A forward-modelling method to infer the dark matter particle mass from strong gravitational lenses

Qiuhan He,[1]⋆ Andrew Robertson,[1] James Nightingale,[1] Shaun Cole,[1] Carlos S. Frenk,[1] Richard Massey,[1] Aristeidis Amvrosiadis,[1] Ran Li,[2,3] Xiaoyue Cao,[3,2] Amy Etherington[1]

[1]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[2]*National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China*
[3]*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*

ABSTRACT
A fundamental prediction of the cold dark matter (CDM) model of structure formation is the existence of a vast population of dark matter haloes extending to subsolar masses. By contrast, other possibilities for the nature of the dark matter, such as a warm thermal relic or a sterile neutrino (WDM) predict a cutoff in the mass function at a mass of $\sim 10^8$ M$_\odot$. We use mock observations to demonstrate the viability of a forward modelling approach to extract information on the cosmological number density of low-mass dark matter haloes along the line-of-sight to galaxy-galaxy strong lenses. This can be used to constrain the mass of a thermal relic dark matter particle, m_{DM}. With 50 strong lenses at Hubble Space Telescope resolution and signal-to-noise (similar to the existing SLACS survey), the expected 2σ constraint for CDM is $m_{DM} > 3.7$ keV. If, however, the dark matter is a warm particle of $m_{DM} = 2.2$ keV, one could rule out $m_{DM} > 3.2$ keV. Our [Approximate Bayesian Computation] method can be extended to the large samples of strong lenses that will be observed by future space telescopes, potentially to rule out the standard CDM model of cosmogony. To aid future survey design, we quantify how these constraints will depend on data quality (spatial resolution and integration time) as well as on the lensing geometry (source and lens redshifts).

Key words: dark matter – gravitational lensing: strong – methods: statistical

1 INTRODUCTION

Constraining the identity and mass of the dark matter particle is one of the most challenging goals in physics today. Astrophysics can play a major role in this endeavour. Cosmological N-body simulations have shown how dark matter haloes form from initial density fluctuations in the Universe (Frenk et al. 1985). The halo mass function (i.e. the abundance as a function of mass) and internal density profiles are determined by the power spectrum of initial fluctuations, which itself depends on the nature of the dark matter particle.

In the case of cold dark matter (CDM), the halo mass function extends as a power law to very low masses (about Earth mass for a 100 GeV particle) (Wang et al. 2020) . This is perhaps the most fundamental prediction of the standard model of cosmology, ΛCDM. By contrast, if the dark matter is "warm" (WDM), the thermal velocities of the particles at early times damp small-mass fluctuations leading to a cutoff in the halo mass function at a scale that depends on the inverse of the particle mass; for particles of keV mass, this is a few times 10^8 M$_\odot$. Thus, a measurement of the small-

mass end of the halo mass function would set strong constraints on the properties of the dark matter particle, including its mass, and could, in principle, rule out some of the main current candidates, including CDM.

However, measuring the properties of dark matter haloes is not trivial. In sufficiently massive haloes, gas can flow to the centre and give birth to a luminous galaxy, thus, in principle, enabling direct number counts. The observed clustering of luminous galaxies provided the first astrophysical constraints on the mass of dark matter particles, ruling out light particle candidates called "hot dark matter", such as light neutrinos, as the predominant form of dark matter (Frenk et al. 1983; White et al. 1983; see Frenk & White 2012 for a review). However, these observations are unable to constrain WDM particles which would require measurements of haloes with masses less than $\sim 10^{10}$ M$_\odot$, that are too small to have made a luminous galaxy (Thoul & Weinberg 1996; Efstathiou 1992; Benson et al. 2002; Sawala et al. 2016; Benitez-Llambay & Frenk 2020). Furthermore, baryonic physics can have a significant effect on dark haloes in this mass range (particularly for those living within a galaxy) adding further complexity to the task. Strong gravitational lensing is an observable that can, in principle, be used to detect dark

⋆ E-mail: qiuhan.he@durham.ac.uk

haloes and constrain the dark matter particle mass. Other observables include the statistics of the Lyman-α forest (Viel et al. 2013; Iršič et al. 2017; Garzilli et al. 2019), features in stellar cluster tidal debris in the Milky Way (Bonaca et al. 2019) and populations of Milky Way satellites (Lovell et al. 2016; Schneider 2016; Lovell et al. 2017; Nadler et al. 2020).

Strong gravitational lensing is the phenomenon whereby light rays from a distant galaxy are deflected by a foreground object, resulting in the formation of large luminous arcs or multiple distinct galaxy images. In principle, any object that happens to lie near the light rays coming from the background source galaxy will leave an imprint on the image, such that small perturbations (or lack thereof) to the lensed source light can be used to infer the presence (or absence) of low mass dark matter haloes. Depending on the source size and the kind of data available, two broad types of analysis are possible.

If the source is an extremely compact object, like a quasar, when lensed, one will observe multiple point sources. Low mass 'invisible' dark matter haloes along the line-of-sight induce anomalies in the flux ratios amongst multiple images and by analysing those anomalies, one can constrain the dark matter particle mass (Mao & Schneider 1998; Metcalf & Madau 2001; Dalal & Kochanek 2002; Amara et al. 2006; Xu et al. 2012; Nierenberg et al. 2014; Xu et al. 2015; Nierenberg et al. 2017; Gilman et al. 2019; Hsueh et al. 2020; Gilman et al. 2020b).

In the case where the source is extended (e.g. a regular galaxy), the lensed source appears as multiple arcs or a complete Einstein ring. Sufficiently massive dark matter haloes along the line-of-sight perturb local regions of the arc in a way that is measurable using techniques such as the pixelized potential correction method (Vegetti et al. 2012, 2014; Hezaveh et al. 2016; Vegetti et al. 2018; Ritondale et al. 2019). By combining results for individual haloes in many systems one can draw conclusions regarding the halo mass function (e.g. Vegetti & Koopmans 2009; Li et al. 2016). However, this modelling has demanding requirements on image quality and resolution and so far only a few haloes have been reported to have been detected with mass below $\sim 10^{10}$ M$_\odot$ (Vegetti et al. 2010, 2012; Hezaveh et al. 2016).

The detection of individual haloes or subhaloes requires their mass and position to be such that they produce an observable perturbation to the lensed image of the source. The resolution and signal-to-noise of the data thus define the 'sensitivity' to halo detection. Haloes of mass $\sim 10^8$ M$_\odot$ are detectable with HST data (Vegetti et al. 2012) while haloes as light as 10^6 M$_\odot$ are detectable with radio data (McKean et al. 2015). Recent studies have proposed that lower masses may be accessible by instead extracting, statistically, the cumulative perturbation of all intervening haloes on strong lensing arcs, even though none of these haloes could be detected individually (Brewer et al. 2016).

Most studies have focused on determining a theoretical relation between the power spectrum of the convergence field and that of the image residuals after fitting a smooth 'macro model' to the lens galaxy mass distribution (Chatterjee & Koopmans 2018; Díaz Rivero et al. 2018; Cyr-Racine et al. 2019). There is one example using real data, in which Bayer et al. (2018) analysed image residuals of the system SDSS J0252+0039, obtaining a detection of power much higher than the prediction of the cold dark matter (CDM) model. There remain a number of questions as to what assumptions are appropriate for this technique. For example, to simplify the theoretical calculations, a random Gaussian field is often assumed for the convergence, potentially omitting multiplane lensing effects (Schneider 2014; McCully et al. 2014). Indeed, recent studies have

shown that the expected signal in a ΛCDM universe is dominated by line-of-sight haloes at different redshifts to the lens galaxy as opposed to subhaloes within the lens galaxy itself (Li et al. 2016, 2017; Despali et al. 2018; Çağan Şengül et al. 2020). Here, we will focus on how the line-of-sight low mass dark matter haloes contribute to perturbations on lensing images.

To form a visual impression of how the line-of-sight low-mass dark haloes perturb the images, in Fig. 1 we show the effective convergence, as derived from the deflection angles (upper panels; Gilman et al. 2019; Çağan Şengül et al. 2020), and the corresponding residuals obtained by fitting the smooth lensing model to the image (lower panels). As may be seen from the effective convergence map, even though all low-mass dark matter haloes are modelled as spherical profiles, some are heavily stretched into arc-like features due to multiplane lensing effects[1], which are difficult to compute analytically. Comparing the left two columns, the 'warm' case in the middle and the 'cold' case in the left have very similar residuals, particularly the large patches, even though the warm case has far fewer low-mass dark matter haloes. This suggests that the residuals are dominated by massive dark matter haloes and to distinguish between different dark matter models, we need to identify the small patches. The system shown in the third column has a higher redshift source and thus many more low-mass dark matter haloes along the line-of-sight, which results in larger and more complicated residuals.

In this work, we investigate whether a forward modeling procedure built around the approximate Bayesian computation (ABC) statistical inference method can extract meaningful information about the properties of low-mass dark matter haloes perturbing strong lensing images. As opposed to determining a concise mathematical expression, our forward modelling method relies on robust modelling of the perturbations on lensing images induced by small dark haloes which can directly build up a relation between the models of interest and the observations without any further assumptions. The approach was first applied to strong lensing by Birrer et al. (2017b) who analysed the strongly lensed quasar RXJ1131-1231. Subsequently, Gilman et al. (2019) used this approach to study flux ratio anomalies of point-like sources with a full realisation of dark matter haloes including line-of-sight dark haloes.

We apply the ABC framework to a simulated sample of 50 strongly lensed extended sources, representative of Hubble Space Telescope imaging of the SLACS survey (Bolton et al. 2006). Each simulation contains a full cosmological realisation of small dark haloes whose redshift distribution requires analysis using multiplane ray tracing. To apply the ABC method we simulate and refit each of these 50 lenses 20 000 times, producing a total of 1 million lens models. The scale of this analysis necessitates an automated framework for lens modelling, for which we use the open source software PyAutoLens[2]. Our aim is to determine whether the cumulative distortions due to the many dark matter haloes perturbing the light of the lensing arcs can be extracted to determine the halo mass function and hence the dark matter particle mass. In particular, we focus on whether this is feasible with existing imaging of the SLACS sample.

Our main goal is to demonstrate that this signal is present in

---

[1] In strong lensing, as viewed from the image plane, not only is the distant light warped by foreground objects, like a luminous source galaxy deflected by a lens, but also the mass on distant planes is warped by near planes, and thus we see some arc-like low-mass dark matter haloes in these convergence maps even though they are all spherical.
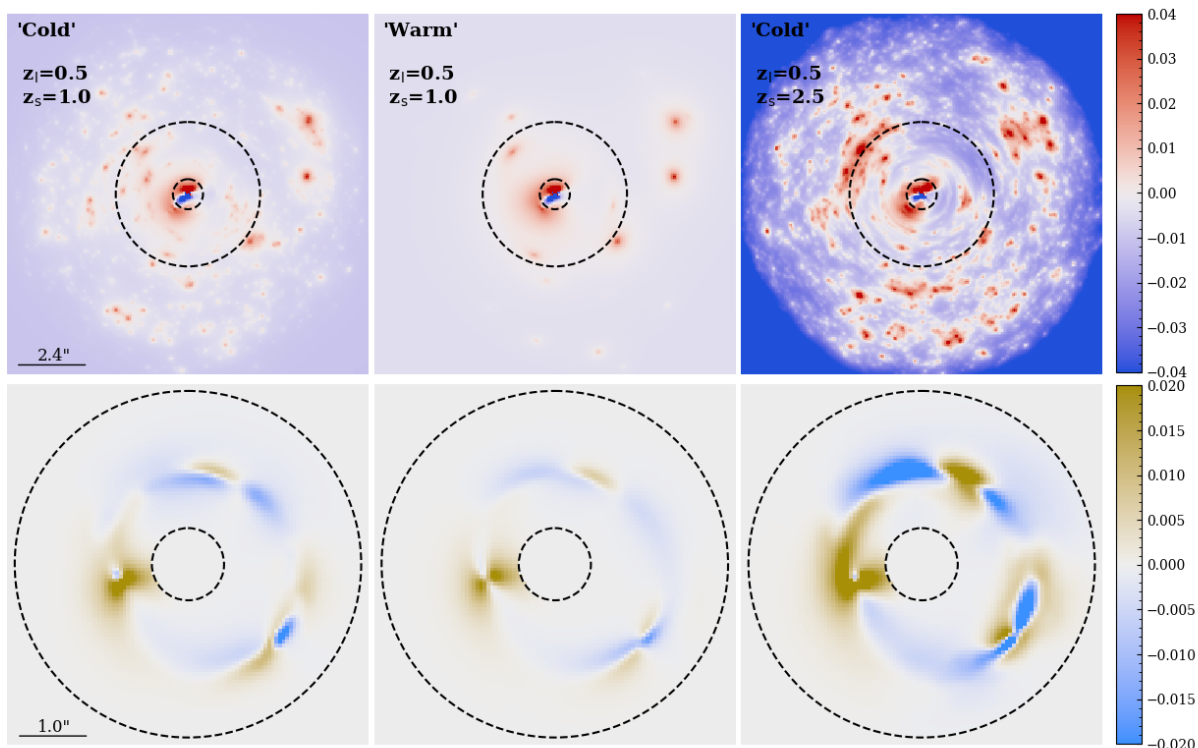
[2] https://github.com/Jammy2211/PyAutoLens

Figure 1. Upper panels: the effective convergence of low mass perturbers. Lower panels: the corresponding best-fit image residuals. The colour scale is in units of e^- pix^{-1} s^{-1}. The parameters of the lens and source galaxies shown in the three columns are the same as those used in Fig. 4, except for the source redshifts. The lensing systems in the left and middle columns have a source at $z = 1$, while that in the right column has a source at $z = 2.5$. The systems in both the left and right columns have a cutoff in the mass function at 10^7 M$_\odot$, while the one in middle column has the cutoff at 10^9 M$_\odot$. The inner and outer dashed circles in each panel have radii, $0.5''$ and $2.4''$, respectively.

HST imaging of strong lenses and that it can, in principle, be extracted using modern lens modeling techniques given realistic levels of noise. However, our study is based on idealized systems: we make a number of simplifying assumptions for the structure of the lens and source and neglect effects such as an imperfect PSF model, correlated noise or inadequate lens light subtraction. These assumptions will need to be relaxed before our methodology can reliably be applied to real data, including, for example, a non-parametric source model (e.g. Warren & Dye 2003), additional complexity in the lens model (Vegetti & Koopmans 2009; Nightingale et al. 2019) and a proper treatment of the PSF and of correlated noise.

We first introduce the forward modelling procedure and lensing models in Sec. 2. In Sec. 3 we show tests of the accuracy of this method, the dependency on different lensing and observational settings, and compare our method to other methods, discussing its possible future applications and shortcomings. Finally, we conclude in Sec. 4. Throughout the paper, we adopt the cosmological parameters given by WMAP9 (Hinshaw et al. 2013).

2   PROCEDURE AND MODELS

In this section we first provide an overview of our forward modelling procedure. We then describe the parametric models we use for the mass distributions of the main lens, dark matter haloes and light distributions of the source galaxies. We then describe how we fit our simulated images with a combination of a smooth parametric lens and source model, and how the residuals of each are used within

an ABC framework to place constraints on the mass function of dark matter haloes.

2.1   The forward modelling scheme

In Fig. 3, we provide an overview of the forward modelling procedure. Starting from an observed strong lensing image (which in this paper is simulated) we begin by fitting it with parametric lens and source models, omitting substructure from the lens model. This procedure gives us best-fit smooth lens mass and source light models, as well as a map of the best-fit image residuals (the observed image minus the best-fit model image). In this work, 'best-fit' refers to the maximum likelihood model determined by means of a non-linear search.

It is impossible to fit all the dark matter substructre in a similar way, because of its low mass and low signal-to-noise (Birrer et al. 2017b). However, we can use the best-fit source and macroscopic lens model to simulate a set of images of this lens system, each including a random realization of dark matter substructure (we call this set of images the 'forward models'). We refit each forward model in the same manner as we fitted the 'observed' image, providing best-fit image residuals for each forward model. We compare the forward-modelled residuals with the observed residuals, as described in detail in Sec 2.2, and apply ABC inference to obtain a constraint on the cutoff in the halo mass function.
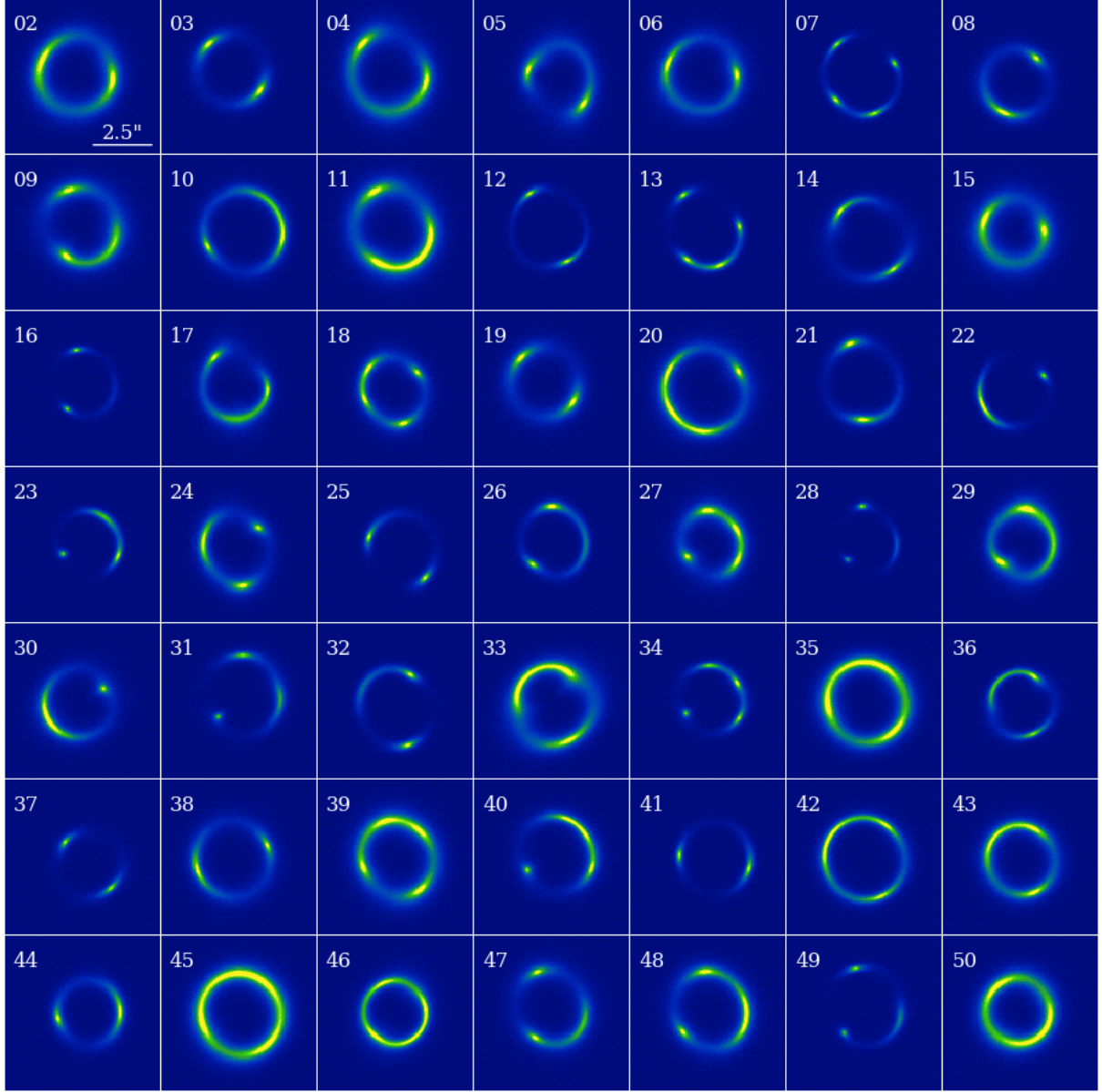
Figure 2. Fiducial mock strong lensing images. A total of 49 mock images are shown, with the last of our set of 50 shown in the upper panel of Fig. 4. The images are simulated with parameters randomly drawn according to Table 1.

2.2 Lensing simulations

The strong lens simulations used in this work represent the lensing system with three components, the lens galaxy, the source galaxy, and the line-of-sight dark matter haloes.

2.2.1 Lens and source

We simulate the smooth mass distribution of our lens galaxy as a 2D singular isothermal ellipsoid (SIE) of the form,

$$
\begin{aligned}
\Sigma(x, y) &= \frac{\sigma^2}{2G} \frac{1}{\sqrt{x^2 q + y^2/q}} \\
&= \frac{c^2}{8\pi G} \frac{D_A(0, z_{source})}{D_A(z_{lens}, z_{source})} \frac{R_E}{\sqrt{x^2 q + y^2/q}},
\end{aligned}
\tag{1}
$$

where R_E is the projected Einstein radius, which can be related to the velocity dispersion of the profile, σ, given the lens and source redshifts (z_{lens} and z_{source}, respectively) and a cosmology. The quantity, $D_A(z_1, z_2)$, is the angular diameter distance between z_1 and z_2; q is the axis ratio.

For the source, we adopt the Sérsic (1963) profile,

$$
I(r) = I' \exp\left[-b_n \left(\frac{r}{r_e}\right)^{1/n}\right],
\tag{2}
$$

where I' is the scale intensity, n is the Sersic index and b_n is a function of n defined such that r_e is the half-light radius. However, the image residuals caused by perturbations to the lens model (in our case the perturbing haloes) are effectively a product of the deflection

Parameter	Range or Value
Lens	Elliptical Isothermal
centre (x, y) ["]	(0.0, 0.0)
R_E ["]	[1.2, 1.6]
axis ratio	[0.6, 0.95]
position angle [°]	30
redshift	0.5
Source	Elliptical Core-Sersic
centre (x, y) ["]	([−0.3, 0.3], [−0.3, 0.3])
r_e ["]	[0.1, 0.5]
n	[1.5, 2.5]
I' [e$^-$ pix^{-1} s^{-1}]	7.4
axis ratio	[0.5, 0.95]
position angle [°]	[10, 100]
r_c ["]	0.01
α	2.0
γ	0.0
redshift	1.0
Dark haloes	Truncated NFW profile
F_{CDM}	Sheth et al. (2001)
c_{CDM}	Ludlow et al. (2016)
$\sigma_{\log_{10} c_{CDM}}$	0.15 dex
tNFW τ	r_{100}/r_s
M_{hf} [$M_\odot$]	10^7
Image	
pixel size ["]	0.05
PSF σ ["]	0.05
t_{exp} [s]	600
background sky [e$^-$ pix^{-1} s^{-1}]	1.0
max S/N of pixels	~ 50

Table 1. Fiducial model parameters. Closed brackets indicate the range from which values are randomly drawn.

	$\{T\}^1_{\rm fiducial}$	$\{F\}^1_{\rm fiducial}$
Lens	Elliptical Isothermal	
centre (x, y) ["]	(0.0, 0.0)	$(-0.002^{+0.002}_{-0.001}, 0.001^{+0.001}_{-0.003})$
R_E ["]	1.342	$1.342^{+0.001}_{-0.003}$
axis ratio	0.618	$0.617^{+0.003}_{-0.003}$
position angle [°]	30.0	$30.0^{+0.1}_{-0.2}$
Source	Elliptical Core-Sersic	
centre (x, y) ["]	(−0.072, 0.259)	$(0.071^{+0.001}_{-0.001}, 0.253^{+0.001}_{-0.001})$
r_e ["]	0.244	$0.240^{+0.005}_{-0.004}$
n	1.85	$1.81^{+0.05}_{-0.05}$
I' [e$^-$ pix^{-1} s^{-1}]	7.4	$7.3^{+0.2}_{-0.2}$
axis ratio	0.85	$0.85^{+0.01}_{-0.01}$
position angle [°]	41	40^{+2}_{-2}

Table 2. The true input and best-fit parameters for the first lensing system. The errors listed in this table are 3σ limits. Notice that we fit a system with additional low-mass perturbers, and thus the inconsistency of the best-fit parameters and the true values is expected. Parameters fixed during fitting are not shown here.
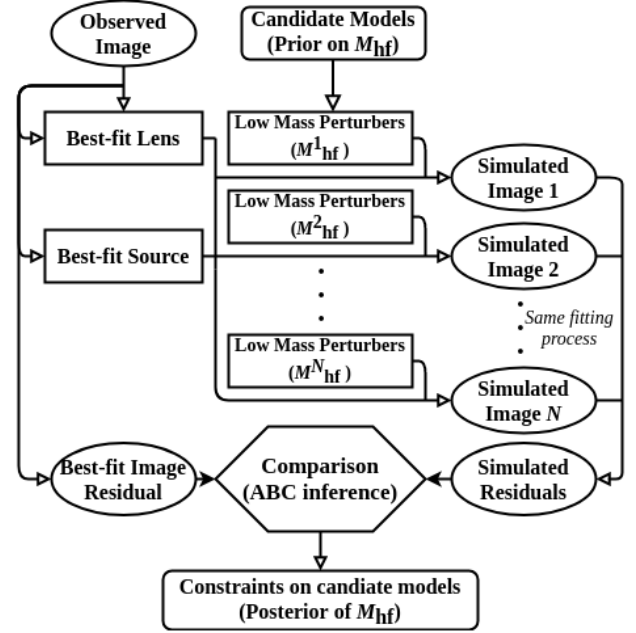
Figure 3. An overview of the forward modelling procedure for one lensing system. The observed image is fitted with a parametric lens mass model and a parametric source light model, producing a best-fit lens model, source model and image residuals. We use this best-fit model to generate N simulated images, including random realisations of low-mass perturbing haloes in the lens model. The number of perturbers as a function of mass depends on the properties of the dark matter as encoded in the 'half-mode mass', M_{hf}. The distribution of M_{hf} over the N 'forward models' follows our prior on M_{hf}. Each of the forward model images is fit in the same way as the observed image to produce N sets of image residuals. These forward model residuals are compared with the observed residuals, using ABC to constrain M_{hf}.

angles due to the perturbation and the gradient of the source surface brightness (Vegetti & Koopmans 2009). The cuspy centre of the Sersic profile creates an infinite surface brightness gradient which numerically causes infinite surface brightness differences. This can be overcome by oversampling the source light, but at the expense of increased computational run time. Instead, we use a Sersic profile with a "core" (Graham et al. 2003; Trujillo et al. 2004) to simulate the background source. This has the form,

$$I(r) = I' \exp\left[-b_n \left(\frac{r^\alpha + r_c^\alpha}{r_e^\alpha}\right)^{1/(n\alpha)}\right], \tag{3}$$

where r_c is the "core radius", which marks the transition radius from the Sersic profile to constant surface brightness. When $r_c \to 0$, the profile reduces to the Sersic form. The parameter, α, quantifies how quickly the profile transitions from a regular Sersic form to one with a constant surface brightness core. To simplify both our simulation and modelling processes, we fixed $\alpha = 2$ and $r_c = 0.01$". With these parameters fixed, the source model has the same three free parameters as the regular Sersic profile: r_e, n and I'.

2.2.2 *Low-mass dark matter haloes*

Low-mass dark matter haloes that can perturb lensing arcs can be of two types: subhaloes within the main lens, or central haloes at different redshifts that happen to lie close to the path of light from
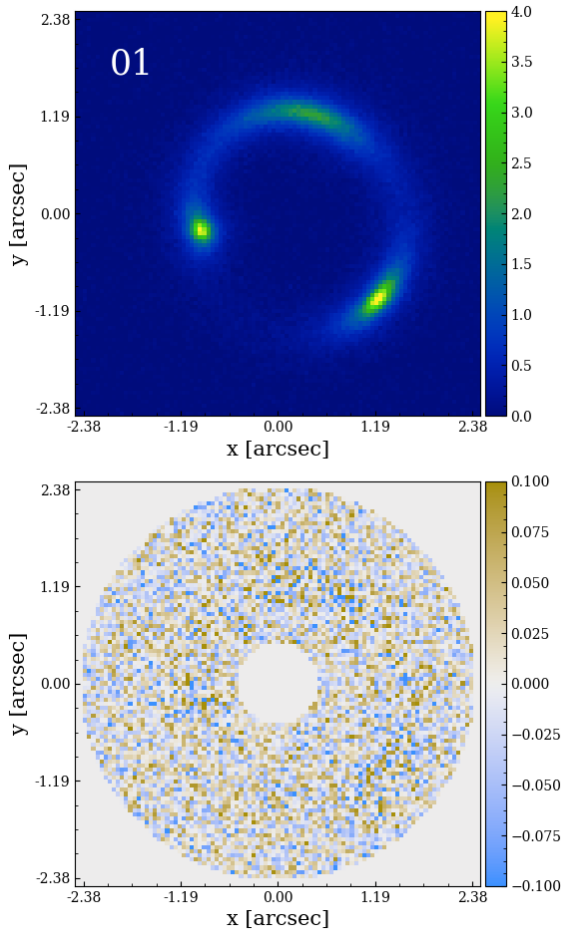
Figure 4. Upper panel: the mock image of the first system. Lower panel: the corresponding residuals after fitting a smooth macro model.

the source galaxy to the observer. We will refer to the latter as line-of-sight haloes.

The number and mass distribution of subhaloes within the lens are somewhat uncertain. Subhaloes are subject to tidal stripping and disruption; these effects are significantly enhanced in hydrodynamic simulations compared to dark matter only simulations due to the presence of the dense stellar component at the centre of the main lens (Sawala et al. 2017; Garrison-Kimmel et al. 2017; Richings et al. 2020a). The degree of disruption also depends on the details of the galaxy formation model (Richings et al. 2020b). In contrast, line-of-sight haloes are not subject to these environmental effects and, in the mass range of interest here, they are entirely 'dark', having never formed stars.

Recent studies (Li et al. 2016, 2017; Despali & Vegetti 2017; Çağan Şengül et al. 2020) have shown that for the lens and source redshifts typical of SLACS lenses, the lensing perturbations mainly arise from line-of-sight haloes. This is fortunate because the irrelevance of uncertain baryon effects, makes line-of-sight haloes a particularly clean probe of a cutoff scale in the halo mass function. In this work, we focus on these low-mass, line-of-sight haloes whose mass function has the form given by Lovell et al. (2014):

$$F(M_{200}, z) \equiv \frac{d^2N}{dM_{200}dV}(M_{200}, z) = F_{\text{CDM}}(M_{200}, z)\left(1 + \frac{M_{\text{hf}}}{M_{200}}\right)^{-1.3}.$$

$$(4)$$

This is composed of two parts, the CDM mass function, $F_{\text{CDM}}(M_{200}, z)$, for which we use the form derived by Sheth et al. (2001), and a cutoff parameterized by the 'half-mode' mass, M_{hf}, a characteristic mass corresponding to the mass scale at which the dark matter transfer function falls to half the CDM transfer function. The half-mode mass can be related to the mass of the dark matter particle, m_{DM}, as,

$$M_{\text{hf}} = 10^{10}\left(\frac{m_{\text{DM}}}{\text{keV}}\right)^{-3.33} \text{M}_\odot \text{h}^{-1}, \qquad (5)$$

where h is the Hubble parameter in units of 100 km/s/Mpc (Lovell et al. 2014). From the perspective of strong lensing, M_{hf} may be regarded as an effective 'cutoff' mass in the dark matter mass function, below which there are very few haloes.

We draw low-mass dark haloes in a lightcone between the observer and the source galaxy within an angular radius, r_{lc}, chosen as,

$$r_{\text{lc}} = \begin{cases} 5.0'' & (z \le z_{\text{lens}}) \\ 5.0'' - 2.0'' \times \frac{D_A(z_{\text{lens}}, z)D_A(0, z_{\text{source}})}{D_A(0, z)D_A(z_{\text{lens}}, z_{\text{source}})} & (z > z_{\text{lens}}) \end{cases},$$

$$(6)$$

where D_A is the angular-diameter distance and z_{lens} and z_{source} are the redshifts of the lens and source galaxies respectively. The angular radius of the lightcone is fixed at $5.0''$ (this is ~ 3 times the Einstein radius, and including haloes at larger radii hardly affects our lensing images) in front of the main lens and gradually decreases behind from $5.0''$ to $3.0''$ at the source redshift, which ensures that all perturbers that might influence the lensed image of the source are included. The lightcone is evenly divided into redshift bins, of width $\Delta z = 0.01$, which balances the accuracy and computational cost of the calculation (see appendix A for details). Within each bin, we sample from the halo mass function, assign density profiles from the halo mass-concentration relation corresponding to that redshift, and place these perturbing haloes randomly on the central redshift plane. Neglecting the clustering of the line-of-sight haloes is a good approximation if, over the angular extent of interest, the typical line-of-sight separation of pairs of haloes is much larger than their correlation length. This is expected to be true because: (a) the intrinsic clustering of low mass haloes is weak (Kaiser 1984; Sheth & Tormen 1999), and (b) the angular extent of interest is very narrow since it is set by the width of the Einstein ring image.

We generate low mass haloes with M_{200} between 10^6 M$_\odot$ and 10^{10} M$_\odot$ within the light cone. Depending on the dark matter model, a significant fraction of the mass in the Universe can exist as collapsed objects in this mass range. To solve correctly the strong lensing equations (which have already assumed a background universe) we add negative mass sheets onto each plane to compensate for the otherwise 'double-counted' mass within perturbers in the lightcone (Petkova et al. 2014; Birrer et al. 2017a; Gilman et al. 2019). For each plane, we include a mass sheet with constant negative surface density equal to the expected density in low mass haloes on that plane. This leads to a mass sheet with a surface density

$$\Sigma_{\text{negative}} = -\frac{1}{A}\int F(M_{200}, z)M(M_{200}, z, s)\, dM_{200}ds\, dV, \qquad (7)$$

where A is the physical area of the lightcone at the plane's redshift, $F(M_{200}, z)$ is the mass function and $M(M_{200}, z, s)$ is the total mass of a dark halo on the plane, which can differ from M_{200} due to halo truncation. The "s" denotes some other parameters of the dark matter profile; in our case, as we discuss below, it represents the "concentration" and "truncation radius".

Instead of modelling haloes as Navarro-Frenk-White (NFW) profiles (Navarro et al. 1996, 1997), which have an infinite total mass (when integrating to infinite radius), we use a truncated NFW profile (Baltz et al. 2009),

$$\rho(r) = \frac{M_0}{4\pi r \, (r_{\rm s} + r)^2} \cdot \frac{r_{\rm t}^2}{r_{\rm t}^2 + r^2}, \tag{8}$$

which has an additional truncation term compared with the standard NFW form. Here, the scale mass, M_0, and scale radius, $r_{\rm s}$, are a pair of parameters that specify an NFW profile and $r_{\rm t}$ is the truncation radius. With the addition of the truncation term, $\rho(r) \propto r^{-5}$ at large radii, and the profile now has a finite mass given by,

$$M_{\rm t} = M_0 \frac{\tau^2}{\left(\tau^2 + 1\right)^2} \left[\left(\tau^2 - 1\right) \ln\tau + \tau\pi - \left(\tau^2 + 1\right) \right], \tag{9}$$

where $\tau = r_{\rm t}/r_{\rm s}$. We set the truncation radius to be the corresponding NFW r_{100}, the radius within which the average density of an NFW profile is 100 times the critical density of the universe. With τ fixed to $r_{100}/r_{\rm s}$, the tNFW profiles are determined by the same two parameters as the NFW profile, M_{200} and the concentration, $c = r_{200}/r_{\rm s}$, which quantifies how centrally concentrated a halo is. Haloes that form earlier typically have higher concentrations (Navarro et al. 1996, 1997). Since the reduction in small-scale power in a WDM model leads to later formation time for low mass haloes these are less concentrated than their CDM counterparts (Lovell et al. 2012; Schneider et al. 2012; Macciò et al. 2013; Bose et al. 2016; Ludlow et al. 2016).

When considering WDM models we adopt the mass-concentration relation of Bose et al. (2016)

$$c(M_{200}, z) = c_{\rm CDM}(M_{200}, z) \left((1 + z)^{0.026z - 0.04} \left(1 + 60 \frac{M_{\rm hf}}{M_{200}} \right)^{-0.17} \right), \tag{10}$$

where $c_{\rm CDM}(m, z)$ is the mass-concentration relation in the CDM case. For CDM, we use the mass-concentration relation of Ludlow et al. (2016), which agrees with recent simulation results from Wang et al. (2020), who measured halo concentrations in ΛCDM simulations over twenty orders of magnitude in halo mass. We use the public package COLOSSUS (Diemer 2018a) to compute the mass-concentration relation. The Ludlow et al. (2016) relation is for the median concentration at a given mass, but simulations predict scatter in concentration at fixed mass which is well described by a log-normal distribution (Neto et al. 2007). As a result, we include log-normal scatter in the mass-concentration relation, with a standard deviation of 0.15 dex, as shown by Wang et al. (2020).

Multi-plane ray tracing is necessary to calculate the effect of line-of-sight haloes on flux ratio anomalies with sufficient precision to match observations (Gilman et al. 2019). We use the publicly available software package PyAutoLens (Nightingale & Dye 2015; Nightingale et al. 2018a, 2019) to simulate lenses using the light and mass profiles described above. To define our fiducial lens sample, we simulate 50 unique lensing systems with different values for the lens SIE mass model (e.g. Einstein radius, axis ratio, etc.) and source core-Sersic profile. The input $M_{\rm hf}$ is 10^7 M$_\odot$. For all 50 systems we assume lens redshifts of 0.5 and source redshifts of 1.

We produce observations representative of Hubble Space Telescope imaging, with a pixel size of $0.05''$, a Gaussian PSF with a standard deviation of $0.05''$ and background sky level of $1 e^-$/pix/second. We set the exposure time to be 600 s. Then the source surface brightness of each simulation is chosen such that

the maximum signal-to-noise ratio (e.g. in the brightest pixel) has a value ~ 50, which is roughly that of the brightest SLACS lenses (Bolton et al. 2006). Table 1 summarizes the models and parameters for our mock observations, where closed brackets indicate the range of values from which each parameter is randomly drawn. For convenience, we use $\{T\}_{\rm fiducial}$ to denote the set of true parameter values used for our fiducial tests and a superscript i to represent the parameters of the i-th lensing system. Mock images of our 50 strong lensing systems are shown in Fig. 2.

2.3 Analysis Process

Our analysis of each mock observation consists of two steps: smooth model fitting followed by ABC inference. Below we describe the detailed steps in this process and illustrate them for one system.

2.3.1 Smooth Model Fitting

First, we fit a simple lens model to the mock lensing image. This consists of a smooth parametric model for the lens galaxy mass distribution (the *macro model*) and a smooth parametric model for the source. We adopt the same parametric forms we used when generating the simulated mock image, an SIE lens mass model and a core-Sersic source.

We use PyAutoLens to fit the mock images, and adopt Py-MultiNest with constant sampling efficiency mode switched on (Feroz et al. 2009a; Buchner et al. 2014a) for the non-linear optimisation. Table 2 shows the inferred parameters for this fit, alongside the true input values for the first mock image. To distinguish between the true inputs and the best-fit values we label the latter as $\{F\}_{\rm fiducial}$, where a superscript, i, denotes the best-fit parameters for the i-th lensing system. This table shows that the inferred and input lens model are close, but not in perfect agreement; this is due to the effects of the low-mass dark matter haloes. We will show later that this small mismatch does not affect our calculation.

In the upper and lower panels of Fig. 4 we show the "observed" image and best-fit residuals for one of our 50 mock observations. To develop intuition on how the low mass perturbers affect the image residuals, we further show the "effective convergence" (the divergence of deflection angles) of low mass perturbers and the corresponding best-fit residuals in Fig. 1. We assume a very small constant noise to show the residuals clearly. The three strong lensing systems shown have the same lens and source galaxy as in Fig. 4, except that the source for the system in of the right-hand column is at $z = 2.5$. For a clear comparison, we make sure that the three systems share the same realisation of low mass perturbers. To achieve this, instead of applying the mass function described by $M_{\rm hf}$ in Eq. 4, we draw haloes from the mass function of Sheth et al. (2001) but discard those with M_{200} below a certain value. Furthermore, we assume no correlation between $M_{\rm hf}$ and the mass-concentration relation. Please note that this change of low mass perturber realisation is only implemented for Fig. 1; in the rest of this paper, we use the mass function and mass-concentration relation described by Eq. 4 and Eq. 10 respectively.

In Fig. 1, the systems in the left and right columns have a cutoff in the mass function at 10^7 M$_\odot$, while the system in the middle column has a cutoff at 10^9 M$_\odot$. As seen from the left two columns, the system with a lower cutoff in the mass function has far more low mass perturbers and thus more subtle structure in the residuals. When comparing the left and the right columns, we see that the number of low mass perturbers increases significantly

with the source redshift and the "profiles" of individual perturbers are more heavily distorted because of the multiplane lensing effect. Also, for a higher redshift source galaxy, the residuals are larger.

2.3.2 ABC inference through forward modelling

Approximate Bayesian Computation (ABC) is a likelihood-free method suitable for problems where the likelihood is difficult to express analytically, but model predictions are relatively easy to simulate. It has been widely applied in astrophysics, for example in studies of large-scale structure, planet surveys and reionization (Akeret et al. 2015; Hahn et al. 2017; Davies et al. 2018; Hsu et al. 2020). Birrer et al. (2017b), Gilman et al. (2019, 2020b,c) and Enzi et al. (2019) have also used ABC to constrain dark matter substructure using strong lensing data. To begin, one defines a summary statistic which measures the similarity between the observational data and the simulations. In principle, any statistic (as long as it contains information of interest) can be used as a measure of similarity, but different statistics may have different requirements on data quality, sampling efficiency, etc. A good statistic captures as much of the characteristic information of different models as possible: in our case, the lensing perturbations from low-mass dark matter haloes. Finding a proper summary statistic is a challenging part of ABC and it is beyond the scope of this work to investigate the many possible candidate summary statistics. Instead, following Bayer et al. (2018), we use the power spectrum of the best-fit image residuals to construct a summary statistic that we hope can extract information on the number of low-mass dark matter haloes perturbing the lens.

Based on the best-fit source and macro model for the lens, $\{F\}_{\mathrm{fiducial}}^{1}$, we simulate images of this best-fit source and lens combination with the addition of low-mass haloes along the line-of-sight. We uniformly sample $\log_{10} M_{\mathrm{hf}}/M_{\odot}$ between 7 and 10 (corresponding to assuming a flat prior on $\log_{10} M_{\mathrm{hf}}$ between $10^7 M_{\odot}$ and $10^{10} M_{\odot}$) and, for each sampled M_{hf} value, we draw a random realisation of low-mass dark matter haloes with the appropriate mass function and mass-concentration relation. We do this for 20,000 values of M_{hf}, producing a corresponding lensed image in each case and then fit each of these 20,000 images to obtain the best-fit image residuals. We have confirmed that our results are converged with 20,000 samples. Note that the forward-modelled images are simulated from the known lens and source galaxies, so when performing the fitting, we add narrow priors centred on the parameter values of the known best-fit lens and source models to accelerate the fitting processes. Since our only goal is to find the maximum likelihood model, once our priors contain the correct solution, the size of the priors does not affect our results. We have made sure that the sizes of the narrow priors are large enough to include the maximum likelihood model.

Next, for the images of the best-fit residuals, we set pixels outside the annular region between $0.5''$ and $2.4''$ to be 0 and then Fourier transform the residual image and azimuthally average to obtain the 1-D power spectrum, $P(k)$. In the upper panel of Fig. 5, we show power spectra for the forward-modelled residuals as black lines (for clarity we only plot 100 of them); the red curve marks the power spectrum of the image residuals of the original observation. Each line is composed of 15 $P(k_i)$ values with the k_i logarithmically spaced between ~ 1.5 arcsec^{-1} and 85 arcsec^{-1} ($\sim 2\pi / (4.2$ arcsec$)$ to $2\pi / (0.07$ arcsec$))$.

Although we have converted the images into power spectra, we cannot simply use the 1D "power spectrum" as our statistic. The reason is that each power spectrum is composed of 15 different values and for ABC to converge at a point in a 15-dimensional space,
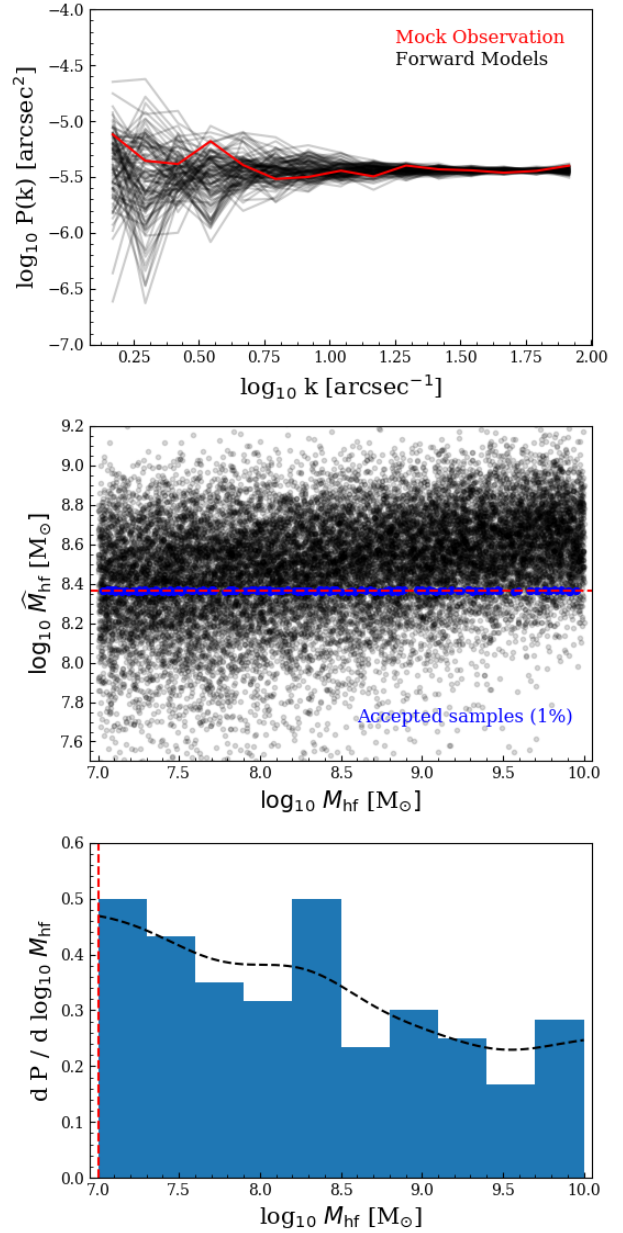

Figure 5. Upper panel: the power spectrum computed from the best-fit image residuals. The red line is the power spectrum of the mock observation. Black lines are for 100 forward models. **Middle panel**: the summary statistic, $\log_{10}\widehat{M}_{\mathrm{hf}}$, computed from the power spectra. The black points are $\log_{10} M_{\mathrm{hf}}$ for the forward models. The horizontal dashed red line marks the value for the mock observation and points 1% closest to it are marked in blue. **Lower panel:** The posterior distribution of M_{hf}, formed by collecting together the M_{hf} of the blue points is shown in the middle panel. The vertical dashed red line marks the true input of M_{hf}. The black dashed curve is a kernel density estimate corresponding to the histogram.

a very small acceptance rate must be set, which, in turn, requires far more forward models. To further reduce the "dimensionality" of the statistic, following Fearnhead & Prangle (2012), we generate a summary statistic from a linear combination of the logarithm of the

power spectrum values at different k,

$$\log_{10}\widehat{M}_{\mathrm{hf}} = \sum_{i=1}^{15} \beta_i \log_{10}\left[P(k_i)/\mathrm{arcsec}^2\right] + \beta_0. \qquad (11)$$

We use $\log_{10}\widehat{M}_{\mathrm{hf}}$ as the summary statistic for the ABC inference. The coefficients $\{\beta_i\}$ (including β_0) are obtained by minimizing the differences between the input $\log_{10}M_{\mathrm{hf}}$ values used in the forward models and their corresponding $\log_{10}\widehat{M}_{\mathrm{hf}}$. Specifically, we find the values of $\{\beta_i\}$ that minimise,

$$\sum_{j=1}^{20000}\left\|\log_{10}\widehat{M}_{\mathrm{hf}}^{(j)} - \log_{10}M_{\mathrm{hf}}^{(j)}\right\|^2, \qquad (12)$$

where $\log_{10}\widehat{M}_{\mathrm{hf}}^{(j)}$ is computed from the j-th forward-modelled power spectrum and $M_{\mathrm{hf}}^{(j)}$ is the corresponding input half-mode mass. Since the forward-modelled power spectrum is systematically different for different macro lensing configurations, we optimize $\{\beta_i\}$ separately for each of our 50 lensing systems. In the middle panel of Fig. 5, we plot the value of $\log_{10}\widehat{M}_{\mathrm{hf}}$ for each forward-modelled simulation as a black point. The horizontal dashed red line marks the value of $\log_{10}\widehat{M}_{\mathrm{hf}}$ given by the mock observation.

We accept the forward-modelled simulations with the 1% of $\log_{10}\widehat{M}_{\mathrm{hf}}$ values closest to the observed $\log_{10}M_{\mathrm{hf}}$ (Eqn. 11), which are shown as blue points in the middle panel of Fig. 5. The set of M_{hf} values associated with those blue points are then a sample drawn from the posterior distribution of M_{hf} (following the ABC method); their density can be used to estimate the posterior density of M_{hf}.

In the lower panel of Fig. 5 we plot the posterior density for $\log_{10}M_{\mathrm{hf}}$, where the vertical dashed red line shows the true input, $M_{\mathrm{hf}} = 10^7\,\mathrm{M}_\odot$. It is clear that with just one system and the data quality of our fiducial setup, we are not able to derive a tight constraint on M_{hf}. While improved data quality might improve this somewhat, a single lens system will always be limited by the stochastic nature of the distribution of the low-mass dark matter haloes we are trying to detect. Just because a particular value of M_{hf} allows perturbers of a given mass this does not mean that there will happen to be one in a location where it produces detectable image residuals. As such, tight constraints will rely on combining results from multiple systems.

3 RESULTS AND DISCUSSION

In this Section, we will first show that the forward modelling procedure described above can correctly recover the input value of M_{hf} when combining results from multiple lenses. Then we will explore how the precision of the constraints depends upon the lensing configuration and image quality of the lenses in an observed sample. We will then compare the strength of our method with the method based on quasar flux ratio anomalies.

3.1 Tests of the accuracy of the method

By repeating the procedure described above for each observed system we can obtain a constraint from our mock observations of 50 lensing systems. Just as for a real set of 50 observed systems, the posterior probability density for M_{hf} will not necessarily peak at the true value. To assess whether our method is biased, we create 500 sets of 50 observed lenses. Running the full procedure described

above 500 times, each time for 50 different lenses, would be prohibitively expensive because of the very large number of forward models this would require. We therefore use the same set of 50 macro lens and source model parameters ($\{T\}_{\mathrm{fiducial}}$) for each of our 500 sets, only changing the realisations of low-mass dark matter haloes between the sets of 50 observed images. This allows us to reuse the same forward models for each of our 500 sets.

In principle, we should generate a new set of forward models for each of the 500 sets (together 25000 lenses) because the particular realisation of the dark matter haloes affects the best-fit macro model, which is the model used to generate the forward model images. We will see later that even though there is a slight mismatch in best-fit models for different realisations, we can still correctly recover the true M_{hf}. Similarly, to examine how the method behaves with different input (true) values for M_{hf}, we further simulate 500 sets of 50 observations with $M_{\mathrm{hf}} = 10^8$ and $10^9\,\mathrm{M}_\odot$ and apply ABC inference to them, all using the same forward models previously simulated.

Note that some care needs to be taken when combining results from multiple ABC calculations. This is because ABC produces an estimate of the posterior distribution, rather than the likelihood. Combining multiple measurements in a traditional Bayesian analysis (with a calculable likelihood function) is a case of multiplying the likelihoods of the different measurements together to get a total likelihood and then multiplying this by the prior in order to find (something proportional to) the posterior distribution. In this work we use a flat prior on $\log_{10}M_{\mathrm{hf}}$, and thus the posterior density (per unit $\log_{10}M_{\mathrm{hf}}$) we obtain is proportional to the likelihood. This means that the posterior densities per unit $\log_{10}M_{\mathrm{hf}}$ for individual systems (such as the blue histograms in the lower panel of Fig. 5) can be multiplied together to produce something proportional to the joint likelihood, which can then be multiplied by a prior (once) to get the posterior.

Before multiplying individual constraints, to reduce the noise and obtain a smooth likelihood, we apply a kernel density estimation method to the distribution of $\log_{10}M_{\mathrm{hf}}$ with a kernel width of 0.3 dex. We have tested smaller kernel sizes, but the constraints quickly become noisy without significant improvement. To correct for boundary effects, we choose the "renormalization" correction for our kernel density estimations.[3] As an example, the kernel density estimation for the fiducial result is shown as the black dashed curve in the lower panel of Fig. 5.

The left column of Fig. 6 shows our test results for three different cases with true input M_{hf} of $10^7\,\mathrm{M}_\odot$, $10^8\,\mathrm{M}_\odot$ and $10^9\,\mathrm{M}_\odot$ from top to bottom, respectively. Each blue line is a constraint from 50 systems (for clarity, we only show 50 blue lines in the plots). Any one set of 50 observations will not necessarily have a posterior that peaks at the true value, so in order to assess whether our results are systematically biased – as opposed to just subject to random error – we plot a mean posterior distribution from the 500 sets of 50 observations. These are the red lines in Fig. 6; they peak close to the input values of M_{hf}, suggesting that our method is unbiased. Using these mean posteriors we compute the ratio of the posterior between the peak value and the values at $M_{\mathrm{hf}} = 10^7\,\mathrm{M}_\odot$ and $M_{\mathrm{hf}} = 10^9\,\mathrm{M}_\odot$. With an input value of $M_{\mathrm{hf}} = 10^7\,\mathrm{M}_\odot$, the recovered constraint shows that the model with $M_{\mathrm{hf}} = 10^9\,\mathrm{M}_\odot$ is disfavoured with a posterior density ~ 1951 times smaller than that of the peak. In a WDM universe (e.g. with $M_{\mathrm{hf}} = 10^9\,\mathrm{M}_\odot$), $M_{\mathrm{hf}} = 10^7\,\mathrm{M}_\odot$ would

---

[3] We use the public package **PyQt-Fit** to implement the kernel density estimation (see "https://pythonhosted.org/PyQt-Fit/").
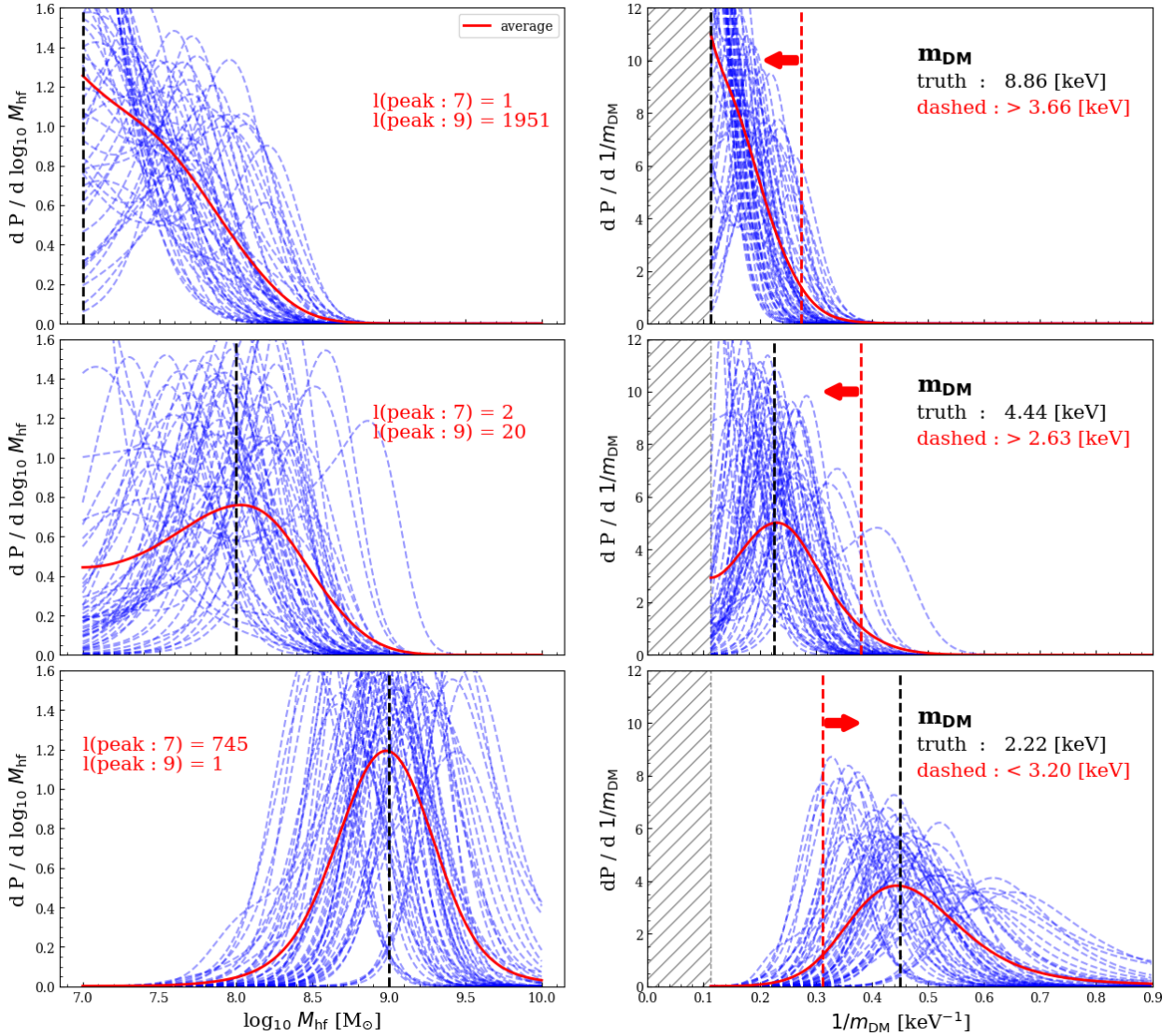
Figure 6. Tests on the fiducial setting. The left column shows the constraints on $M_{\rm hf}$ (adopting a flat prior on $\log M_{\rm hf}$) and the right columns the constraints on $1/m_{\rm DM}$ (adopting a flat prior on $1/m_{\rm DM}$). From top to bottom, panels correspond to true $M_{\rm hf}$ values of 10^7, 10^8, 10^9 M$_\odot$, marked by the black vertical dashed lines. Blue dashed curves are individual constraints for each set of 50 lensing systems (for clarity, we only plot 50 out of 500 here). The red curve is the average constraint from the 500 sets. In the right column we use the mean posteriors to place 95% credible interval limits (upper limits on $1/m_{\rm DM}$ for the top 2 panels and a lower limit for the bottom one). The hatched regions are where $M_{\rm hf} < 10^7$ M$_\odot$, which is outside the range probed by our forward models.

be ruled out, with a mean posterior density ~ 745 times smaller than at the peak.

In the right column of Fig. 6, we show our posterior distribution in terms of $1/m_{\rm DM}$, which is the way in which constraints on the DM particle mass from the Lyman-α forest are typically expressed (Iršič et al. 2017). Note that the $1/m_{\rm DM}$ posteriors are not simply the $M_{\rm hf}$ posteriors transformed to a new parameterisation. Instead, we transform the likelihood as a function of $M_{\rm hf}$ to the likelihood as a function of $1/m_{\rm DM}$ following Eq. 5, and then adopt a flat prior on $1/m_{\rm DM}$ (as done in Lyman-α studies). This is different from a flat prior on $\log_{10} M_{\rm hf}$ and so the posteriors are not actually the same in the two columns of Fig. 6.

Depending on the "behaviour" of the red curves, we place either a 95% upper or lower limit on $1/m_{\rm DM}$. For the cases with true $M_{\rm hf} = 10^7$ and 10^8 M$_\odot$, we place an upper limit, while for the case with $M_{\rm hf} = 10^9$ M$_\odot$ we place a lower limit. These limits

are shown as vertical dashed red lines in the figure. We can see that with 50 fiducial-like lensing systems, at 2σ level, one can rule out particle candidates with $m_{\rm DM}$ less than 3.66 keV and 2.63 keV in universes with true $m_{\rm DM} = 8.86$ and 4.44 keV respectively, or rule out particles with $m_{\rm DM} > 3.20$ keV in a universe with true $m_{\rm DM} = 2.22$ keV. We view the 2σ limits in the first two panels as conservative because we do not assign any posterior mass to the shaded regions in our computation. If we had simply assumed that the posterior density in the shaded region was the same as at ~ 0.12 keV^{-1}, the upper limits would have been tighter.

3.2 Dependency on the lensing configuration

Having shown that with a large number of lensing systems the forward modelling procedure can correctly recover the true $M_{\rm hf}$ (albeit with fairly broad posteriors), we now explore how the constraints
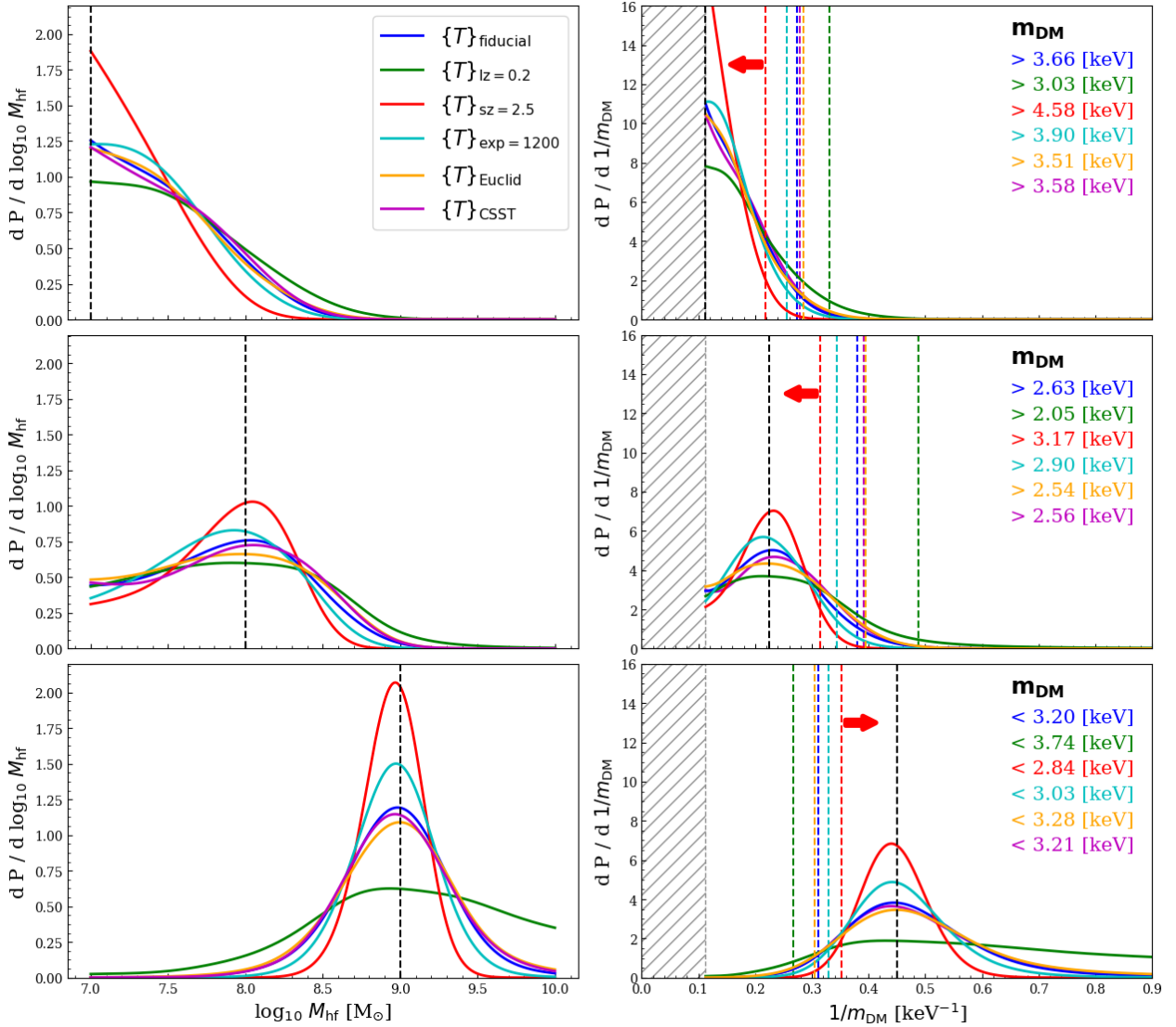
Figure 7. Average constraints from 50 lensing systems with different settings. As in Fig. 6, the left panels show the constraints on $M_{\rm hf}$ and the right panels the constraints on $1/m_{\rm DM}$. Blue, green, red, cyan, orange and purple lines correspond to the results of settings, $\{T\}_{\rm fiducial}$, $\{T\}_{\rm lz=0.2}$, $\{T\}_{\rm sz=2.5}$, $\{T\}_{\rm exp=1200}$, $\{T\}_{\rm Euclid}$ and $\{T\}_{\rm CSST}$ respectively. The vertical dashed black lines mark the true input values and the color lines the corresponding 2σ upper limits (top 2 panels) or lower limit (bottom panel) on $1/m_{\rm DM}$.

change when the properties of the lenses and sources are varied. In particular, we vary the lens and source redshifts, the image S/N ratio, and the image resolution.

To show clearly how the constraints depend upon a specific parameter, we only change one parameter value at a time. For example, to investigate how the result changes for a higher redshift source, we only change the source redshift, and leave the image resolution, Einstein radius, surface brightness, etc., unchanged from the $\{T\}_{\rm fiducial}$ values. Notice that the parameters listed in Table 1 are observational rather than intrinsic physical quantities, so by fixing a parameter, we refer to fixing a particular observational quantity. For example, keeping the source surface brightness unchanged from a redshift of 1 to 2.5 means that the source galaxy at $z = 2.5$ is actually intrinsically brighter because of cosmological dimming. The reason for changing quantities in this way is that we want to focus on how the results change from an observational perspective and thus provide a basic idea of what type of lensing configuration has most

constraining power, which may help inform future observational designs.

In total, we have carried out five additional tests: placing the lens at a lower redshift of 0.2, denoted as $\{T\}_{\rm lz=0.2}$; placing the source at a higher redshift of 2.5, denoted as $\{T\}_{\rm sz=2.5}$; doubling the exposure time, denoted as $\{T\}_{\rm exp=1200}$; lowering the resolution, with pixel size of $0.1''$ and PSF σ of $0.08''$, similar to the expected resolution of images from the Euclid Space Telescope (Collett 2015), denoted as $\{T\}_{\rm Euclid}$; assuming a similar resolution to that of the China Space Station Telescope (CSST), with pixel size of $0.075''$ and PSF σ of $0.08''$, denoted $\{T\}_{\rm CSST}$. To focus on the changes caused by different settings, without being affected by set-of-observations to set-of-observations noise, we compare the mean posteriors from 500 sets of 50 observed systems (i.e. the equivalents of the red lines in Fig. 6).

In the left column of Fig. 7 we plot the expected $M_{\rm hf}$ posteriors for different lensing configuration using different colours.

Blue, green, red, cyan, orange and purple correspond to the settings, $\{T\}_{\text{fiducial}}$, $\{T\}_{\text{lz=0.2}}$, $\{T\}_{\text{sz=2.5}}$, $\{T\}_{\text{exp=1200}}$, $\{T\}_{\text{Euclid}}$ and $\{T\}_{\text{CSST}}$, respectively. Different rows show results assuming different true inputs of M_{hf}, which are marked as vertical dashed black lines. As in Fig. 6, we plot $1/m_{\text{DM}}$ posteriors in the right column and place upper (in the top two panels) or lower (in the bottom panel) 2σ limits on $1/m_{\text{DM}}$, which are marked as vertical dashed lines. The red, green and blue curves show how the results vary with the lens and source redshifts.

Increasing either the source or lens redshift improves the constraints because (at fixed Einstein radius) the volume in which low-mass dark matter haloes are projected close to the observed Einstein ring increases as either redshift is increased. The comparison between the cyan and blue curves shows that increasing the S/N ratio (exposure time) results in a better constraint. Also, better angular resolution gives better constraints when comparing the fiducial setting (blue curve) and the Euclid / CSST resolution setting (orange / purple curves), but the improvement is not significant. Although a worse constraint is obtained with lower image resolution, we see that even with lower angular resolution we can distinguish between models with M_{hf} of 10^7 M$_\odot$ and 10^8 M$_\odot$. The advantage of Euclid and CSST will be sample sizes that are vastly larger than 50.

To summarize, the tests shown here agree with the expectation that lensing systems with longer exposure times, higher resolution, and more low-mass dark matter haloes (higher source redshifts and a larger area around the Einstein arcs) give better constraints.

3.3   Model assumptions and limitations

Our method makes a number of simplifying assumptions which we now summarize and describe in more detail.

By simulating low-mass dark matter haloes with the mass function of Eq. 4 we neglect lensing perturbations from subhaloes within the lens galaxy. This is a good approximation since, as Li et al. (2017) have shown, for realistic lensing configurations the signal is dominated by line-of-sight haloes, rather than by subhaloes in the lens. In any case, this assumption makes our results conservative, as including subhaloes would boost the lensing perturbations, increasing the signal we can extract from the lens model residuals. On the other hand, there are uncertainties regarding subhaloes that do not affect line-of-sight haloes, for example, the extent of tidal disruption (Despali et al. 2018; Richings et al. 2020b); marginalising over these uncertainties would weaken the constraint on M_{hf}.

We neglect any uncertainties on the amplitude and shape of the (CDM) halo mass function. The amplitude is fixed by the value of the cosmological parameter, σ_8, which is known to better than 1% from cosmic microwave background data (Planck Collaboration et al. 2020). The shape is known very precisely from cosmological simulations (Wang et al. 2020). There is, however, some degeneracy between the amplitude of the halo mass function and the half-mode mass. Some lensing studies, such as that of Gilman et al. (2019) based on flux ratio anomalies, have included the amplitude of the mass function as a free parameter to be fit at the same time as the subhalo mass function.

In this work we have assumed an SIE lens mass model, whereas in studies of real lenses a power-law mass model with external shear is widely used (e.g. Vegetti et al. 2014; Dye et al. 2014). This gives the mass model more freedom to fit the residuals, reducing the signal left over from substructures. Even then, the mass model may only provide an approximate fit to the lens and result in residuals not associated with subhaloes, at which point even more complex mass profiles (Nightingale et al. 2019), or a potential corrections-based

approach may be required (Vegetti & Koopmans 2009). We assume a core-Sersic model for the source light, whereas studies using real data often use a non-parametric approach that reconstructs a source's irregular morphology (see Warren & Dye 2003; Suyu et al. 2006). Gilman et al. (2020a) show how such an approach may absorb part of the residual signal of the low-mass perturbers, reducing the available information on the halo mass function. PyAutoLens has all the necessary functionality to test these assumptions and this will be the topic of future work.

Our method implicitly attributes all image residuals to the presence of perturbing dark matter haloes when, in reality, there could be other sources of mismatch between a true lens and source model and the best-fit macro model. In the idealized setup used here we have shown that this allows us to obtain a correct measurement of the input half-mode mass. In reality, other sources of perturbations will contribute to the power spectrum of the image residuals, including a deficient model for the smooth component of the lens, an inaccurate description of the telescope PSF and artefacts or correlated noises introduced by the data reduction process. Not taking these effects into account may bias our inference of the subhalo mass function. Currently, the impact of any individual effect is unclear and further investigation is required, noting that all effects that impact the signal can, in principle, be included in our forward modeling procedure and marginalized over.

On a positive note, our use of the power spectrum to define a summary statistic to extract the signal from the residuals is likely to be suboptimal. A more carefully crafted summary statistic or a machine learning-based approach (see Brehmer et al. 2019; Diaz Rivero & Dvorkin 2020) can potentially improve the signal that can be extracted from a lens system and thus provide better constraints on M_{hf} than shown here. It could also make the estimation less sensitive to systematic sources of residuals as shown by Birrer et al. (2017b), and so increase the constraining power of the forward modelling method.

3.4   Comparison with flux ratio anomalies

A theoretical investigation of the constraining power of flux ratio anomalies was performed by Gilman et al. (2019), who subsequently applied their method to real observations, placing constraints on both the dark matter mass function and the mass-concentration relation (Gilman et al. 2020b,c). It is interesting to compare our approach with that of Gilman et al. (2019). Note, however, that, as we just discussed, these authors treat the amplitude and slope of the halo mass function as free parameters, whereas our tests assume a halo mass function with only one free parameter, M_{hf}. Gilman et al. (2019) assume an SIE plus external shear mass model whereas we only assume an SIE. In our tests, for a universe with true $M_{\text{hf}} = 10^7$ M$_\odot$, our fiducial calculation gives an upper 2σ constraint, $M_{\text{hf}} = 10^{8.28}$ M$_\odot$, which is slightly smaller than the upper constraints quoted by Gilman et al., $M_{\text{hf}} = 10^{8.34}$ M$_\odot$, for a 2% uncertainty on flux measurements (see Figure 8 of Gilman et al. 2019). Given the different assumptions in the two studies this comparison serves to show that the constraints that they provide are roughly comparable.

When considering the relative performance of our method and that based on flux ratio anomalies it is pertinent to consider the mass scales to with each method is sensitive. The "coldest" case in our tests assumes $M_{\text{hf}} = 10^7$ M$_\odot$ and our signal is heavily influenced by the "detections" of larger perturbers with $M_{200} \sim 10^{7-8}$ M$_\odot$. The small detection area of flux ratio anomalies results in a lack of sensitivity to subhaloes of such high masses, but the

point-like sources involved allows sensitivity down to dark matter haloes of mass, $M_{200} \sim 10^6$ M$_\odot$, a scale that is not accesible with our approach. The two approaches are therefore complementary and, together, can provide constraints on the halo mass function over a broad range of masses.

4   CONCLUSIONS

The existence of a large population of dark matter haloes of all masses down to subsolar values is a fundamental prediction of the CDM model that distinguishes it from other currently popular models such as WDM in which the halo mass function is truncated below a mass $\sim 10^8$ M$_\odot$. Detecting the predicted population is a key test of the standard cosmological model which would be ruled out if it could conclusively be shown that low-mass haloes are absent or that they are present in lower numbers than predicted.

In this work we have presented a method based on forward modelling applied to the analysis of resolved strong lensing arc data that can constrain a cutoff in the dark matter halo mass function. The key idea is to simulate a large number of strong lensing images (forward models) based on the best-fit macro model, with the addition of perturbations to the lens model from low-mass dark matter haloes (whose number depends on the assumed nature of the dark matter). These images can then be fit in the same manner as the original observed image, and the best-fit image residuals for the observed system can be compared with the best-fit image residuals for the forward models. This comparison is made by means of the power spectrum of the image residuals within the Approximate Bayesian Computation framework and leads to a posterior distribution for the half-mode mass ($M_{\rm hf}$) which describes the cutoff in the halo mass function. Our main results may be summarized as follows:

• For mock observed lenses constructed from parametric source light profiles and lens mass models, we confirm that information on low-mass dark matter haloes can be extracted from the power spectrum of the best-fit image residuals. As demonstrated in Fig. 6, with a large number of observed lenses, a forward modelling procedure can correctly recover the true input half-mode mass, $M_{\rm hf}$ (the red curves). However, the scatter in the constraint is significant even though results for 50 systems have been combined together (blue curves), which may be attributed to the large intrinsic scatter in realisations of low-mass dark matter haloes. Averaging the constraints from 50 lensing systems, for our fiducial settings (see Table 1), we find that if the true (thermal) $m_{\rm DM}$ is 8.86 keV ($M_{\rm hf}$ = 10^7 M$_\odot$), particles with $m_{\rm DM}$ less than 3.66 keV can be ruled out at the 2σ level. Conversely, in a WDM universe where the true (thermal) $m_{\rm DM}$ is 2.22 keV ($M_{\rm hf} = 10^9$ M$_\odot$), particles with $m_{\rm DM}$ larger than 3.20 keV can be ruled out at the 2σ level.

• We have tested the dependency of the method on different lensing configurations and image quality settings. As shown in Fig. 7, the dependency agrees with expectations: higher redshift sources and/or larger areas around the Einstein arcs, and better data quality (longer exposure time and/or higher resolution) all result in a tighter constraint. Among our tests, the one with sources placed at $z = 2.5$ produced the strongest constraint: in a universe with true $m_{\rm DM}$ = 8.86 keV ($M_{\rm hf}$ = 10^7 M$_\odot$), particles with mass less than 4.58 keV can be ruled out at the 2σ level. Although a slightly worse constraint is obtained from images with the resolution of Euclid/CSST, 50 systems can still provide a constraint on $M_{\rm hf}$, and with Euclid/CSST data we will have many more than 50 strong lenses.

Throughout this study we have made several simplifying assumptions, particularly in using parametric models for the lens mass distributions and the source light profiles. The effects of removing these assumptions will need to be investigated before we can safely apply our technique to real observations. We have shown that, in principle, information on low-mass dark matter haloes can be statistically extracted from the power spectrum of the image residuals. In the near future, we hope that using a more powerful summary statistic together with more advanced lens modelling techniques (such as pixelized sources) can improve the power of this technique. With thousands of new strong lensing observations expected from future space telescopes, there is every prospect of pinning down the mass of the dark matter particles using this forward modelling technique.

SOFTWARE CITATIONS

This work uses the following software packages:

- Astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018)
- Colossus (Diemer 2018b)
- corner.py (Foreman-Mackey 2016)
- hmf (Murray et al. 2013)
- matplotlib (Hunter 2007)
- numba (Lam et al. 2015)
- NumPy (van der Walt et al. 2011)
- PyAutoFit (Nightingale & Hayes 2020a)
- PyAutoLens (Nightingale & Dye 2015; Nightingale et al. 2018b; Nightingale & Hayes 2020b)
- PyMulitNest (Feroz et al. 2009b; Buchner et al. 2014b)
- PyQt-Fit (Reuille 2015)
- pyquad (Kelly 2020)
- Python (Van Rossum & Drake 2009)
- scikit-image (Van der Walt et al. 2014)
- scikit-learn (Pedregosa et al. 2011)
- Scipy (Virtanen et al. 2020)

REFERENCES

Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, J. Cosmology Astropart. Phys., 2015, 043

Amara A., Metcalf R. B., Cox T. J., Ostriker J. P., 2006, MNRAS, 367, 1367

Astropy Collaboration et al., 2013, A&A, 558, A33

Baltz E. A., Marshall P., Oguri M., 2009, J. Cosmology Astropart. Phys., 2009, 015

Bayer D., Chatterjee S., Koopmans L. V. E., Vegetti S., McKean J. P., Treu T., Fassnacht C. D., 2018, arXiv e-prints, p. arXiv:1803.05952

Benitez-Llambay A., Frenk C., 2020, arXiv e-prints, p. arXiv:2004.06124

Benson A. J., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2002, MNRAS, 333, 177

Birrer S., Welschen C., Amara A., Refregier A., 2017a, J. Cosmology Astropart. Phys., 2017, 049

Birrer S., Amara A., Refregier A., 2017b, J. Cosmology Astropart. Phys., 2017, 037

Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, ApJ, 638, 703

Bonaca A., Hogg D. W., Price-Whelan A. M., Conroy C., 2019, ApJ, 880, 38

Bose S., Hellwing W. A., Frenk C. S., Jenkins A., Lovell M. R., Helly J. C., Li B., 2016, MNRAS, 455, 318

Brehmer J., Mishra-Sharma S., Hermans J., Louppe G., Cranmer K., 2019, ApJ, 886, 49

Brewer B. J., Huijser D., Lewis G. F., 2016, MNRAS, 455, 1819

Buchner J., et al., 2014a, A&A, 564, A125

Buchner J., et al., 2014b, Astronomy and Astrophysics, 564, A125

Chatterjee S., Koopmans L. V. E., 2018, MNRAS, 474, 1762

Collett T. E., 2015, ApJ, 811, 20

Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., 2019, Phys. Rev. D, 100, 023013

Dalal N., Kochanek C. S., 2002, ApJ, 572, 25

Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018, ApJ, 855, 106

Despali G., Vegetti S., 2017, MNRAS, 469, 1997

Despali G., Vegetti S., White S. D. M., Giocoli C., van den Bosch F. C., 2018, MNRAS, 475, 5424

Diaz Rivero A., Dvorkin C., 2020, Phys. Rev. D, 101, 023515

Díaz Rivero A., Dvorkin C., Cyr-Racine F.-Y., Zavala J., Vogelsberger M., 2018, Phys. Rev. D, 98, 103517

Diemer B., 2018a, ApJS, 239, 35

Diemer B., 2018b, The Astrophysical Journal Supplement Series, 239, 35

Dye S., et al., 2014, MNRAS, 440, 2013

Efstathiou G., 1992, MNRAS, 256, 43P

Enzi W., Vegetti S., Despali G., Hsueh J. W., Metcalf R. B., 2019, arXiv e-prints, p. arXiv:1911.02581

Fearnhead P., Prangle D., 2012, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74, 419

Feroz F., Hobson M. P., Bridges M., 2009a, MNRAS, 398, 1601

Feroz F., Hobson M. P., Bridges M., 2009b, Monthly Notices of the Royal Astronomical Society, 398, 1601

Foreman-Mackey D., 2016, The Journal of Open Source Software, 1, 24

Frenk C. S., White S. D. M., 2012, Annalen der Physik, 524, 507

Frenk C. S., White S. D. M., Davis M., 1983, ApJ, 271, 417

Frenk C. S., White S. D. M., Efstathiou G., Davis M., 1985, Nature, 317, 595

Garrison-Kimmel S., et al., 2017, MNRAS, 471, 1709

Garzilli A., Magalich A., Theuns T., Frenk C. S., Weniger C., Ruchayskiy O., Boyarsky A., 2019, MNRAS, 489, 3456

Gilman D., Birrer S., Treu T., Nierenberg A., Benson A., 2019, MNRAS, 487, 5721

Gilman D., Birrer S., Treu T., 2020a, arXiv e-prints, p. arXiv:2007.01308

Gilman D., Birrer S., Nierenberg A., Treu T., Du X., Benson A., 2020b, MNRAS, 491, 6077

Gilman D., Du X., Benson A., Birrer S., Nierenberg A., Treu T., 2020c, MNRAS, 492, L12

Graham A. W., Erwin P., Trujillo I., Asensio Ramos A., 2003, AJ, 125, 2951

Hahn C., Vakili M., Walsh K., Hearin A. P., Hogg D. W., Campbell D., 2017, MNRAS, 469, 2791

Hezaveh Y. D., et al., 2016, ApJ, 823, 37

Hinshaw G., et al., 2013, ApJS, 208, 19

Hsu D. C., Ford E. B., Terrien R., 2020, arXiv e-prints, p. arXiv:2002.02573

Hsueh J. W., Enzi W., Vegetti S., Auger M. W., Fassnacht C. D., Despali G., Koopmans L. V. E., McKean J. P., 2020, MNRAS, 492, 3047

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Iršič V., et al., 2017, Phys. Rev. D, 96, 023522

Kaiser N., 1984, ApJ, 284, L9

Kelly A. J., 2020, pyquad, doi:10.5281/zenodo.3936959, https://doi.org/10.5281/zenodo.3936959

Lam S. K., Pitrou A., Seibert S., 2015, Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15, pp 1–6

Li R., Frenk C. S., Cole S., Gao L., Bose S., Hellwing W. A., 2016, MNRAS, 460, 363

Li R., Frenk C. S., Cole S., Wang Q., Gao L., 2017, MNRAS, 468, 1426

Lovell M. R., et al., 2012, MNRAS, 420, 2318

Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300

Lovell M. R., et al., 2016, MNRAS, 461, 60

Lovell M. R., et al., 2017, MNRAS, 468, 4285

Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, MNRAS, 460, 1214

Macciò A. V., Ruchayskiy O., Boyarsky A., Muñoz-Cuartas J. C., 2013, MNRAS, 428, 882

Mao S., Schneider P., 1998, MNRAS, 295, 587

McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2014, MNRAS, 443, 3631

McKean J., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 84 (arXiv:1502.03362)

Metcalf R. B., Madau P., 2001, ApJ, 563, 9

Murray S. G., Power C., Robotham A. S. G., 2013, Astronomy and Computing, 3, 23

Nadler E. O., et al., 2020, arXiv e-prints, p. arXiv:2008.00022

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Neto A. F., et al., 2007, MNRAS, 381, 1450

Nierenberg A. M., Treu T., Wright S. A., Fassnacht C. D., Auger M. W., 2014, MNRAS, 442, 2434

Nierenberg A. M., et al., 2017, MNRAS, 471, 2224

Nightingale J. W., Dye S., 2015, MNRAS, 452, 2940

Nightingale J., Hayes R., 2020a, PyAutoFit: A classy Probablistic Programming Language for Complex Model-fitting, https://github.com/rhayes777/PyAutoFit

Nightingale J., Hayes R., 2020b, PyAutoLens: Open-source Strong Gravitational Lensing, https://github.com/Jammy2211/PyAutoLens

Nightingale J. W., Dye S., Massey R. J., 2018a, MNRAS, 478, 4738

Nightingale J. W., Dye S., Massey R. J., 2018b, Monthly Notices of the Royal Astronomical Society, 478, 4738

Nightingale J. W., Massey R. J., Harvey D. R., Cooper A. P., Etherington A., Tam S.-I., Hayes R. G., 2019, MNRAS, 489, 2049

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Petkova M., Metcalf R. B., Giocoli C., 2014, MNRAS, 445, 1954

Planck Collaboration et al., 2020, A&A, 641, A6

Price-Whelan A. M., et al., 2018, AJ, 156, 123

Reuille P. B., 2015, PyQt-fit, doi:https://pythonhosted.org/PyQt-Fit/index.html, https://pythonhosted.org/PyQt-Fit/index.html

Richings J., Frenk C., Jenkins A., Robertson A., 2020a, arXiv e-prints, p. arXiv:2005.14495

Richings J., et al., 2020b, MNRAS, 492, 5780

Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2019, MNRAS, 485, 2179

Sawala T., et al., 2016, MNRAS, 456, 85

Sawala T., Pihajoki P., Johansson P. H., Frenk C. S., Navarro J. F., Oman K. A., White S. D. M., 2017, MNRAS, 467, 4383

Schneider P., 2014, Astronomy and Astrophysics, 568, L2

Schneider A., 2016, J. Cosmology Astropart. Phys., 2016, 059

Schneider A., Smith R. E., Macciò A. V., Moore B., 2012, MNRAS, 424, 684

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Sheth R. K., Tormen G., 1999, MNRAS, 308, 119

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Suyu S. H., Marshall P. J., Hobson M. P., Bland ford R. D., 2006, MNRAS, 371, 983

Thoul A. A., Weinberg D. H., 1996, ApJ, 465, 608

Trujillo I., Erwin P., Asensio Ramos A., Graham A. W., 2004, AJ, 127, 1917

Van Rossum G., Drake F. L., 2009, Python 3 Reference Manual. CreateSpace, Scotts Valley, CA

Van der Walt S., Schönberger J. L., Nunez-Iglesias J., Boulogne F., Warner J. D., Yager N., Gouillart E., Yu T., 2014, PeerJ, 2, e453

Vegetti S., Koopmans L. V. E., 2009, MNRAS, 400, 1583

Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2010, MNRAS, 408, 1969

Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, Nature, 481, 341

Vegetti S., Koopmans L. V. E., Auger M. W., Treu T., Bolton A. S., 2014, MNRAS, 442, 2017

Vegetti S., Despali G., Lovell M. R., Enzi W., 2018, MNRAS, 481, 3661

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502

Virtanen P., et al., 2020, Nature Methods, 17, 261

Wang J., Bose S., Frenk C. S., Gao L., Jenkins A., Springel V., White S. D. M., 2020, Nature, 585, 39

Warren S. J., Dye S., 2003, ApJ, 590, 673

White S. D. M., Frenk C. S., Davis M., 1983, ApJ, 274, L1

Xu D. D., Mao S., Cooper A. P., Gao L., Frenk C. S., Angulo R. E., Helly J., 2012, MNRAS, 421, 2553

Xu D., Sluse D., Gao L., Wang J., Frenk C., Mao S., Schneider P., Springel V., 2015, MNRAS, 447, 3189

Çağan Şengül A., Tsang A., Diaz Rivero A., Dvorkin C., Zhu H.-M., Seljak U., 2020, arXiv e-prints, p. arXiv:2006.07383

van der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science Engineering, 13, 22

APPENDIX A: DETERMINING MULTIPLANE RESOLUTION

We simulate the effect of line-of-sight haloes by dividing the light-cone from the observer to the source galaxy into a number of intervals and approximating low-mass dark matter haloes within a a given redshift interval as lying on a single plane at the central redshift. Reducing the redshift interval, Δz, between two neighbouring planes (increasing the number of multiplanes) will increase the accuracy of the approximation to the line-of-sight effects, but will also increase the computational cost. The smallest allowable Δz should be determined by the accuracy one would like to achieve in the inference of M_{hf}. In practice, one needs to make sure the "difference" caused by approximating line-of-sight effects into multiplanes is smaller than the "difference" caused by changing M_{hf} by the amount one would like to distinguish.

Fig. A1 shows an example of the amplitudes of the differences in deflection angles when changing the values of Δz and M_{hf}. The two panels are computed from the same macro settings (a spherical isothermal sphere model with R_E of $1.5''$ at $z = 0.5$ and a source at $z = 1.0$) and the same low-mass dark matter haloes (generated from a distribution with $M_{hf} = 10^7 \, M_\odot$). To simplify the test, instead of using the mass function described in the main body of the paper, we impose a "sharp cut" on the mass function, such that no haloes are drawn with M_{200} smaller than M_{hf}. Also, we assume that there is no correlation between M_{hf} and the mass-concentration relation to make sure massive haloes are unchanged when changing M_{hf}.

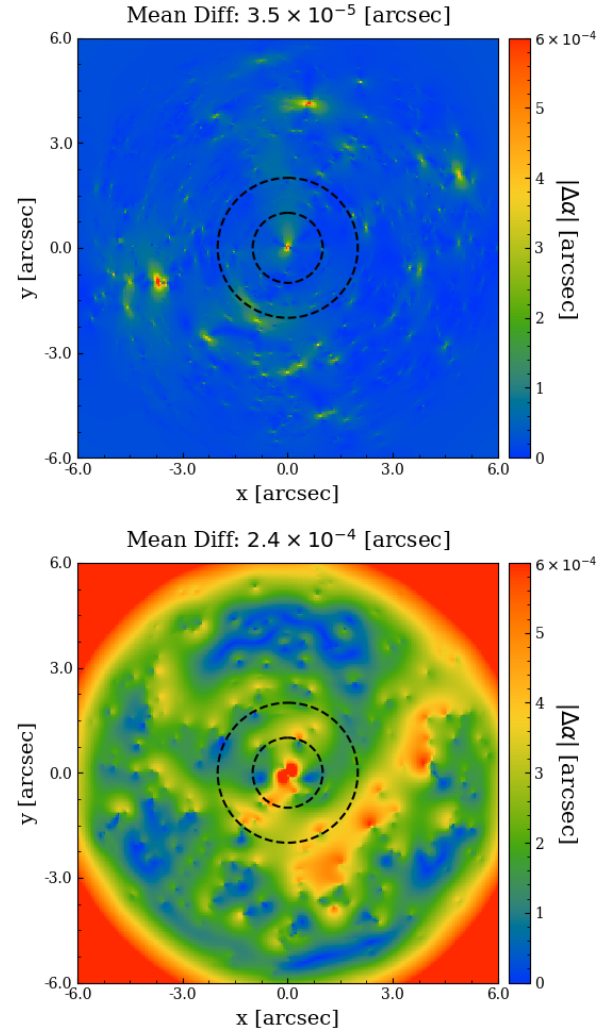The upper panel of Fig. A1 shows the amplitudes of the differences in deflection angles when changing Δz from 0.01 to 0.001

Figure A1. Upper panel: amplitude of the differences in deflection angles caused by changing Δz from 0.01 to 0.001. **Lower panel**: amplitude of the differences in deflection angles caused by changing M_{hf} from $10^7 \, M_\odot$ to $10^{7.1} \, M_\odot$. The region between two dashed circles is the annulus between $1.0''$ and $2.0''$. The unit of the colourbar is arcsecs.

(a very high resolution setting which we take as the "accurate" result). The lower panel shows the amplitudes of the differences in deflection angles when changing M_{hf} from $10^7 \, M_\odot$ to $10^{7.1} \, M_\odot$. As seen from the figure, the patterns of the differences caused by changing Δz and M_{hf} are different. We use the mean amplitude of the difference in deflection angles within the annular region $1.0''$ – $2.0''$ (the region between the two dashed circles in the figure) to quantify the "differences". The mean difference caused by changing Δz is 3.47×10^{-5} arcsec, which is smaller than the mean difference caused by changing M_{hf} by 0.1 dex, 2.42×10^{-4} arcsec, suggesting that Δz of 0.01 would not affect the ability to distinguish between models with $M_{hf} = 10^7 \, M_\odot$ and $10^{7.1} M_\odot$.

The results shown in Fig. A1 are for a special case, so we repeat the same procedure for 128 different realisations of low-mass dark matter haloes. The orange histogram in Fig. A2 shows the distribution of the $\log_{10}$ ratio of the mean difference in amplitudes defined previously, that is the ratio of the mean difference in the deflection angles between models with $M_{hf} = 10^7$ and $M_\odot = 10^{7.1}$

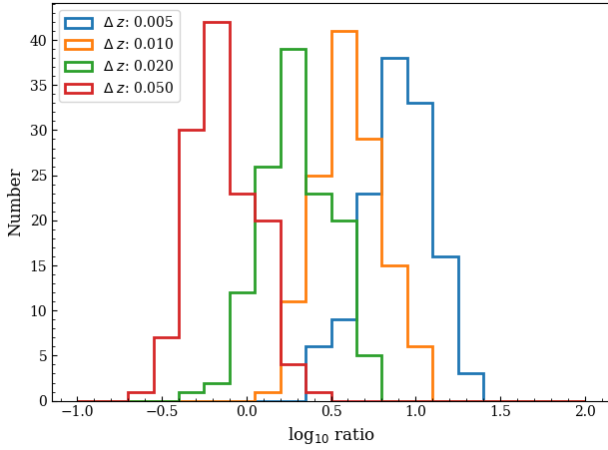**Figure A2.** Histograms of the $\log_{10}$ ratio of the mean amplitude of the differences in deflection angles caused by changing Δz and $M_{\rm hf}$. Blue, orange, green and red colors shows results for $\Delta z = 0.005$, 0.01, 0.02 and 0.05 respectively.

divided by the mean difference in the deflection angles for the Δz given in the legend. As shown by the orange histogram, the ratios from all of the 128 realisations are larger than one, indicating that the differences caused by approximating line-of-sight effects using multiplanes separated by 0.01 in redshift is smaller than the difference cause by changing $M_{\rm hf}$ by 0.1 dex. Also, we plot the histogram of ratios for other values of Δz. As may be seen, increasing Δz increases the differences caused by changing Δz and, in some cases, the differences are larger than those caused by changing $M_{\rm hf}$ by 0.1 dex. To summarize, the value of Δz we adopt for our tests, $\Delta z = 0.01$, does not affect our inference of $M_{\rm hf}$ to an accuracy of 0.1 dex.

This paper has been typeset from a TeX/LaTeX file prepared by the author.