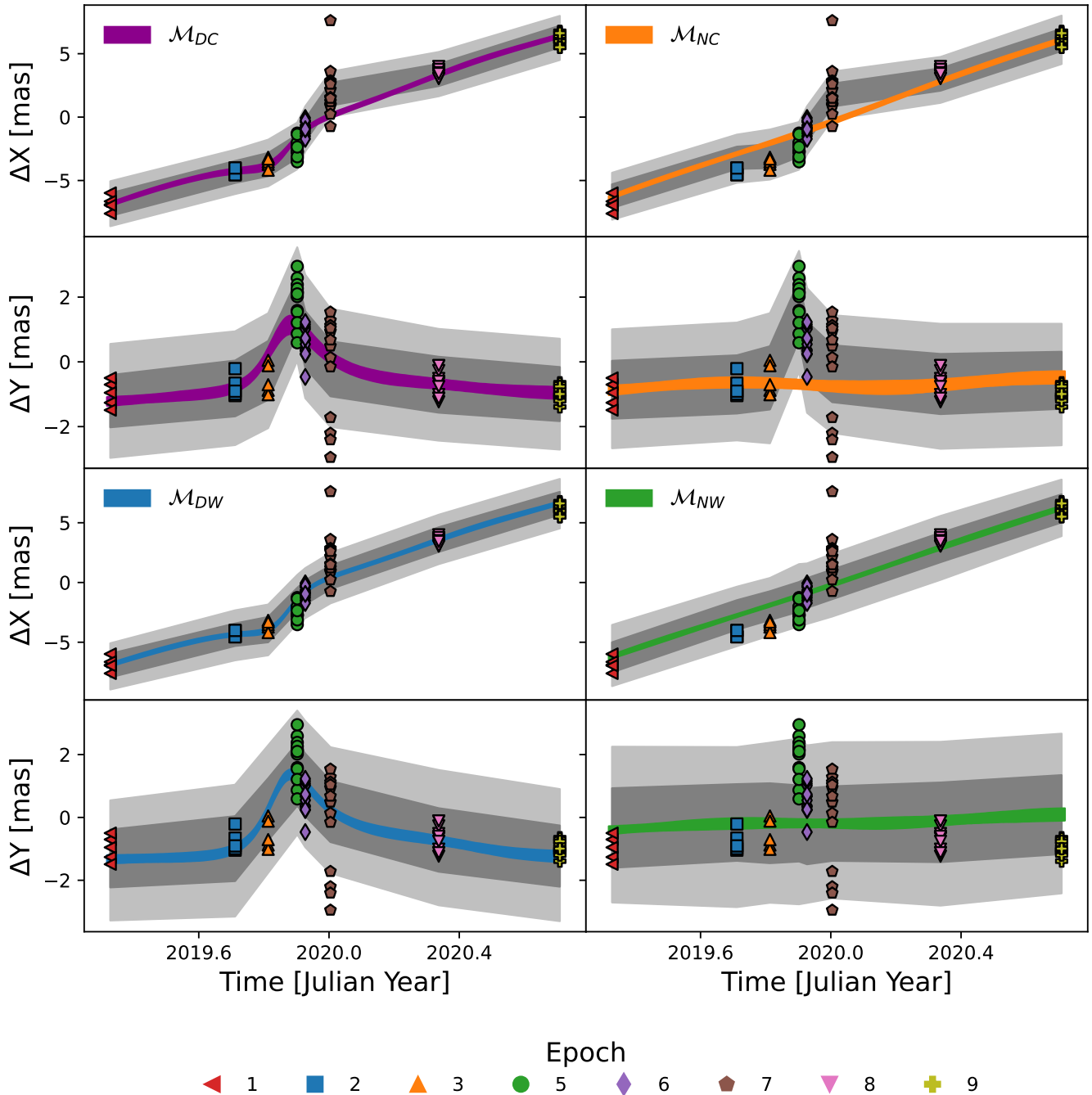
	$\Delta LOO_{row, column}$			
	$M_{DC}$	$M_{DW}$	$M_{NC}$	$M_{NW}$
$M_{DC}$	-	29.5(3.5)	8.6(1.5)	76.5(4.6)
$M_{DW}$	-29.5(3.5)	-	-20.9(2.2)	47.1(3.4)
$M_{NC}$	-8.6(1.5)	20.9(2.2)	-	67.9(5.4)
$M_{NW}$	-76.5(4.6)	-47.1(3.4)	-67.9(5.4)	-
$LOO_M$	-215.6	-245.0	-224.1	-292.1
$se(LOO_M)$	38.3	37.8	33.2	26.0

**Table 6.** Pairwise difference in LOO scores for all models considered. Positive number means the model in the row is preferred. Significance,  $\text{sig}(\Delta LOO_{row, column})$ , of the scores are indicated in with the parentheses. Total LOO score for each more and its standard error are shown in the bottom two rows. In order of descending LOO score (highest to lowest predictive accuracy) the models are  $M_{DC}$ ,  $M_{NC}$ ,  $M_{DW}$  and  $M_{NW}$ .

to-noise ratios of all the epochs. Overall, this means correlated noise can mimic the deflection signal. It is noted that in epoch 7 and for the white noise both  $M_{NC}$  and  $M_{DC}$  inflate the noise terms relative to the prior. This suggests that the priors underestimate both of these quantities. It is also noted that for models missing either the correlated noise or the deflection signal, the white noise is increased to compensate relative to  $M_{DC}$  and the prior.

Fig. 10 shows the GEDR3 priors and posterior distributions of the source astrometric parameters. While there is broad agreement between all models for the source proper motion in the  $X$  direction and the source parallax,  $M_{NC}$  and  $M_{DC}$  disagree on the  $Y$  direction proper motion posterior.  $M_{DC}$  infers  $\mu_{Y,S} = 0.028^{+0.047}_{-0.047}$  mas/year, whereas  $M_{NC}$  infers  $\mu_{Y,S} = 0.159^{+0.046}_{-0.046}$  mas/year. The relatively high value inferred by  $M_{NC}$  is caused by  $M_{NC}$  trying to explain



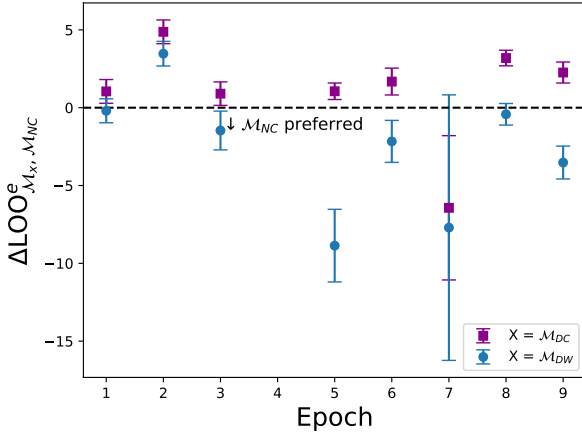


**Figure 7.** Posterior realizations plotted over the data in the X and Y directions and for each of the considered models. Coloured bands show the 84th-16th posterior percentiles on the inferred source trajectory for each model and direction. Dark (light) grey bands are the posterior data realizations 16th-84th (2nd-98th) percentile bands. Specifically this includes the trajectory and the noise model component realizations. The posterior data realizations are discontinuous between epoch because the noise model is only defined within an epoch and on the data grid.

away the positive Y direction deflection (see e.g., Fig. 7) by altering the source trajectory. This means that further data taken after the event to pin down  $\mu_{Y,S}$  could completely rule out  $\mathcal{M}_{NC}$  as a viable model. Encouragingly, we note that for all models the lens and source astrometric parameters are consistent with the GEDR3 priors (see Figs. B2 and B1 in Appendix B).

The reason for the better performance of  $\mathcal{M}_{DW}$  compared with  $\mathcal{M}_{NC}$  in epoch 2 is a high signal-to-noise deflection (as mentioned

earlier), combined with the inflexibility of the source trajectory to be altered to explain away an offset in the negative X direction. This is due to the asymmetrical deflection in the X direction being in the negative (before the closest approach) then positive (after the closest approach) X direction. The source trajectory cannot be altered in  $\mathcal{M}_{NC}$  to account for both of these offsets, so  $\mathcal{M}_{NC}$  performs worse than  $\mathcal{M}_{DC}$  in epoch 2. The better performance of  $\mathcal{M}_{NC}$  compared to  $\mathcal{M}_{DW}$  in epoch 7 is due to the outlying data within that epoch



**Figure 8.** LOO score of models  $M_{DC}$  and  $M_{DW}$  compared to  $M_{NC}$  within each epoch. Error bars are one standard error.

(see Fig. 7). These outlying data are located at lower  $Y$  values which are further away from the deflection trajectory of  $M_{DC}$  than the unlensed trajectory of  $M_{NC}$ .

For the per-epoch comparison of  $M_{NC}$  and  $M_{DW}$ , Fig. 8 shows that  $M_{NC}$  beats  $M_{DW}$  in every epoch apart from epoch 2. The relative performance of  $M_{NC}$  and  $M_{DW}$  can be explained by the same reasoning used above, for the  $M_{DC}$  and  $M_{NC}$  epoch comparison. In epoch 5, Fig. 8 shows  $M_{DW}$  clearly performs worse than  $M_{NC}$ , despite there being a large deflection signal in epoch 5. This is due to epoch 5 also having a large correlated noise estimate ( $m\sigma_{e, \text{corr}} = 0.6$  mas) which white noise  $M_{DW}$  cannot explain away, even by inflating the size of the white noise (see Fig. 9).

Fig. 5 shows that epoch 7 has a large scatter in the data compared with all of the other epochs. This is due to the source lying close to a column bleed as shown in Fig. 2, which meant a specialized measuring procedure to fit only the uncorrupted pixels had to be used. This could mean that the data in epoch 7 are particularly unreliable. Furthermore, because all of the data in epoch 7 are potentially effected by the source’s proximity to the charge bleed column a LOO analysis is not sensitive to this, since withholding one epoch 7 data point leaves in the remaining 12. Therefore as a safety check we compare the two best models  $M_{DC}$  and  $M_{NC}$  whilst leaving all of epoch 7 out of the analysis. In this case  $\text{LOO}_{M_{DC}} = -95.2 \pm 7.9$  and  $\text{LOO}_{M_{NC}} = -103.8 \pm 7.2$ . This means the model with the deflection and correlated noise is still preferred over  $M_{NC}$ , but with slightly higher significance,  $\text{sig}(\Delta \text{LOO}) = 2.7$ . Overall, the model selection conclusions are not sensitive to including or withholding of epoch 7 for  $M_{DC}$  and  $M_{NC}$ .

## 5.2 Inference on the angular Einstein radius

Both models including the astrometric lensing deflection signal ( $M_{DW}$  and  $M_{DC}$ ), provide a posterior inference on  $\Theta_E$ . Fig. 11 shows the  $\Theta_E$  marginal posterior distribution for both the  $M_{DW}$  and  $M_{DC}$  models, along with the prior used in both models. For  $M_{DW}$  and  $M_{DC}$ , the inferred values are  $\Theta_E = 34.2^{+1.4}_{-1.4}$  mas and  $\Theta_E = 31.4^{+2.1}_{-2.2}$  mas, respectively. Here the values and upper and lower error bars represent the 50th, 84th-50th, and 16th-50th posterior percentiles, respectively. Fig. 11 shows that  $M_{DW}$  provides a tighter constraint and slightly higher value of  $\Theta_E$  compared with

$M_{DC}$ . It is also shown that the  $M_{DC}$   $\Theta_E$  posterior distribution is slightly asymmetrical with more probability mass towards lower  $\Theta_E$  values.

The difference in the  $\Theta_E$  posteriors between  $M_{DW}$  and  $M_{DC}$  is consistent with the findings in Section 4. In Section 4, it was shown that the correlated noise part of the model (with help from a positive  $\mu_{Y,S}$ ) can mimic or explain away some of the microlensing signal. Therefore, jointly fitting both the correlated noise and the microlensing signal means the observed offset on position explanation is shared amongst the microlensing signal and correlated noise, causing a slightly lower posterior median  $\Theta_E$  (as  $\delta_+ \propto \Theta_E$  for  $u \gg 1$ ) for  $M_{DC}$  compared with  $M_{DW}$ . Conversely, because the white noise model in  $M_{DW}$  cannot explain correlated noise at each epoch, it inflates the values of  $\Theta_E$  to compensate. The larger spread in the  $\Theta_E$  posterior for  $M_{DC}$  is likely similarly due to the fact the correlated noise can mimic the microlensing signal, mean that there is some degeneracy between them, which ultimately leads to a less certain inference on  $\Theta_E$  in  $M_{DC}$ .

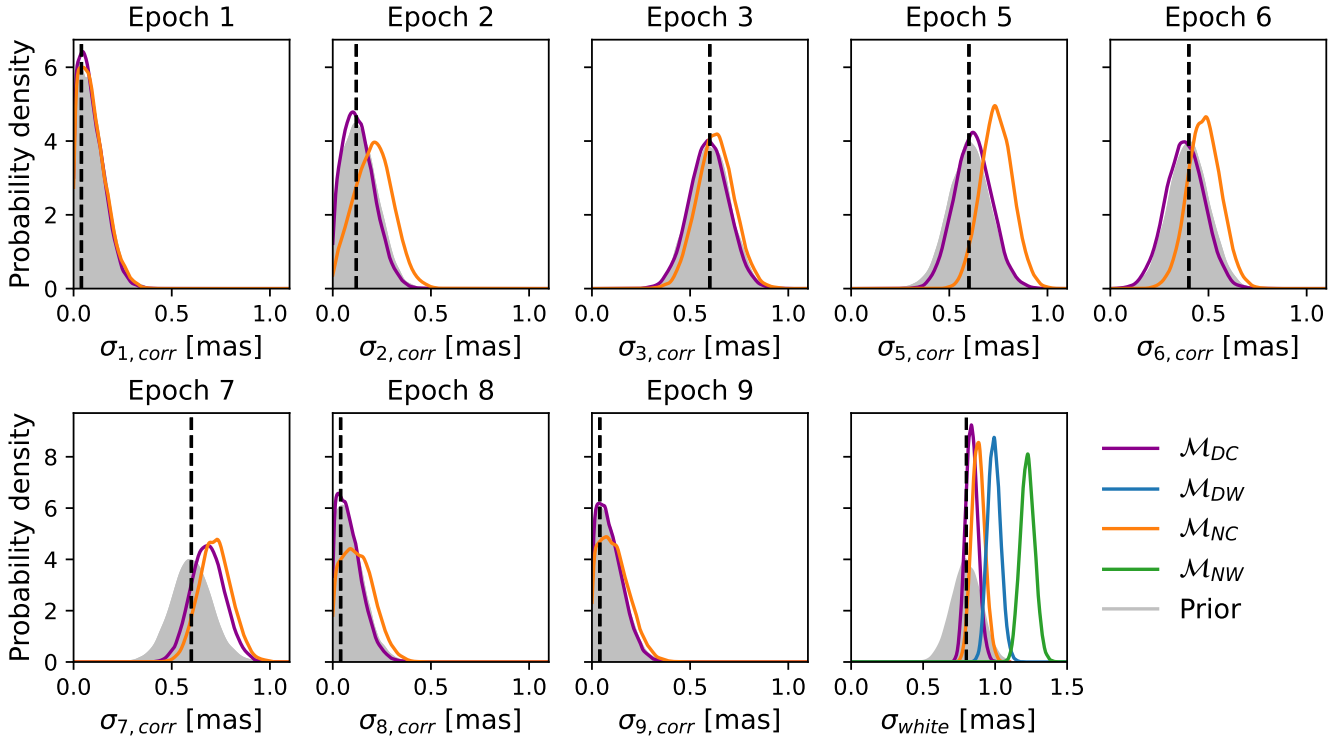
We now use the posterior samples to calculate an inferred value for  $M_L$ , the mass of LAWD 37. Inverting the expression for  $\Theta_E$ , we can write the  $M_L$  in terms  $\Theta_E$  and  $\pi_L - \pi_S$ . Of course, we run into the same difficulties described in Section 3.1 of negative  $\pi_S$  values entering as distance terms in  $\Theta_E$ . This is not a problem because the source is so distant,  $\pi_S \approx 0.1$  mas, compared with the lens,  $\pi_L \approx 215$  mas, we can approximate  $\pi_L - \pi_S \approx \pi_L$ . Under this approximation we get  $M_L$  for the correlated noise model,  $M_L = 0.56 \pm 0.08 M_\odot$ , and  $M_L$  for the white noise model,  $M_L = 0.66^{+0.06}_{-0.05} M_\odot$ . Here, we report the 50th, 84th-16th, and 50th-16th posterior percentiles. We note inclusion of  $\pi_S$  both with negative posterior samples and with the negative posterior values truncated, does not change the reported mass values or uncertainties for either model.

We also find that leaving out all of the data in epoch 7 (for the reasons mentioned at the end of Section 5.1) does not significantly change the inference on the lens mass for  $M_{DC}$  –  $M_L = 0.54 \pm 0.06 M_\odot$  when all of epoch 7 is withheld. The mass uncertainty is slightly smaller indicating epoch 7’s large spread is inflating the uncertainty in the final mass inference. Overall, the lens mass inference is not sensitive to the including or withholding of epoch 7.

## 5.3 Prior sensitivity

We can check how sensitive the posterior inference on  $\Theta_E$  is to the prior assumptions in  $M_{DC}$ . Specifically, we can test how tightening or relaxing the prior parameter distributions affects the  $\Theta_E$  posterior distribution with the view of determining which of our assumptions strongly affect our inferences. For both the lens and source unlensed trajectory, we used informative Gaussian priors from GEDR3. The right panel of Fig. 12 shows how changing (inflating or shrinking by a multiplicative factor) the GEDR3 prior covariance matrix on the source ( $\Sigma_S^G$ ) or lens ( $\Sigma_L^G$ ), changes the  $\Theta_E$  posterior constraint. Fig. 12 shows that the inference on  $\Theta_E$  is insensitive to changing  $\Sigma_L^G$ . Specifically, the posterior constraint on  $\Theta_E$  does not degrade until  $\Sigma_L^G$  is inflated by a factor of  $1000^2$  (or equivalently multiplying the standard deviation by a factor of 1000). Moreover, it is also shown that shrinking  $\Sigma_L^G$  (even shrinking by a factor of  $1000^2$ ) does not improve the posterior constraint on  $\Theta_E$ . This means that further data on the lens position, either by future *Gaia* data releases or further *HST* monitoring, is unlikely to improve the constraint on  $\Theta_E$ .

The right panel of Fig. 12 also shows the effect of changing the source unlensed trajectory prior covariance matrix  $\Sigma_S^G$ , specifically



**Figure 9.** Prior and marginal posterior probability density functions for the noise parameters in the different models. Vertical dashed lines show the values of these parameters estimated from the PSF subtraction simulations (Table 2). The prior probability density function is shaded to aid differentiation with the posteriors. Note that the models with no correlated noise,  $\mathcal{M}_{DW}$  and  $\mathcal{M}_{NW}$ , do not have correlated noise parameters and therefore, only the posteriors on  $\sigma_{white}$  are shown.

that the outcome is somewhat more sensitive to the prior on the source unlensed trajectory compared with the lens trajectory. Fig. 12 shows that inflating the prior by a factor of  $10^2$  immediately starts to degrade the inference on  $\Theta_E$ , although the degradation does level off past this point. This is likely because at an inflation level beyond  $100^2$ , the source unlensed trajectory is completely determined by the *HST* data and the source unlensed trajectory prior becomes uninformative.

More importantly however, shrinking the prior on the unlensed source trajectory does improve the posterior constraint on  $\Theta_E$ . Specifically, an improvement of a factor of  $10^2$  in the prior covariance of the source unlensed trajectory would lead to a maximum improvement in the 16th–84th posterior constraint on  $\Theta_E$  of  $\approx 1$  mas (corresponding to a 7% improvement on 84th–16th percentile constraint on  $M_L$ ). While an improvement of a factor of  $10^2$  of the source astrometry covariance matrix is unlikely to be achieved, this means that further astrometric data pinning down the unlensing source trajectory either by future *Gaia* data releases or *HST* will likely improve the posterior constraint on  $\Theta_E$ . Assuming a more realistic improvement of  $2^2$  on the source astrometry covariance matrix, corresponds to a 3% improvement on 16th–84th percentile constraint on  $M_L$ .

The left panel of Fig. 12 shows the effect of inflating or shrinking the standard deviation on the Gaussian prior for the noise parameters ( $\sigma_n$ ), both correlated and white. The means of these priors were informed by lens PSF subtraction simulations and WFC3/UVIS instrument precision, in the case of the correlated and white noise terms, respectively. The standard deviations of these priors, ( $\sigma_n$ ), were however chosen to be 0.1 mas. Fig. 12 shows that the analysis is sensitive to changing  $\sigma_n$ . Specifically, if  $\sigma_n$  is increased from

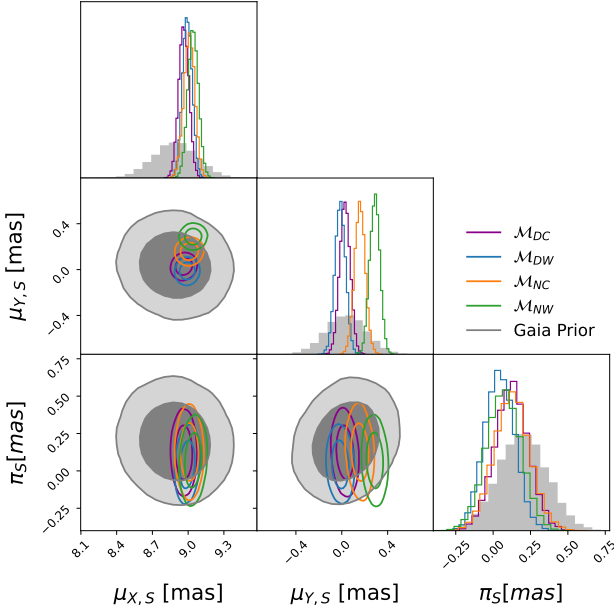
0.1 mas to 0.4 mas, the  $\Theta_E$  16th–84th posterior percentile degrades by  $\approx 1$  mas. In contrast, shrinking  $\sigma_n$  to 0.01 mas only marginally improves the  $\Theta_E$  16th–84th posterior percentile, by  $\approx 0.2$  mas. This sensitivity is reflected in the posterior distributions of the correlated noise parameters shown in Fig. 9. This is because for the majority of epochs the size of the correlated noise is not constrained by the data and defaults to the prior distribution, for the adopted value of  $\sigma_n = 0.1$ , indicating the priors are informative.

However, for epoch 7, the data does inform the value of the size of the correlated noise for  $\mathcal{M}_{DC}$  and the posterior is slightly shifted to higher values than the prior. This highlights the important balance between picking a reasonably tight prior to inform the size of the correlated noise but one that reflects reasonable uncertainty so the data can alter the value if there is constraining information. The adopted value of  $\sigma_n = 0.1$  mas in the analysis strikes a reasonable balance between our confidence in the size of the correlated noise at each epoch, whilst also giving sufficient flexibility for the data to further inform the size of the noise. Overall, the posterior constraint on  $\Theta_E$  is sensitive to our choice of  $\sigma_n = 0.1$  mas and our prior knowledge on the correlated component of the noise in general. We also investigated changing the boundaries of the flat prior on  $\Theta_E$  from 20 – 60 mas to 0 – 100 mas. We found this has no effect on any of the resulting reported parameter posterior percentiles.

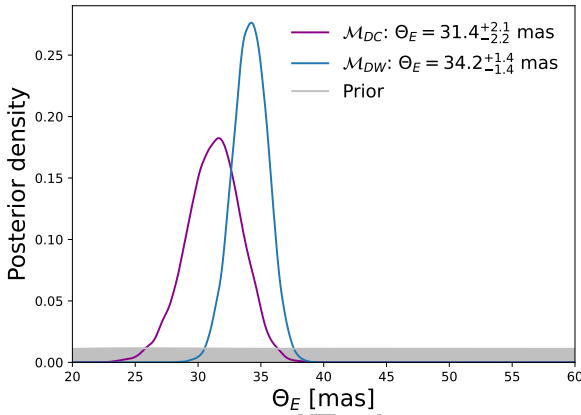
## 6 THE ASTROPHYSICS OF LAWD 37

At a distance of 4.6 pc, LAWD 37 (WD1142-645, 2MASS J11454297-6450297, HIP 57367, Gaia DR3 5332606522595645952,

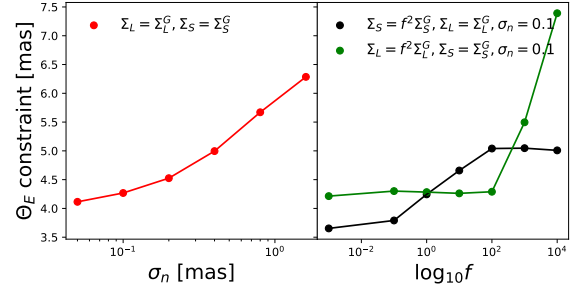




**Figure 10.** Comparison of the source GEDR3 astrometry used as a prior in all of the models compared with the posterior on the astrometry from each of the models. All panels show a probability density. For the 2D plots, the inner and outer contours contain 68% and 95% of the probability mass, respectively. The histograms show the marginal probability densities for each parameter. We have omitted the GEDR3 prior and model posteriors for the source reference positions ( $X_{0,S}$ ,  $Y_{0,S}$ ) because we found good agreement between GEDR3 priors and all the model posteriors (see Fig. B1 for the full corner plot).



**Figure 11.** Posterior distribution on  $\Theta_E$  for the deflection model with correlated noise  $M_{DC}$  and the deflection model with white noise  $M_{DW}$ . Reported  $\Theta_E$  values are the 50th posterior percentile and the 84th–16th posterior percentile uncertainties. The prior used for  $\Theta_E$  in both models is shown in grey.

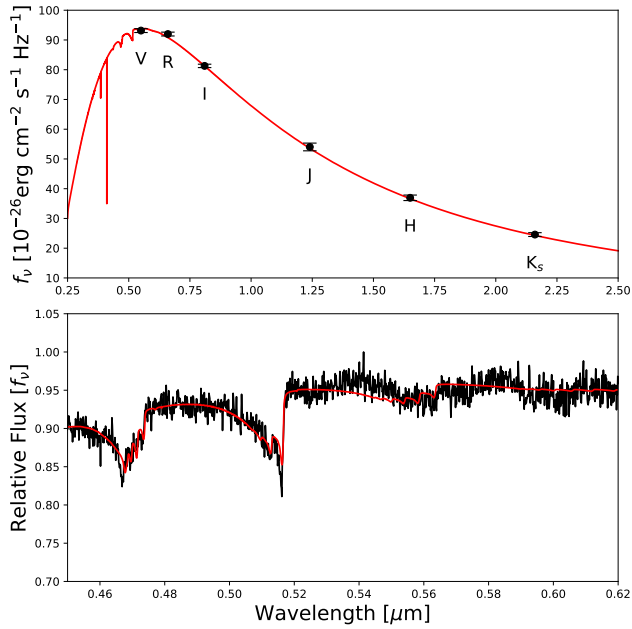


**Figure 12.** Sensitivity of the posterior constraint on  $\Theta_E$  (16th–84th posterior percentile) to changes in the priors in the deflection and correlated noise model  $M_{DC}$ . **Left:** Posterior constraint versus  $\sigma_n$ , the standard deviation on the Gaussian prior on each of the correlated noise and white noise parameters. In this case, the priors on the source and lens astrometry are fixed to the GEDR3  $\Sigma_G^S$  and  $\Sigma_G^L$ , respectively. **Right:** Posterior constraint versus altering the source or lens astrometric Gaussian prior covariance matrix by a multiplicative factor  $f$ , whilst keeping all other priors fixed.

WISEA J114547.32-645033.2), is the second nearest single white dwarf to the sun after van Maanen’s Star. Direct imaging of the region around LAWD 37 with *HST* shows no evidence of visible companions down to detection limits (Schroeder et al. 2000). Comparing Hipparcos and GDR2 astrometry, Kervella et al. (2019) identified a proper motion anomaly for LAWD 37 which could be explained by a massive companion. However, upon comparison between Hipparcos and the more precise GEDR3 astrometry for LAWD 37, this has largely been ruled out (Lindgren & Dravins 2021). Overall, at the present time it appears that LAWD 37 has no detectable companions.

Being a close by and bright target (V-band  $\approx 11.50$ ), LAWD 37 has been the subject of numerous studies (e.g., Koester & Weidemann 1982; Weidemann & Koester 1995; Dufour et al. 2005; Bergeron et al. 2001; Holberg et al. 2008; Subasavage et al. 2009; Giammichele et al. 2012; Sion et al. 2009; Subasavage et al. 2017; Coutu et al. 2019). As a result, a wealth of photometric and spectroscopic information has been gathered. LAWD 37 has a spectral type DQ indicating a helium dominated atmosphere and the presence of carbon lines in its spectrum. The presence of carbon in DQ white dwarf atmospheres is well explained by models of carbon settling and then being caught up by the helium convection zone bringing it to the surface (Bédard et al. 2022). In line with all helium rich white dwarfs LAWD 37 is expected to have no or trace amounts of hydrogen which is represented by a thin hydrogen layer in models,  $q_H = 10^{-10}$  (Dufour et al. 2005).

The atmospheric parameters of LAWD 37 are determined by fitting DQ-type white dwarf atmospheric models (Blouin et al. 2018; Blouin & Dufour 2019) to the broad-band photometry (V,R,I; Subasavage et al. 2017, J,H,Ks; 2MASS Skrutskie et al. 2006), and spectroscopy of LAWD 37. The solid angle and  $T_{\text{eff}}$  are first adjusted by fitting the model grid to the photometric data assuming an ad hoc photospheric carbon abundance. During this process, the surface gravity (and mass) of the star is calculated using the inferred solid angle of LAWD 37, its distance (from the GEDR3 parallax), and a theoretical mass—radius relationship suitable for CO core white dwarfs with thin hydrogen layers (Bédard et al. 2020). Once a photometric solution is found, the carbon abundance is then adjusted to match the spectroscopy while keeping the other parameters fixed. The photometric fit is then performed once again with this revised carbon abundance, and the whole process (photometric and spectroscopic fits) is repeated until a self-consistent solution is reached. We find



**Figure 13.** Best fit atmospheric model (red) over the photometric and spectroscopic data (black) of LAWD 37. **Top:** Photometric fit with VRI bands from Subasavage et al. (2017) and JHK<sub>s</sub> bands from 2MASS (Skrutskie et al. 2006). **Bottom:** Optical spectrum of LAWD 37 from (Subasavage et al. 2017). The photometry and spectroscopy of LAWD 37 were fit jointly using the method detailed in Section 4.1 of Subasavage et al. (2017), and we can see that the models and data are in agreement.

$T_{\text{eff}} = 7837^{+83}_{-82}$  K,  $\log g = 7.983 \pm 0.016$ ,  $R = 0.0127 \pm 0.0003 R_{\odot}$ ,  $M = 0.57 \pm 0.01 M_{\odot}$ , and  $\log[C/He] = -5.61 \pm 0.14$  (see Fig. 13). The use of the MRR in this fitting procedure mainly constrains  $\log g$  and has a minimal effect on  $T_{\text{eff}}$ , the solid angle, and therefore the photometric radius determination. This is because the energy distribution is not sensitive to  $\log g$ .

The direct gravitational mass determination from the astrometric microlensing caused by LAWD 37 is completely independent of all atmospheric and evolutionary modeling assumptions. This gravitational mass can therefore be used to test the theoretical models. Fig. 14 shows the position of LAWD 37 on the theoretical MRRs, obtained from the Montreal theoretical cooling tracks<sup>6</sup>, using the gravitational mass from the astrometric microlensing event. Fig. 14 shows excellent agreement between the gravitational microlensing mass from model  $M_{DC}$  and predicted value from the evolutionary theory of CO core white dwarfs,  $0.57 \pm 0.01 M_{\odot}$ , and assuming a thin hydrogen layer. Also plotted in Fig. 14 is the MRR for CO core white dwarfs with a thick hydrogen layer (not expected for LAWD 37). For comparison, also shown in Fig. 14, is the theoretical MRR for zero temperature white dwarfs with an iron (Fe) core (Hamada & Salpeter 1961), which is definitively ruled out by the microlensing mass.

Coutu et al. (2019) observed that the main peak of the mass distribution for DQ white dwarfs in their sample was shifted  $\approx 0.05 M_{\odot}$  below to that obtained for the DA and DB white dwarfs, suggesting

<sup>6</sup> Obtained from interpolation of the Montreal theoretical cooling tracks available at <https://www.astro.umontreal.ca/~bergeron/CoolingModels/>.

that there could be a problem with the DQ models used in their mass determinations (the same models as those used here). This suspicion was supported by comparing their mass for Procyon B determined with the photometric/spectroscopic method ( $0.554 M_{\odot}$ ) with the dynamical mass determination of  $0.592 M_{\odot}$  determined by Bond et al. (2015). An alternative explanation for the discrepancy between the mass distributions of DQ white dwarfs and DA/DB stars has recently been offered by (Bédard et al. 2022, see their Section 4.3) who argued that although all DB stars likely experience carbon dredge-up, this phenomenon is more significant — and thus more easily detected — in lower-mass objects, thus explaining the relative deficit of high-mass DQ stars. The good agreement between our gravitational mass for LAWD 37 and the hybrid photometric/spectroscopic value reported here yields support to this interpretation, although the uncertainty on the microlensing mass remains too large to rule out any systematic problem with the current DQ model atmospheres, and this work centers around the analysis of only one object.

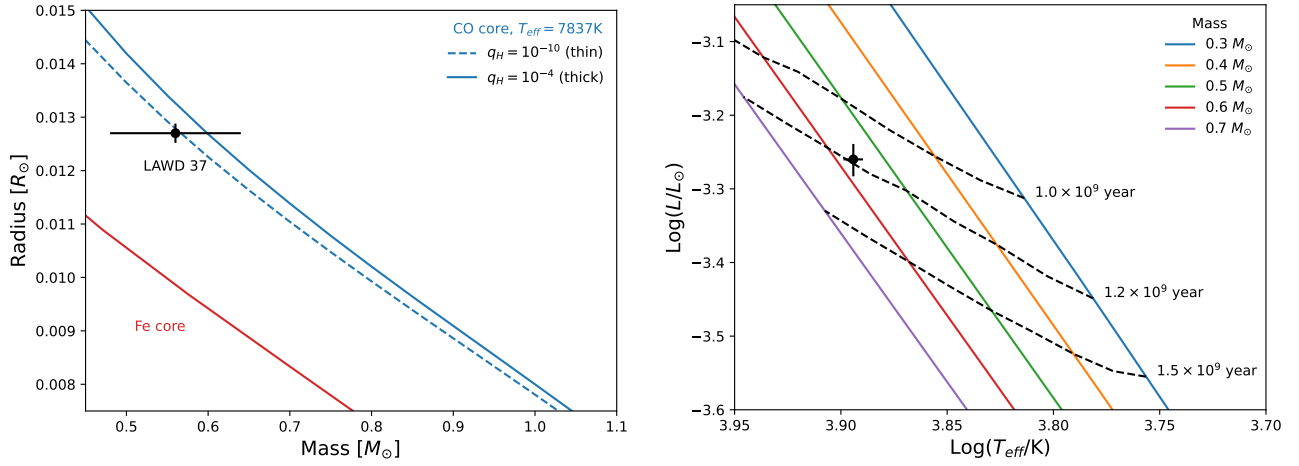
Fig. 14 shows the position of LAWD 37 on the theoretical Hertzsprung–Russell diagram (luminosity versus effective temperature) along with theoretical cooling tracks (assuming  $q_H = 10^{-10}$ ) from the Montreal database for a range of masses. The position of LAWD 37 in the Hertzsprung–Russell diagram is in agreement with the expected position of a white dwarf with the inferred microlensing mass of  $0.56 \pm 0.08 M_{\odot}$ . Also shown in Fig. 14 are isochrones for 1.0, 1.2, and  $1.5 \times 10^9$  years. By interpolation of the Montreal theoretical cooling tracks, the implied cooling age of LAWD 37 is  $1.15 \pm 0.04 \times 10^9$  years.

## 7 DISCUSSION AND CONCLUSIONS

We have analyzed *HST* follow-up data of a predicted astrometric microlensing event caused by the nearby white dwarf, LAWD 37. Specifically, we used WFC3/UVIS *HST* astrometric data of the source in combination with *Gaia* astrometry (GEDR3) of the source and lens to infer a gravitational mass for LAWD 37 of  $0.56 \pm 0.08 M_{\odot}$ . We consider and fit four different models to the data. Models with and without the astrometric deflection term, and then each with and without a correlated noise component due to the lens PSF subtraction within an epoch ( $M_{DC}$ ,  $M_{DW}$ ,  $M_{NC}$ ,  $M_{NW}$ ). We find the model with the deflection term and correlated noise ( $M_{DC}$ ) best explains the data according to the overall LOO score.

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

The model with the deflection and just white noise,  $M_{DW}$ , provides a comparatively poor explanation of the data according to the LOO score.  $M_{DW}$  fails to explain the data in epoch 5 where there is a predicted high amount of correlated noise from the lens PSF subtractions. The failure in epoch 5 is clear, when compared to  $M_{NC}$ , despite  $M_{DW}$  increasing the size of the white noise and epoch 5 having a large deflection signal. Consequently, because the correlated



**Figure 14.** Left: Comparison of the mass of LAWD 37 inferred from the astrometric lensing event with theoretical MRR relationships for white dwarfs. Blue lines show the MRR for CO core white dwarfs with a thin ( $q_H = 10^{-10}$ ), and thick ( $q_H = 10^{-4}$ ) hydrogen layer, and effective temperature equal to that of LAWD 37. For comparison, also shown is the theoretical MRR for a zero temperature white dwarf with an iron (Fe) core (Hamada & Salpeter 1961). Right: Hertzsprung-Russell diagram for LAWD 37. Coloured lines show the Montreal cooling tracks for a selection of white dwarf masses. Dashed lines indicate isochrones. The position of LAWD 37 agrees with the inferred microlensing mass of  $0.56 \pm 0.08 M_{\odot}$ . The implied age of LAWD 37 is  $1.15 \pm 0.04 \times 10^9$  years.

noise can mimic the deflection signal,  $M_{DW}$  interprets some of the unmodeled correlated noise as additional deflection signal and infers a high mass for LAWD 37 of  $M_L = 0.66^{+0.06}_{-0.05} M_{\odot}$ . Overall, we rule out this model, and its mass inference, due to its poor LOO score compared with  $M_{DC}$  and  $M_{NC}$ .

We performed checks on the sensitivity of our inferences in our chosen model,  $M_{DC}$ , to our prior modeling assumptions. We find that, for  $M_{DC}$ , our inference on  $M_L$  is not sensitive to the lens GEDR3 prior. This means that collecting additional follow-up astrometric data on the lens (LAWD 37) to improve its astrometric solution is unlikely to improve the inference on  $M_L$ . In contrast, we find that, for  $M_{DC}$ , the inference on  $M_L$  is sensitive to the GEDR3 prior on the source astrometry. Therefore, we conclude that collecting additional astrometric data to further constrain the source astrometric solution is likely to improve the inference on  $M_L$  by  $\approx 3\%$ . For the noise parameters in  $M_{DC}$ , both correlated and white, we found our inference on  $M_L$  is sensitive to their chosen prior distributions. We found that simultaneously tightening the prior distributions on all the noise parameters does not significantly improve the constraint on  $M_L$ .

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

Using the posterior distribution of  $M_{DC}$ , we note our results are consistent with the event parameters originally predicted by McGill et al. (2018). The inferred event parameters (vs the predictions from McGill et al. 2018) are: a closest lens-source approach of  $396.65 \pm 0.17$  ( $380 \pm 10$ ) mas, a time of closest lens-source approach of  $2019.86422^{+0.00004}_{-0.00005}$  ( $2019.86 \pm 0.01$ ) Julian year, and a maxi-

mum astrometric shift due to lensing of  $2.46 \pm 0.34$  ( $2.8 \pm 0.1$ ) mas. The high predicted shift can be attributed to the high photometric mass estimate for LAWD 37 ( $0.61 M_{\odot}$ ) used in McGill et al. (2018). The difference in predicted closest approach can be explained by the poorer quality source astrometry used by McGill et al. (2018), which was the only source astrometry available at the time.

In conclusion, the gravitational mass for LAWD 37 obtained via astrometric microlensing in this work is in agreement with theoretical MRR and cooling tracks expected from the evolutionary theory of CO core white dwarfs. LAWD 37 is also predicted to cause many more astrometric microlensing events over the coming decades (Bramich 2018; Klüter et al. 2018b), which may offer further opportunities to increase the precision of the gravitational mass measurement. This work provides the first ever semi-empirical test of the white dwarf MRR for a single, isolated white dwarf and lends support to current white dwarf evolutionary theory. This work also marks only the third time that the astrometric microlensing effect has been detected via the prediction channel.

This analysis reveals how the followup of future predicted astrometric microlensing events (Bramich 2018; Bramich & Nielsen 2018; Klüter et al. 2018b; Nielsen & Bramich 2018b) could be improved. Generally, predicted microlensing events permit targeted and optimised followup campaigns (e.g., Sahu et al. 2017; Zurlo et al. 2018; McGill et al. 2019b), because the time of maximum signal can be predicted and hence measurements can be clustered around it. This seems like a sensible approach in the first instance. However, typical predictable microlensing events are caused by nearby bright lenses (Dominik & Sahu 2000). This means usually we will encounter a scenario similar to this analysis, i.e. where the lens PSF subtraction introduces significant correlated noise into the data. This correlated noise is usually worse at closer lens-source separations (see e.g., Table 2), when the deflection signal is largest. Moreover, this analysis shows that the correlated noise can mimic the astrometric deflection signal. Ultimately this meant that epoch 2 in this analysis, which is in the tails of the event, was critical for measuring the deflection signal, more so than the data at the predicted maximum. The utility of epoch 2 was due to a high signal to noise offset in a direction that could not be explained by alteration of the source’s astrometric parameters.



With the power of hindsight, the importance of taking observations at epoch 2's time could have been determined ahead of planning the *HST* observations. This is because all of the lensing systems parameters (lens and source astrometry and lens mass estimate) were already known, and the size of the correlated noise introduced by the lens PSF subtraction could be roughly estimated ahead of time. However, at the time of closest approach, LAWD 37 was at the edge of the “sun avoidance zone”, which caused some increased noise, and also a shortened visibility period during an orbit. The small range of available ORIENTs at this configuration made it impossible to keep the diffraction spike too far from the source. This turned out to be the most important factor in elevating the noise, and there was no way to avoid this. This aspect of the observation planning is also extremely difficult to simulate in advance.

In principle, an observation strategy with shorter intervals between epochs, and with fewer exposures per epoch could be used. This would provide more constraints on the source astrometry at a greater range of epochs and hence a degree of redundancy of epoch in the case that residuals prove to be irremovable. However, in the case of followup campaigns carried out with *HST*, including the followup of the event presented in this paper, one orbit is the minimum that can be used per visit. A future line of research could work out if *HST* (or perhaps even the *James Webb Space Telescope*; Gardner et al. 2006) observation times which optimally constrain the lens mass and/or rule out purely correlated noise explanations of the data can be determined ahead of time. This could be tested on future predicted microlensing events.

## ACKNOWLEDGEMENTS

We would like to thank the two referees for their detailed reports that both improved the clarity of the paper. Based in part on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at STScI, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. Support for this research was provided by NASA through grants from STScI. *HST* data used in this paper are available from the Mikulski Archive for Space Telescopes at STScI (<https://archive.stsci.edu/hst/search.php>) under proposal IDs 15705, 15961 and 16251. We thank Howard Bond and Ed Nelan for their help at various stages of the project. We thank Daniel Bramich, Tom Brown, Ed Nelan, Łukasz Wyrzykowski, and Paul Hewett for useful comments that improved the clarity of the paper. PM acknowledges studentship funding support from the United Kingdom Science and Technology Facilities Council (STFC) and the Cambridge Centre for Doctoral Training in Data Intensive Science (CDT-DIS). MBN acknowledges support from the UK Space Agency. This research is based on observations made with the NASA/ESA *Hubble Space Telescope* obtained from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* Multi-Lateral Agreement (MLA). The *Gaia* mission website is <https://www.cosmos.esa.int/gaia>. The *Gaia* Archive website is <http://archives.esac.esa.int/gaia>.

## DATA AVAILABILITY

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

## REFERENCES

- Alcock C., et al., 2000, *ApJ*, 542, 281
- Anderson J., 2016, Empirical Models for the WFC3/IR PSF, Instrument Science Report WFC3 2016-12, 42 pages
- Anderson J., 2018, Focus-Diverse PSFs for Five Commonly Used WFC3/UVIS Filters, Instrument Science Report WFC3 2018-14, 19 pages
- Anderson J., 2021, Table-Based CTE Corrections for flt-Format WFC3/UVIS, Instrument Science Report WFC3 2021-13
- Andrae R., Schulze-Hartung T., Melchior P., 2010, arXiv e-prints, p. [arXiv:1012.3754](https://arxiv.org/abs/1012.3754)
- Arenou F., et al., 2018, *A&A*, 616, A17
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Bédard A., Bergeron P., Fontaine G., 2017, *ApJ*, 848, 11
- Bédard A., Bergeron P., Brassard P., Fontaine G., 2020, *ApJ*, 901, 93
- Bédard A., Bergeron P., Brassard P., 2022, arXiv e-prints, p. [arXiv:2203.12045](https://arxiv.org/abs/2203.12045)
- Bellini A., Anderson J., Bedin L. R., 2011, *PASP*, 123, 622
- Bellini A., et al., 2014, *ApJ*, 797, 115
- Bellini A., Anderson J., Bedin L. R., King I. R., van der Marel R. P., Piotto G., Cool A., 2017, *ApJ*, 842, 6
- Belokurov V. A., Evans N. W., 2002, *MNRAS*, 331, 649
- Bennett D. P., et al., 2018, *AJ*, 156, 113
- Bergeron P., Leggett S. K., Ruiz M. T., 2001, *ApJS*, 133, 413
- Bergeron P., Dufour P., Fontaine G., Coutu S., Blouin S., Genest-Beaulieu C., Bédard A., Rolland B., 2019, *ApJ*, 876, 67
- Blouin S., Dufour P., 2019, *MNRAS*, 490, 4166
- Blouin S., Dufour P., Allard N. F., 2018, *ApJ*, 863, 184
- Bond I. A., et al., 2004, *ApJ*, 606, L155
- Bond H. E., et al., 2015, *ApJ*, 813, 106
- Bond H. E., et al., 2017a, *ApJ*, 840, 70
- Bond H. E., Bergeron P., Bédard A., 2017b, *ApJ*, 848, 16
- Bramich D. M., 2018, *A&A*, 618, A44
- Bramich D. M., Nielsen M. B., 2018, *Acta Astron.*, 68, 183
- Bürkner P.-C., Gabry J., Vehtari A., 2020, *Computational Statistics*, pp 1–19
- Chandra V., Hwang H.-C., Zakamska N. L., Cheng S., 2020, *ApJ*, 899, 146
- Chandrasekhar S., 1935, *MNRAS*, 95, 207
- Chwolson O., 1924, *Astronomische Nachrichten*, 221, 329
- Coutu S., Dufour P., Bergeron P., Blouin S., Loranger E., Allard N. F., Dunlap B. H., 2019, *ApJ*, 885, 74
- Dominik M., Sahu K. C., 2000, *ApJ*, 534, 213
- Dong S., et al., 2019, *ApJ*, 871, 70
- Dufour P., Bergeron P., Fontaine G., 2005, *ApJ*, 627, 404
- Einstein A., 1916, *Annalen der Physik*, 354, 769
- Einstein A., 1936, *Science*, 84, 506
- Everall A., Boubert D., Koposov S. E., Smith L., Holl B., 2021, *MNRAS*, 502, 1908
- Fabircius C., et al., 2016, *A&A*, 595, A3
- Falcon R. E., Winget D. E., Montgomery M. H., Williams K. A., 2010, *ApJ*, 712, 585
- Feibelman W. A., 1966, *Science*, 151, 73
- Feibelman W. A., 1986, *PASP*, 98, 1199
- Fong E., Holmes C., 2020, *Biometrika*, 107, 489
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Foreman-Mackey D., et al., 2021, *The Journal of Open Source Software*, 6, 3285
- Gaia Collaboration et al., 2016a, *A&A*, 595, A1
- Gaia Collaboration et al., 2016b, *A&A*, 595, A2
- Gaia Collaboration et al., 2018, *A&A*, 616, A1

- Gaia Collaboration et al., 2021, *A&A*, **649**, A1
- Gardner J. P., et al., 2006, *Space Sci. Rev.*, **123**, 485
- Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A., Rubin D. B., 2013, Bayesian data analysis. CRC press
- Gelman A., Hwang J., Vehtari A., 2014, *Statistics and computing*, **24**, 997
- Giammichele N., Bergeron P., Dufour P., 2012, *ApJS*, **199**, 29
- Hamada T., Salpeter E. E., 1961, *ApJ*, **134**, 683
- Hansen B. M. S., et al., 2002, *ApJ*, **574**, L155
- Harding A. J., Stefano R. D., Lépine S., Urama J., Pham D., Baker C., 2018, *MNRAS*, **475**, 79
- Higson E., Handley W., Hobson M., Lasenby A., 2019, *Statistics and Computing*, **29**, 891
- Høg E., Novikov I. D., Polnarev A. G., 1995, *A&A*, **294**, 287
- Holberg J. B., Sion E. M., Oswalt T., McCook G. P., Foran S., Subasavage J. P., 2008, *AJ*, **135**, 1225
- Homan M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, **15**, 1593–1623
- Husseiniouva A., McGill P., Smith L. C., Evans N. W., 2021, *MNRAS*, **506**, 2482
- Iben I. J., Tutukov A. V., 1984, *ApJ*, **282**, 615
- Joyce S. R. G., Barstow M. A., Holberg J. B., Bond H. E., Casewell S. L., Burleigh M. R., 2018, *MNRAS*, **481**, 2361
- Kaczmarek Z., McGill P., Evans N. W., Smith L. C., Wyrzykowski L., Howil K., Jabłońska M., 2022, *MNRAS*, **514**, 4845
- Kains N., et al., 2017, *ApJ*, **843**, 145
- Kervella P., Arenou F., Mignard F., Thévenin F., 2019, *A&A*, **623**, A72
- Kilic M., Bergeron P., Kosakowski A., Brown W. R., Agüeros M. A., Blouin S., 2020, *ApJ*, **898**, 84
- Kim S.-L., et al., 2016, *Journal of Korean Astronomical Society*, **49**, 37
- Klüter J., Bastian U., Demleitner M., Wambsgans J., 2018a, *A&A*, **615**, L11
- Klüter J., Bastian U., Demleitner M., Wambsgans J., 2018b, *A&A*, **620**, A175
- Klüter J., Bastian U., Demleitner M., Wambsgans J., 2021, arXiv e-prints, p. arXiv:2112.12152
- Koester D., Weidemann V., 1982, *A&A*, **108**, 406
- Kumar R., Carroll C., Hartikainen A., Martin O., 2019, *The Journal of Open Source Software*, **4**, 1143
- Lam C. Y., et al., 2022, *ApJ*, **933**, L23
- Lépine S., DiStefano R., 2012, *ApJ*, **749**, L6
- Lindgren L., Dravins D., 2021, *A&A*, **652**, A45
- Lindgren L., et al., 2021, *A&A*, **649**, A2
- Luberto J., Martin E. C., McGill P., Leauthaud A., Skemer A. J., Lu J. R., 2022, arXiv e-prints, p. arXiv:2210.14414
- McGill P., Smith L. C., Evans N. W., Belokurov V., Smart R. L., 2018, *MNRAS*, **478**, L29
- McGill P., Smith L. C., Evans N. W., Belokurov V., Zhang Z. H., 2019a, *MNRAS*, **483**, 4210
- McGill P., Smith L. C., Evans N. W., Belokurov V., Lucas P. W., 2019b, *MNRAS*, **487**, L7
- McGill P., Everall A., Boubert D., Smith L. C., 2020, *MNRAS*, **498**, L6
- Miralda-Escude J., 1996, *ApJ*, **470**, L113
- Miyamoto M., Yoshii Y., 1995, *AJ*, **110**, 1427
- Mróz P., et al., 2019, *ApJS*, **244**, 29
- Mustill A. J., Davies M. B., Lindgren L., 2018, *A&A*, **617**, A135
- Nielsen M. B., Bramich D. M., 2018a, *Acta Astron.*, **68**, 351
- Nielsen M. B., Bramich D. M., 2018b, *Acta Astron.*, **68**, 351
- Ofek E. O., 2018, *ApJ*, **866**, 144
- Paczynski B., 1995, *Acta Astron.*, **45**, 345
- Paczynski B., 1996, *Acta Astron.*, **46**, 291
- Parsons S. G., et al., 2016, *MNRAS*, **458**, 2793
- Parsons S. G., et al., 2017, *MNRAS*, **470**, 4473
- Pasquini L., Pala A. F., Ludwig H. G., Leao I. C., de Medeiros J. R., Weiss A., 2019, *A&A*, **627**, L8
- Proft S., 2016, PhD thesis, University of Heidelberg
- Proft S., Demleitner M., Wambsgans J., 2011, *A&A*, **536**, A50
- Refsdal S., 1964, *MNRAS*, **128**, 295
- Rybicki K. A., Wyrzykowski L., Klencki J., de Bruijne J., Belczyński K., Chruślińska M., 2018, *MNRAS*, **476**, 2013
- Sabbi E., Bellini A., 2013, UVIS PSF Spatial & Temporal Variations, Instrument Science Report WFC3 2013-011, 10 pages
- Sahu K. C., Chaney E., Graham J., Kane S., Wieldt D., 1998, in American Astronomical Society Meeting Abstracts #192. p. 07.01
- Sahu K. C., Bond H. E., Anderson J., Dominik M., 2014, *ApJ*, **782**, 89
- Sahu K. C., et al., 2017, *Science*, **356**, 1046
- Sahu K. C., et al., 2022, *ApJ*, **933**, 83
- Salim S., Gould A., 2000, *ApJ*, **539**, 241
- Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, *PeerJ Computer Science*, **2**, e55
- Schmidt H., 1996, *A&A*, **311**, 852
- Schroeder D. J., et al., 2000, *AJ*, **119**, 906
- Sion E. M., Holberg J. B., Oswalt T. D., McCook G. P., Wasatonic R., 2009, *AJ*, **138**, 1681
- Skilling J., et al., 2006, *Bayesian analysis*, **1**, 833
- Skrutskie M. F., et al., 2006, *AJ*, **131**, 1163
- Smith M. C., Belokurov V., Evans N. W., Mao S., An J. H., 2005, *MNRAS*, **361**, 128
- Speagle J. S., 2020, *MNRAS*, **493**, 3132
- Subasavage J. P., Jao W.-C., Henry T. J., Bergeron P., Dufour P., Janna P. A., Costa E., Méndez R. A., 2009, *AJ*, **137**, 4547
- Subasavage J. P., et al., 2017, *AJ*, **154**, 32
- Sundararajan S., Keerthi S. S., 2001, *Neural computation*, **13**, 1103
- Tremblay P. E., Bergeron P., 2008, *ApJ*, **672**, 1144
- Tremblay P.-E., et al., 2017, *MNRAS*, **465**, 2849
- Urban S. E., Seidelmann P. K., 2014, in American Astronomical Society Meeting Abstracts #223. p. 247.20
- Vehtari A., Simpson D., Gelman A., Yao Y., Gabry J., 2015, arXiv preprint arXiv:1507.02646
- Vehtari A., Gelman A., Gabry J., 2017, *Statistics and computing*, **27**, 1413
- Vehtari A., Gelman A., Simpson D., Carpenter B., Bürkner P.-C., 2021, *Bayesian analysis*, **1**, 1
- Walker M. A., 1995, *ApJ*, **453**, 37
- Weidemann V., Koester D., 1995, *A&A*, **297**, 216
- Zurlo A., et al., 2018, *MNRAS*, **480**, 236

## APPENDIX A: LEAVE-ONE-OUT CROSS-VALIDATION APPROXIMATION

In this Section we explain how the LOO PSIS approximation is applied to the models considered in this work. This section closely follows Vehtari et al. (2017) and Bürkner et al. (2020). The goal of the approximation, given that we have obtained  $S$  posterior samples from fitting  $\mathcal{M}_{TN}$  to the full data set  $\mathcal{D}$ ,  $\{\theta^s\}_{s=1}^S$ , we wish to approximate,

$$p(D_i | \mathcal{D}_{-i}, \mathcal{M}_{TN}) = \int p(D_i | \theta, \mathcal{D}_{-i}, \mathcal{M}_{TN}) p(\theta | \mathcal{D}_{-i}, \mathcal{M}_{TN}) d\theta. \quad (\text{A1})$$

Specifically, we want to use  $\{\theta^s\}_{s=1}^S$  to approximately evaluate Eq. (A1) instead of having to refit the model (which is expensive) with the  $i$ th data point left-out to obtain  $\{\theta_{-i}^s\}_{s=1}^S$ . This is achieved by using an importance sampling approximation and re-weighting  $\{\theta^s\}_{s=1}^S$  accordingly.

Importance sampling is a Monte Carlo method used to compute expectations. We would like to evaluate the expectation,

$$\mathbb{E}_f[h(\theta)] = \int h(\theta) f(\theta) d\theta, \quad (\text{A2})$$

where  $f$  is some hard-to-sample-from distribution. Instead of using samples from  $f$  we can use samples from an easier-to-sample-from proposal distribution  $g$ ,  $\{\theta_g^s\}_{s=1}^S$ , provided we know the the ratio between  $f$  and  $g$  which is  $r(\theta_g^s) = f(\theta_g^s)/g(\theta_g^s)$ . The importance

sampling approximation is then,

$$\mathbb{E}_f[h(\theta)] \approx \frac{\sum_{s=1}^S r(\theta_g^s) h(\theta_g^s)}{\sum_{s=1}^S r(\theta_g^s)}. \quad (\text{A3})$$

The quality of this approximation is sensitive to the distribution of the importance weights,  $r(\theta_g^s)$ . If the proposal distribution,  $g$ , is not representative of the target distribution  $f$ , then the importance weights become unstable. This instability comes from the importance weights being dominated by a few extreme values which leads to them having a large or infinite variance. Ultimately this leads to a poor importance sampling approximation (Vehtari et al. 2015). To mitigate this problem,  $r(\theta_g^s)$  are fitted to a Pareto distribution and the extreme values are removed and are replaced with draws from the fitted Pareto distribution,  $\tilde{r}(\theta_g^s)$ . These smoothed weights,  $\tilde{r}(\theta_g^s)$ , are dropped in as a replacement for  $r(\theta_g^s)$  in Eq. (A3). This is called Pareto Smoothed Importance Sampling (PSIS; Vehtari et al. 2015). In addition to stabilising the importance sample approximation, PSIS also provides a diagnostic on the quality of the importance sampling approximation. This diagnostic is the fitted shape parameter,  $k$ , of the Pareto distribution.  $k$  traces the number of finite moments of the importance weights distribution and therefore the quality of the PSIS approximation. Vehtari et al. (2015) find that  $k < 0.7$  indicates PSIS will work well whereas a value of  $k > 0.7$  indicates the PSIS approximation is likely to be poor and should not be used.

For approximating the LOO score, our hard-to-sample-from target distribution is  $p(\theta|\mathcal{D}_{-i}, \mathcal{M}_{TN})$ . This is the posterior distribution obtained when fitting the model to the data set with the  $i$ th data point left-out. This is hard to sample from because it is expensive and we would have to refit the model, which we want to avoid.  $h$  is  $p(D_i|\theta, \mathcal{D}_{-i}, \mathcal{M}_{TN})$ . Our easy to sample from proposal distribution is  $p(\theta|\mathcal{D}, \mathcal{M}_{TN})$  which we readily have samples from -  $\{\theta^s\}_{s=1}^S$ . The only remaining task is to compute the ratio between the target and proposal distributions.

$$r(\theta) = \frac{p(\theta|\mathcal{D}_{-i}, \mathcal{M}_{TN})}{p(\theta|\mathcal{D}, \mathcal{M}_{TN})} \propto \frac{p(\theta|\mathcal{M}_{TN})p(\mathcal{D}_{-i}|\theta, \mathcal{M}_{TN})}{p(\theta|\mathcal{M}_{TN})p(\mathcal{D}|\theta, \mathcal{M}_{TN})} \quad (\text{A4})$$

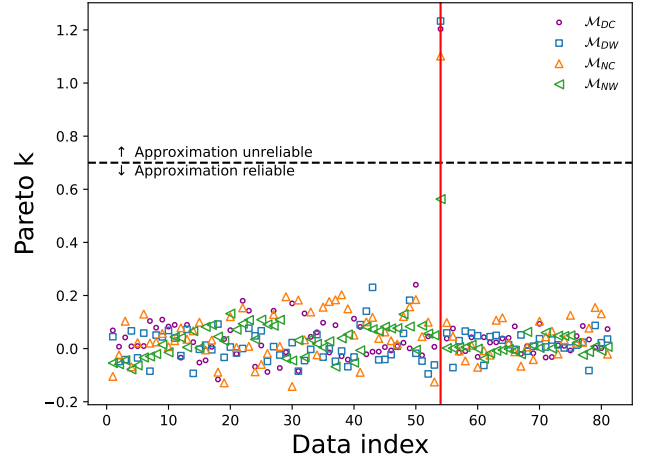
$$= \frac{p(\mathcal{D}_{-i}|\theta, \mathcal{M}_{TN})}{p(\mathcal{D}|\theta, \mathcal{M}_{TN})}. \quad (\text{A5})$$

Here we have used Bayes rule which states that the ratio of the posteriors is proportional to the ratio of the prior  $\times$  likelihood, and the priors terms have cancelled. For all models the likelihood factorises over the different epochs (Eq. 10). If the left-out data point,  $D_i$ , is the  $j$ th data point in epoch  $\tilde{e}$ , all other epochs cancel from the likelihood ratio in Eq. (A5) and,

$$r(\theta) \propto \frac{p(\mathcal{D}_{-i}|\theta, \mathcal{M}_{TN})}{p(\mathcal{D}|\theta, \mathcal{M}_{TN})} = \frac{1}{p(D_{\tilde{e},j}|D_{\tilde{e},-j}, \theta, \mathcal{M}_{TN})}. \quad (\text{A6})$$

The ratio is proportional to the inverse of the likelihood of the left-out data point (Bürkner et al. 2020). In all models considered in this work,  $p(D_{\tilde{e},j}|D_{\tilde{e},-j}, \theta, \mathcal{M})$  is the product of two multivariate Gaussian distributions (in the  $X$  and  $Y$  directions) as described in Eq. (9). Using the results from Sundararajan & Keerthi (2001), Bürkner et al. (2020) and Eq. (9), this likelihood can be computed efficiently as,

$$\begin{aligned} \log p(D_{\tilde{e},j}|D_{\tilde{e},-j}, \theta, \mathcal{M}_{TN}) &= -\log 2\pi\tilde{\sigma}_{\tilde{e},j} \\ &\quad - \frac{1}{2\tilde{\sigma}_{\tilde{e},j}} \left( \left( [X_{\tilde{e}}]_j - \tilde{X}_{\tilde{e},j} \right)^2 \right. \\ &\quad \left. + \left( [Y_{\tilde{e}}]_j - \tilde{Y}_{\tilde{e},j} \right)^2 \right). \end{aligned} \quad (\text{A7})$$



**Figure A1.** Pareto  $k$  values in the PSIS-LOO approximation used for each model. Different colours indicate different models. Different sized markers are used so points on top of each other can be differentiated. Dashed horizontal line indicates the threshold value of Pareto  $k$  (0.7) used to judge if the importance sampling approximation is reliable. For all models, the importance sampling approximation fails for the data point with index 54 indicated as a vertical red line. To obtain the LOO score for data point 54 the model was refit.

Here,  $[\bullet]_j$  is the  $j$ th component of vector  $\bullet$ . The conditional means,  $\tilde{X}_{\tilde{e},j}$  and  $\tilde{Y}_{\tilde{e},j}$ , and conditional standard deviation  $\tilde{\sigma}_{\tilde{e},j}$  are given by Bürkner et al. (2020) as,

$$\tilde{X}_{\tilde{e},j} = [X_{\tilde{e}}]_j - \frac{[(\Sigma_{\tilde{e}}^N)^{-1}(X_{\tilde{e}} - X_{\tilde{e}}^T)]_j}{[(\Sigma_{\tilde{e}}^N)^{-1}]_{jj}}, \quad (\text{A8})$$

$$\tilde{Y}_{\tilde{e},j} = [Y_{\tilde{e}}]_j - \frac{[(\Sigma_{\tilde{e}}^N)^{-1}(Y_{\tilde{e}} - Y_{\tilde{e}}^T)]_j}{[(\Sigma_{\tilde{e}}^N)^{-1}]_{jj}}, \quad (\text{A9})$$

and,

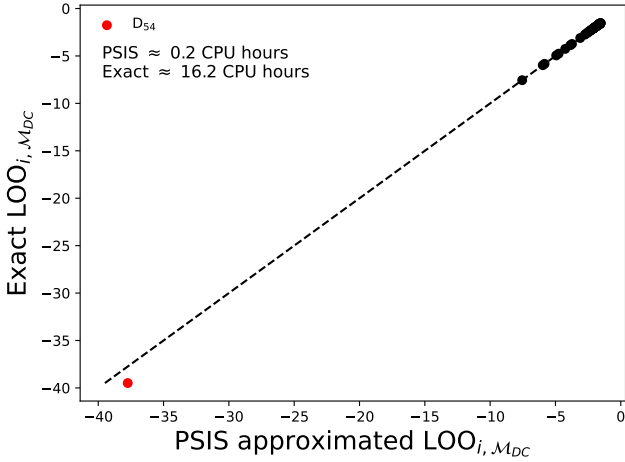
$$\tilde{\sigma}_{\tilde{e},j} = \frac{1}{[(\Sigma_{\tilde{e}}^N)^{-1}]_{jj}}. \quad (\text{A10})$$

Here, we have dropped the dependence of  $\Sigma_{\tilde{e}}^N$ ,  $X_{\tilde{e}}^T$  and  $Y_{\tilde{e}}^T$  on  $\theta$  for brevity.  $[\bullet]_{jj}$  denotes the  $j$  diagonal element of the matrix  $\bullet$ .

Using these weights with PSIS, we can then approximate Eq. (A1) for each left-out data point using the posterior samples from the full dataset,  $\{\theta^s\}_{s=1}^S$ . For each left-out data point, we obtain the Pareto  $k$  diagnostic. If for any left-out data point  $k > 0.7$ , we do not use the PSIS approximation, and instead perform a full refit to evaluate Eq. (A1) exactly. Fig. A1 shows the Pareto  $k$  diagnostic for each left-out data point for all models. For every model the approximation was reliable for all left-out data points apart from  $D_{53}$ . For all models we therefore computed this LOO term exactly by refitting the model with  $D_{53}$  left-out. Overall, with PSIS, we obtained an approximate LOO score for each model with only one additional refit of the model compared to the 81 refits required to compute LOO exactly.

To further check the safety of the PSIS approximation, we computed all LOO terms for the model,  $\mathcal{M}_{DC}$ , exactly and compared them to the PSIS approximation. Fig. A2 shows the comparison of the PSIS approximation and the exactly computed LOO terms from  $\mathcal{M}_{DC}$ . The PSIS approximation matches the exact LOO computation





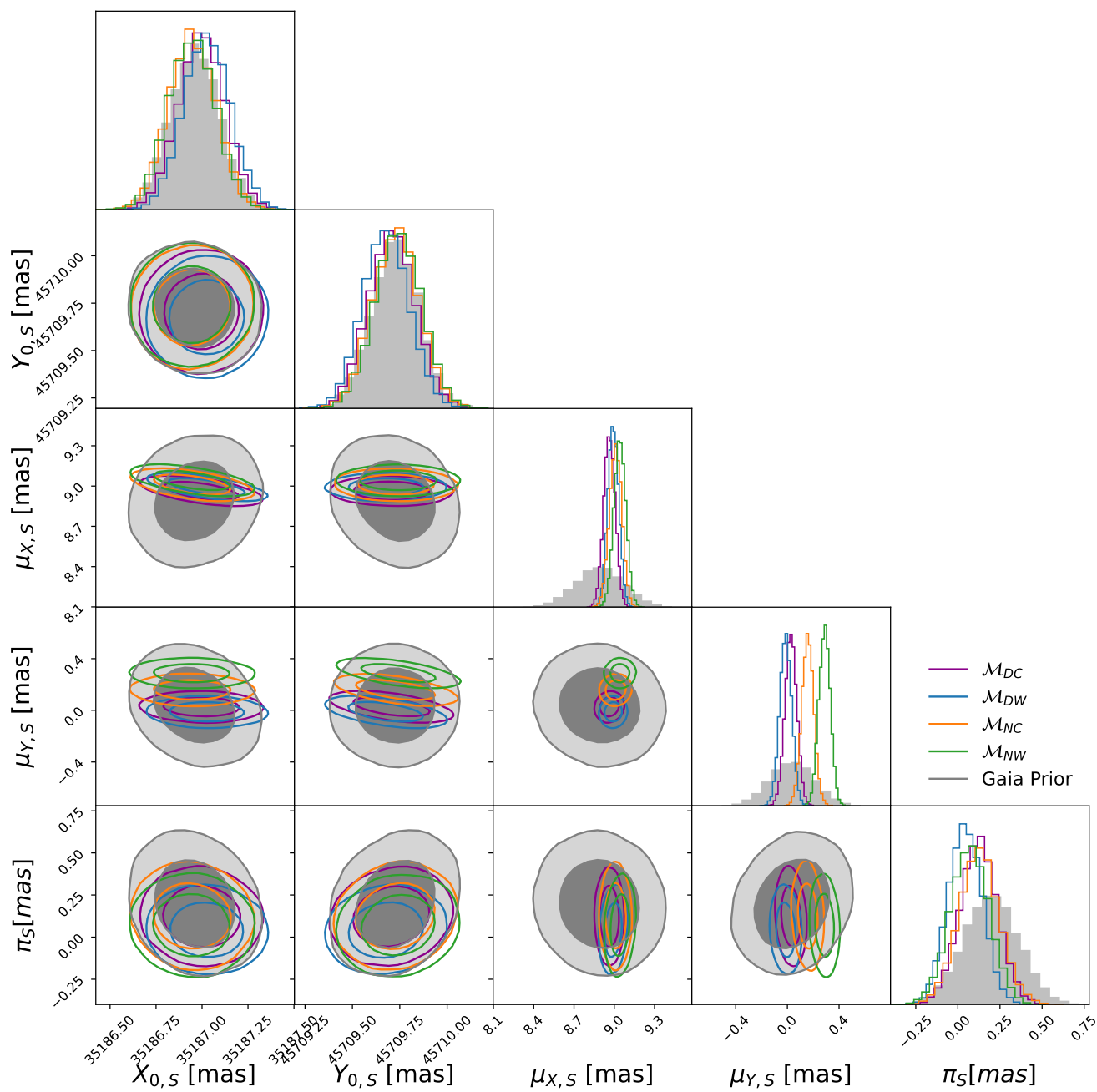
**Figure A2.** PSIS approximation versus exact computation of the LOO terms for  $M_{DC}$ . The PSIS is in good agreement with the exact values. The dashed line is the one-to-one relationship. The data point that failed the PSIS approximation  $D_{53}$ , is marked in red. The approximate CPU computation times for the exact and PSIS methods are shown as text.

well for all left-out data points apart from  $D_{54}$ . Encouragingly,  $D_{54}$  is the same data point that was flagged by PSIS as the approximation being unreliable (Fig. A1). Consequently, the model was refitted and the exact LOO value was used for  $D_{54}$ . Fig A2 also shows that the PSIS approximation was  $\approx 80$  times faster than the exact LOO computation.

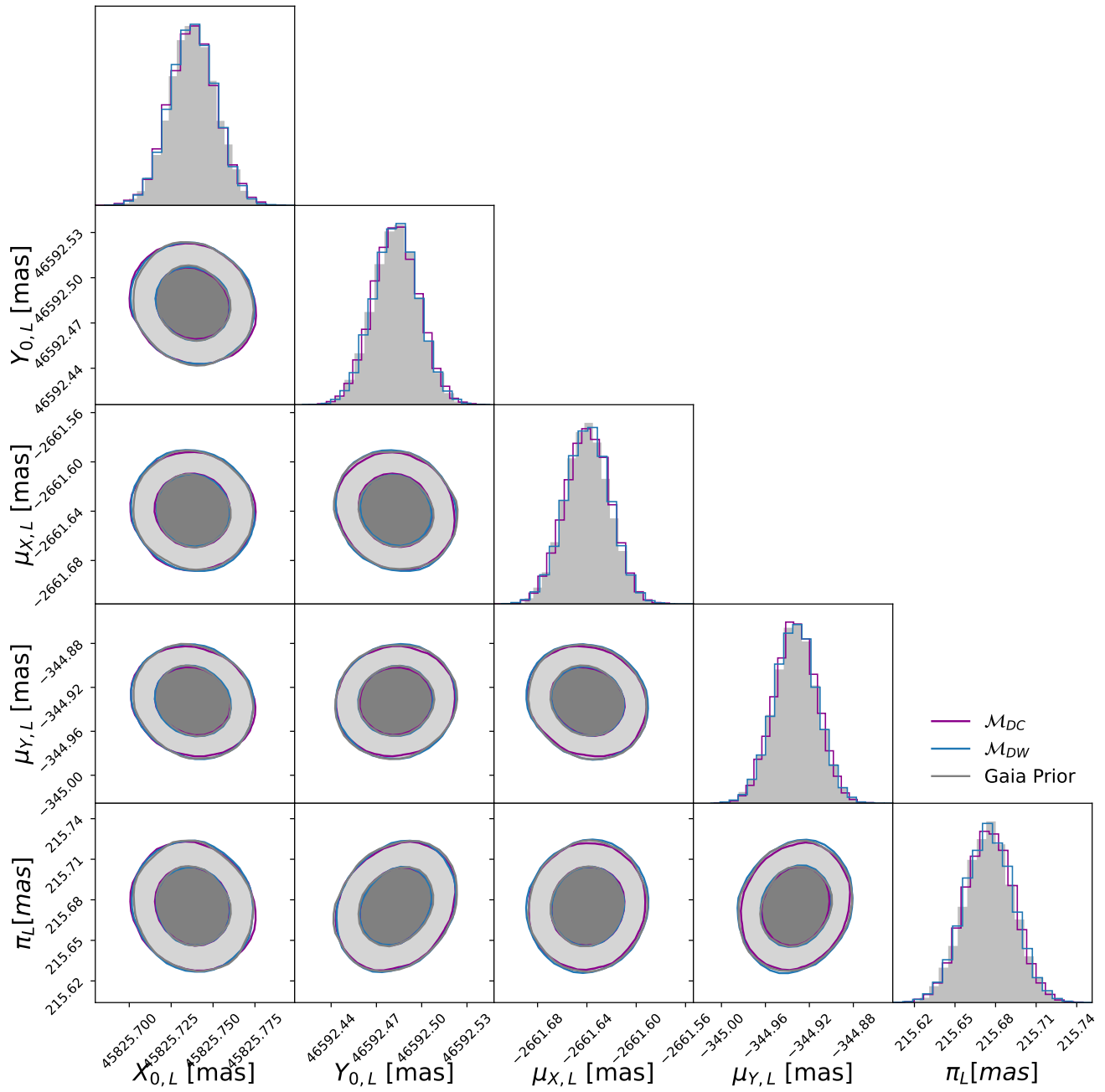
The large negative value of  $LOO_{54, M_{DC}}$  indicates that  $D_{54}$  is an outlying data point which is not well predicted by  $M_{DW}$ . The failure of PSIS approximation for  $D_{54}$  means that the removal of this data point changes the posterior by a large amount compared with the removal of all other data points. This is unsurprising as  $D_{54}$  belongs to epoch 7 where the source was close to a charge bleed column (Fig. 2), and the estimated size of the correlated noise is high (Table 2), due to the specialized fitting procedure used in this epoch (see Section 2.1).  $D_{54}$  can also be seen to be the furthest outlying brown pentagon marker in Fig. 7. Other than  $D_{54}$ 's membership to epoch 7, there is nothing else in the reduction that explains its significant outlying position. One possible explanation could be that if a cosmic ray landed at the very center of the source PSF when  $D_{54}$  was taken, it could have corrupted the source position and would be very hard to identify.

## APPENDIX B: CONSISTENCY WITH *Gaia*

This Section contains the full corner plots for both the source (Fig. B1) and lens (Fig. B2) prior and posterior distributions of their astrometric parameters. For both the lens and source, and all models, it can be seen that the posteriors are in good agreement with the GEDR3 priors.



**Figure B1.** GEDR3 priors versus posterior inferences on all source astrometric parameters for all the models considered. The plot is a full version of the densities shown in Fig. 10.



**Figure B2.** GEDR3 priors versus posterior inferences on all the lens (LAW 37) astrometric parameters for the two deflection models which contain the astrometry as free parameters. For both these models, the *HST* data provides no further constraint from GEDR3.

ORIGINAL