Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **518,** 2746–2760 (2023) https://doi.org/10.1093/mnras/stac3215
Advance Access publication 2022 November 7

Estimating the warm dark matter mass from strong lensing images with truncated marginal neural ratio estimation

Noemi Anau Montel [1]★ Adam Coogan,[1,2,3]★ Camila Correa [1] Konstantin Karchev [1,4] and Christoph Weniger[1]★

[1]*GRAPPA (Gravitation Astroparticle Physics Amsterdam), University of Amsterdam, Science Park 904, NL-1098 XH Amsterdam, the Netherlands*
[2]*Département de Physique, Université de Montréal, 1375 Avenue Thérèse-Lavoie-Roux, Montréal, QC H2V 0B3, Canada*
[3]*Mila – Quebec AI Institute, 6666 St-Urbain, 200, Montreal, QC H2S 3H1, Canada*
[4]*SISSA (Scuola Internazionale Superiore di Studi Avanzati), via Bonomea 265, I-34136 Trieste, Italy*

Accepted 2022 November 2. Received 2022 November 1; in original form 2022 June 7

ABSTRACT

Precision analysis of galaxy–galaxy strong gravitational lensing images provides a unique way of characterizing small-scale dark matter haloes, and could allow us to uncover the fundamental properties of dark matter's constituents. Recently, gravitational imaging techniques made it possible to detect a few heavy subhaloes. However, gravitational lenses contain numerous subhaloes and line-of-sight haloes, whose subtle imprint is extremely difficult to detect individually. Existing methods for marginalizing over this large population of subthreshold perturbers to infer population-level parameters are typically computationally expensive, or require compressing observations into hand-crafted summary statistics, such as a power spectrum of residuals. Here, we present the first analysis pipeline to combine parametric lensing models and a recently developed neural simulation-based inference technique called truncated marginal neural ratio estimation (TMNRE) to constrain the warm dark matter halo mass function cut-off scale directly from multiple lensing images. Through a proof-of-concept application to simulated data, we show that our approach enables empirically testable inference of the dark matter cut-off mass through marginalization over a large population of realistic perturbers that would be undetectable on their own, and over lens and source parameter uncertainties. To obtain our results, we combine the signal contained in a set of images with *Hubble Space Telescope* resolution. Our results suggest that TMNRE can be a powerful approach to put tight constraints on the mass of warm dark matter in the multi-keV regime, which will be relevant both for existing lensing data and in the large sample of lenses that will be delivered by near-future telescopes.

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

1 INTRODUCTION

Over the past several decades, numerous astrophysical probes, including rotational curves of spiral galaxies (Rubin, Ford & Thonnard 1980), galaxy-cluster dynamics (Zwicky 1933), cosmic microwave background (Ade et al. 2016), and gravitational lensing observations (Taylor et al. 1998), have established dark matter (DM) as one of the major components of the Universe. However, up to the present time, the fundamental nature of DM is still an unresolved puzzle. For many years, the cold dark matter (CDM) paradigm (Peebles 1982) has been able to accurately reproduce vastly disparate large-scale observations across all epochs. In this model, DM is massive, neutral, non-relativistic, and collisionless. The main prediction of the CDM paradigm is that structure formation is due to a hierarchical clustering process, guided by gravitational instability of DM density perturbations, originated from quantum fluctuations during inflation.

Despite providing a stunning description of the observed distribution of matter on large scales ($> \mathcal{O}$(Mpc)), the agreement between CDM predictions and observations at galactic and subgalactic scales has been less clear. One of the most well-known small-scale discrepancies of CDM is the missing satellites problem (Moore et al. 1999). Numerical CDM simulations predict that a large population of DM subhaloes, spanning a wide range of masses, should be orbiting around all main DM haloes. However, we have observed a lot fewer small galaxies in the Local Group than the predicted subhaloes below $10^9\,M_\odot$ (Klypin et al. 1999).

Solutions to this tension include the impact of baryonic processes or alternative DM physics. Baryonic processes from supernova feedback and reionization processes suppress star formation in low-mass galaxies (Bullock 2010). As a result, most DM subhaloes would not contain sufficiently bright galaxies and thus are more difficult to detect. The other approach requires an alteration of DM particle physics, such that large-scale predictions remain unaffected, but the number of small-scale substructures is suppressed. One of the alternative models that has been proposed is warm dark matter (WDM; Colin, Avila-Reese & Valenzuela 2000; Lovell et al. 2014). Moreover, its main particle candidates, sterile neutrinos (Boyarsky

★ E-mail: n.anaumontel@uva.nl (NAM); adam.coogan@umontreal.ca (AC); c.weniger@uva.nl (CW)

et al. 2019) and gravitinos (Bond, Szalay & Turner 1982), are well motivated from a particle physics perspective. In WDM models, DM particles have non-negligible thermal velocities that allow them to free-stream out of density perturbations, effectively preventing small-scale structure formation. The scale at which this happens depends on model parameters and is parametrized by the half-mode mass $M_{\rm hm}$ in the halo mass function (HMF). Therefore, one of the viable way to discriminate between CDM and alternative DM models is to constrain the low-mass end of the HMF by probing small-scale DM haloes that are completely devoid of stars and truly *dark*, whose only signature is then gravitational.

Strong lensing images analysis. In strong gravitational lensing, the gravitational field of a mass distribution acts as a lens by distorting and magnifying the light flux coming from a background source (Kochanek 2004). This effect is sensitive only to how matter is distributed, regardless of its physical nature (baryonic/DM). Hence, it provides a direct way of probing the distribution of DM at small scales, by means of the distortions to the images due to substructures on top of the main lens mass distribution. Therefore, gravitational lensing provides a pristine probe of small-scale structures and can in principle distinguish between CDM and WDM scenarios.

Various different methods have been suggested to analyse the effects of small-scale structures on lensing images (Drlica-Wagner et al. 2019). These methods usually target two different types of lensing systems that differ in the lensed source: extended background galaxies that get lensed into extended arcs or complete Einstein rings, and almost point-like quasars that get lensed into multiple point-like projections.

Quasar lensing was first used in Mao & Schneider (1998) to constrain the amount of DM substructures by analysing the deviations in the relative fluxes of the multiple source projections from a smooth lens model. Later, Dalal & Kochanek (2002) derived a statistical constraint on the substructure fraction in the lensing galaxies using a small sample of seven lensed quasars. Nierenberg et al. (2014) showed that flux-ratio anomalies can also be used to detect individual low-mass subhaloes. Several studies derived upper limits on the half-mode mass $M_{\rm hm}$ (Nierenberg et al. 2017; Gilman et al. 2018), also including perturbations due to line-of-sight (LOS) haloes (Gilman et al. 2019a,b). Further investigations (Hsueh et al. 2016, 2017, 2019) pointed out the importance of correctly modelling baryonic structure in the main lens, in order to avoid systematic errors while constraining DM substructure abundance with flux-ratio anomalies.

On the other hand, in strong lens systems where the background source is a galaxy, massive substructures can leave a signature in the form of per cent-level variations in the shape of the predicted lensed light based on a smooth lens model. The gravitational imaging technique, which models these distortions, was first introduced in Koopmans (2005) and further developed in Vegetti & Koopmans (2009a,b). Its application to real data has led to several detections of individual heavy ($> 10^8\,{\rm M}_\odot$) subhaloes (Vegetti, Czoske & Koopmans 2010a; Vegetti et al. 2010b, 2012; Hezaveh et al. 2016a). Moreover, samples of gravitational lens systems have been analysed in Vegetti et al. (2014, 2018), and, including LOS halo modelling, in Ritondale et al. (2019), in order to derive constraints on the HMF using detections and non-detections of individual substructures.

A population of low-mass haloes can collectively cause perturbations to images that can be detected statistically in order to constrain the HMF. In reality, constraining collective substructure properties from gravitational lensing images is an extremely difficult problem. In fact, inferring marginal posteriors for the HMF cut-off requires marginalizing over all source, lens, and substructure parameters to get the marginal likelihood for the population-level parameter of interest, thus involving a time-consuming exploration of a very high-dimensional parameter space for complex realistic models. Therefore, Markov chain Monte Carlo (MCMC) or nested sampling methods would imply an intractable sampling from the high-dimensional joint posterior.

To partially overcome traditional likelihood-based methods' challenges, Brewer, Huijser & Lewis (2016) and Daylan et al. (2018) performed inference on subhaloes using a likelihood-based method called transdimensional Bayesian inference. This approach uses transdimensional MCMC sampling over the union of different models, with different numbers of subhaloes, to infer a probability for the subhalo catalogue.

In order to reduce the dimensionality of the problem and enable inference of the collective effects of a large number of low-mass substructures at the statistical level, Hezaveh et al. (2016b) proposed to use the power spectrum (PS) of the lensed deflection field. Subsequently, Diaz Rivero, Cyr-Racine & Dvorkin (2018a) developed a theoretical general formalism to compute the convergence PS for different subhalo populations from first principles, which was adopted in Díaz Rivero et al. (2018b) and Brennan et al. (2019). This formalism has been recently expanded to account for LOS haloes in Şengül et al. (2020). However, this approach is not directly applicable to observations, because we do not have access to the true displacement field from the data. Chatterjee & Koopmans (2017), Bayer et al. (2018), and Cyr-Racine, Keeton & Moustakas (2019) developed statistical formalisms to relate PS of the surface brightness fluctuations in strong lens images to the lens potential fluctuations arising from DM distribution that contribute to the convergence PS, and Bayer et al. (2018) applied it to a real observation.

Instead, Birrer, Amara & Refregier (2017) and He et al. (2022a) employed the residual PS summary statistic, given by the subtraction of a smooth lens model from the data, to constrain the half-mode mass $M_{\rm hm}$. For the analysis, they used approximate Bayesian computation, a likelihood-free inference method based on a rejection algorithm (Grazian & Fan 2019).

Another class of methods that has developed in recent years uses neural networks to measure lens parameters (Hezaveh, Levasseur & Marshall 2017; Perreault Levasseur, Hezaveh & Wechsler 2017; Morningstar et al. 2019), quantifying the structure of gravitational lens potential (Vernardos, Tsagkatakis & Pantazis 2020), detect individual subhaloes (Diaz Rivero & Dvorkin 2020), and distinguish different types of DM substructure (Alexander et al. 2020). Still, these methods need lots of data to amortize over all possible variations in lensing systems. In fact, amortized methods learn the posterior for any data, generated by any parameter over the whole range of the prior (Cranmer, Brehmer & Louppe 2020). However, learning an amortized posterior is unnecessary if only a small range of parameters are consistent with a target observation.

In this work, we present the first analysis pipeline that combines parametric lensing models with recent neural simulation-based inference developments (Cranmer et al. 2020) to infer the DM mass cut-off scale from a set of realistic simulated galaxy–galaxy strong lenses, by combining their signal. In fact, there are currently around a hundred strong lensing observations suitable for substructure inference, most of which come from the SLACS (Bolton et al. 2006) and BELLS (Brownstein et al. 2011) surveys. In the near future, new and future telescopes like *JWST* (Gardner et al. 2006), *ELT* (Simon et al. 2019), *Euclid* (Refregier et al. 2010; Laureijs et al. 2011), *SKA* (Koopmans, Browne & Jackson 2004), and *LSST* (Abolfathi et al. 2021) will deliver thousands of very high precision galaxy–galaxy lensing images (McKean et al. 2015). It is then extremely important to be

Table 1. Summary of model parameters used for the simulated images in this work. When a prior distribution is not specified, the parameter is fixed to the true value.

Parameter	True value	Prior	Description
Main lens			SPLE
$r_{\rm Ein}$ (arcsec)		$\mathcal{U}(1, 2)$	Einstein radius
$\xi_{0,x}$ (arcsec)		$\mathcal{U}(-0.2, 0.2)$	Lens centre x-axis
$\xi_{0,y}$ (arcsec)		$\mathcal{U}(-0.2, 0.2)$	Lens centre y-axis
q_l		$\mathcal{U}(0.1, 1)$	Axial ratio
ϕ_l (rad)		$\mathcal{U}(0, 2\pi)$	Rotation angle
γ	2.1	–	Slope
$z_{\rm lens}$	0.5	–	Lens redshift
External shear			
γ_1		$\mathcal{U}(-0.05, 0.05)$	First component
γ_2		$\mathcal{U}(-0.05, 0.05)$	Second component
Source			Sérsic
$I_{\rm e}$		$\mathcal{U}(0, 4)$	Surface intensity
$r_{\rm e}$ (arcsec)		$\mathcal{U}(0.1, 2.5)$	Effective radius
x_0 (arcsec)		$\mathcal{U}(-0.1, 0.1)$	Source centre x-axis
y_0 (arcsec)		$\mathcal{U}(-0.1, 0.1)$	Source centre y-axis
q_s		$\mathcal{U}(0.1, 1)$	Axial ratio
ϕ_s (rad)		$\mathcal{U}(0, 2\pi)$	Position angle
n		$\mathcal{U}(0.1, 4)$	Index
$z_{\rm src}$	2	–	Source redshift
Subhaloes			tNFW
$\boldsymbol{p}$ (arcsec)	$\in [-2.5, 2.5]$	$\mathcal{U}_{\rm 2D}(-2.5, 2.5)$	Position
m_{200} ($\rm M_\odot$)	$\in [10^7, 10^{10}]$	Giocoli et al. (2010)	Virial mass
c_{200}	15	–	Concentration
τ	6	–	Truncation
LOS haloes			Projected tNFW
$\boldsymbol{p}$ (arcsec)	$\in [-2.5, 2.5]$	$\mathcal{U}_{\rm 2D}(-2.5, 2.5)$	Position
m_{200} ($\rm M_\odot$)	$\in [10^7, 10^{10}]$	Tinker et al. (2008)	Virial mass
$z_{\rm LOS}$	$\in [0, z_{\rm src}]$	Tinker et al. (2008)	LOS redshift
c_{200}	15	–	Concentration
τ	6	–	Truncation
WDM			
$M_{\rm hm}$ ($\rm M_\odot$)		$\log \mathcal{U}(10^7, 10^{10})$	Half-mode mass

able to combine the information coming from different observations in the statistical analysis.

For the statistical analysis, we employ truncated marginal neural ratio estimation (TMNRE). Developed by Hermans, Begy & Louppe (2020) and Miller et al. (2020), marginal neural ratio estimation (MNRE) is a neural simulation-based inference method that makes it possible to learn the marginal posterior approximation for a specified subset of model parameters of interest directly from the full input data that, in our case, correspond to the observed lensed images, without the need for hand-crafted summary statistics. This method improves the simulator efficiency and the quality of inference. Moreover, MNRE is amortized, which enables important statistical consistency tests, which would have been extremely expensive with likelihood-based inference, like the expected coverage test we employ in Section 4.5. Up to now, this approach has been applied in simplified modelling frameworks: Hermans et al. (2020) focus on recovering the Einstein radius of a gravitational lens marginalizing over 15 source and lens mass distribution parameters, whereas Brehmer et al. (2019) estimate the slope and normalization of a CDM subhalo mass function. Simulation-based inference using neural posterior density estimator and hierarchical inference have also been employed in Wagner-Carena et al. (2022) to infer the CDM subhalo mass function normalization from a set of strong lensing images, generated using real galaxy images as a source model, including realistic observational noise effects from *Hubble Space Telescope* (*HST*) and accounting for the mean expected convergence from LOS haloes.

TMNRE is able to *target* the inference to a specific observation at hand rather than amortize over all possible parameter combinations, by successively focusing simulations on the parameter regions that are most relevant for the inference problem (Miller et al. 2022). This targeted approach is more efficient when most posterior density is concentrated compared to the prior density, which is the case for lens and source parameters. This truncation method applied to strong lensing images was proposed in Karchev, Coogan & Weniger (2021) and used in Coogan, Karchev & Weniger (2020) and Coogan et al. (2022) to learn marginal posterior approximations for individual subhalo parameters, marginalizing over lens and source uncertainties given an observation. It has also been recently applied to analysis of the CMB (Cole et al. 2022).

The main goal of this work is to demonstrate that our TMNRE approach is sensitive to the HMF half-mode mass $M_{\rm hm}$ given a set of *HST* resolution observations, and it is able to efficiently and accurately infer its statistic.

The paper is structured as follows. In Section 2, we describe how we model strong lens observations with analytic source, lens, and substructure population that accounts for both subhaloes and LOS haloes. In Section 3, we discuss the inference methodology employed in the statistical analysis: TMNRE. Finally, we show our results in Section 4 and conclude in Section 6. This work paves the way for combining the presented statistical analysis with more realistic strong lensing source models and for future applications to real high-resolution data in upcoming works.

2 STRONG LENSING MODEL

In this section, we review how we model strong lensing images. In strong lensing systems, the mass distribution of a foreground galaxy gravitationally lenses the light rays coming from a background source, resulting in an arc-like image in the case of an extended galaxy source. Under the assumptions of the thin-lens formalism (Meneghetti 2016), the lens-plane and source-plane coordinates of a light ray, respectively, $\boldsymbol{\xi}$ and $\boldsymbol{x}$ are related by the simple ray-tracing equation:

$$\boldsymbol{x} = \boldsymbol{\xi} - \boldsymbol{\alpha}(\boldsymbol{\xi}). \tag{1}$$

The displacement field $\boldsymbol{\alpha}$ can be computed as

$$\boldsymbol{\alpha}(\boldsymbol{\xi}) = \frac{4G}{c^2}\frac{D_{ls}}{D_l D_s}\int \mathrm{d}^2(D_l\boldsymbol{\xi}')\frac{\boldsymbol{\xi}-\boldsymbol{\xi}'}{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2}\Sigma(\boldsymbol{\xi}'), \tag{2}$$

where we have introduced the angular diameter distance from the observer to the lens D_l, from the observer to the source D_s, and from the lens to the source D_{ls}. The projected mass density is given by the integral of the 3D lensing mass density ρ:

$$\Sigma(\boldsymbol{\xi}) = \int \mathrm{d}z\,\rho(\boldsymbol{\xi}, z), \tag{3}$$

where z is the coordinate perpendicular to the lens plane. It is also useful to define the convergence κ in terms of the critical surface density $\Sigma_{{\rm cr},l}$ on the lens plane as

$$\kappa(\boldsymbol{\xi}) = \frac{\Sigma(\boldsymbol{\xi})}{\Sigma_{{\rm cr},l}}, \quad \Sigma_{{\rm cr},l} \equiv \frac{c^2}{4\pi G}\frac{D_s}{D_l D_{ls}}, \tag{4}$$

where c is the speed of light and G is the gravitational constant. It can be shown that the convergence κ is related to the trace of the

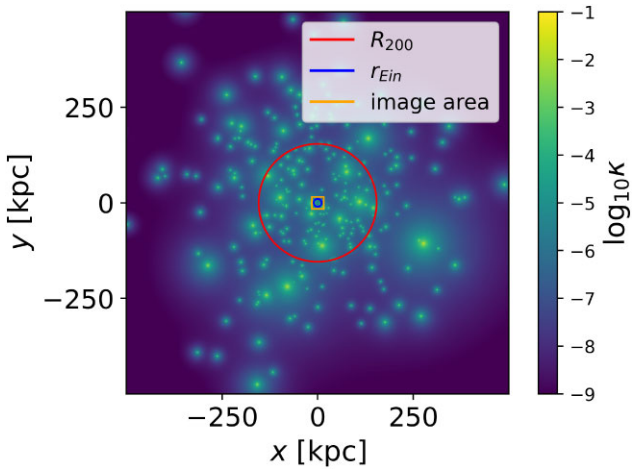

Figure 1. Convergence map for a CDM subhalo population in the adopted mass range. The convergence map shows how the deflecting mass from all the subhalo lenses is distributed. The full map size is 1 Mpc × 1 Mpc. We mark in red the virial radius of the main lens halo, in blue its Einstein radius, and in orange the 5 arcsec × 5 arcsec lensing image area.

Jacobian of the lensing transformation, and it represents the lens mass distribution.

Strong lensing systems are then represented by two main ingredients: the lens model, which describes the total mass distribution of the lens, and the source model, which describes the surface brightness profile of the background source. It is common to split the lens model into a macroscopic smooth component (main lens and external shear) and a substructure[1] component, due to subhaloes and LOS haloes. Each ingredient can be directly superimposed by summing their respective displacement fields in the lens plane:

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\rm lens} + \boldsymbol{\alpha}_{\rm ext} + \sum_{i=1}^{N_{\rm sub}} \boldsymbol{\alpha}_{{\rm sub},i} + \sum_{i=1}^{N_{\rm los}} \boldsymbol{\alpha}_{{\rm los},i}. \tag{5}$$

In the following subsections, we will describe each component of the model, which we summarize in Table 1.

2.1 Main lens model

We model the lens mass distribution smooth component with a singular power-law ellipsoid (SPLE) lens (Suyu et al. 2009) plus external shear. The latter accounts for matter in the lens surroundings and describes large-scale effects constant across the image. For a detailed description of these two ingredients, we refer the reader to Karchev et al. (2021). We end up with eight parameters in total that we collect in the vector $\boldsymbol{\theta}_l \equiv \{r_{\rm Ein}, \xi_{0,x}, \xi_{0,y}, q_l, \phi_l, \gamma, \gamma_1, \gamma_2\}$, the first six from the SPLE for the main-lens mass distribution and the last two for external shear.

When simulating data (see Section 4.1), we always use the same SPLE slope that produced the mock observation[2] we are analysing (as fixed in Table 1) instead of inferring it for simplicity.[3]

[1]Throughout our work, we use the terms 'small-scale structures', 'substructures', and 'low-mass haloes' when considering both subhaloes of the main lens and LOS haloes.

[2]Throughout our work, we use the terms 'simulated data' for data used during inference and 'mock observations' for the simulated data that we analyse.

[3]In principle, inferring the slope is possible, but it requires more training data and leads to increased uncertainties in both lens and source parameters.

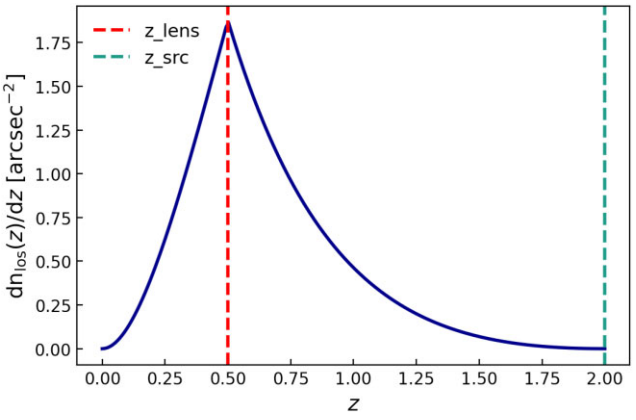

Figure 2. LOS halo distribution in redshift for our source and lens redshifts configuration, described in Table 1.

2.2 Source model

To model the surface brightness of the source galaxy, we adopt a Sérsic profile (Sérsic 1963). The surface brightness distribution is given by

$$\beta(x, y) = I_{\rm e} \exp -k_n \left[\left(\frac{r(x, y)}{r_{\rm e}} \right)^{1/n} - 1 \right], \tag{6}$$

where $I_{\rm e}$ is the surface intensity at the half-light radius $r_{\rm e}$, $r(x, y) = \sqrt{r_x^2 + r_y^2}$ is the elliptical radial coordinate, and the normalization k_n depends on the index n. We give more details about the Sérsic profile modelling and parameters in Appendix A. In total, the analytic source is parametrized with seven variables that we collect in the vector $\boldsymbol{\theta}_s \equiv \{I_{\rm e}, r_{\rm e}, x_0, y_0, q_s, \phi_s, n\}$.

2.3 Small-scale structures model

Substructures can be divided into two categories: subhaloes that orbit around the main halo at the lens redshift, and LOS haloes distributed between the source and the observer. LOS haloes are a more direct probe of free-streaming-induced small-scale structure suppression, because they are less affected by baryonic processes and environmental effects, such tidal stripping interactions with the main halo (Despali et al. 2018). For this reason and the fact that they are expected to be more abundant than subhaloes in a lensing system (Despali et al. 2018; He et al. 2022b), it is very important to model them as well, in order to correctly estimate the collective effects of all substructures on the lensing image.

2.3.1 Density profile

To model the density profiles of small-scale DM haloes, we adopt the smoothly truncated universal 3D mass density profile from Baltz, Marshall & Oguri (2009):

$$\rho_{\rm tNFW}(r) = \frac{\rho_s}{r/r_s(1 + r/r_s)^2} \frac{1}{1 + (r/r_t)^2}. \tag{7}$$

Here, r is the three-dimensional distance from the centre of the halo, ρ_s and r_s are, respectively, the scale density and scale radius that specify an Navarro–Frenk–White (NFW) profile (Navarro, Frenk & White 1997), and $r_{\rm t} \equiv \tau r_s$ is the tidal truncation radius that depends on the history of the subhalo. Typical values of the truncation scale τ

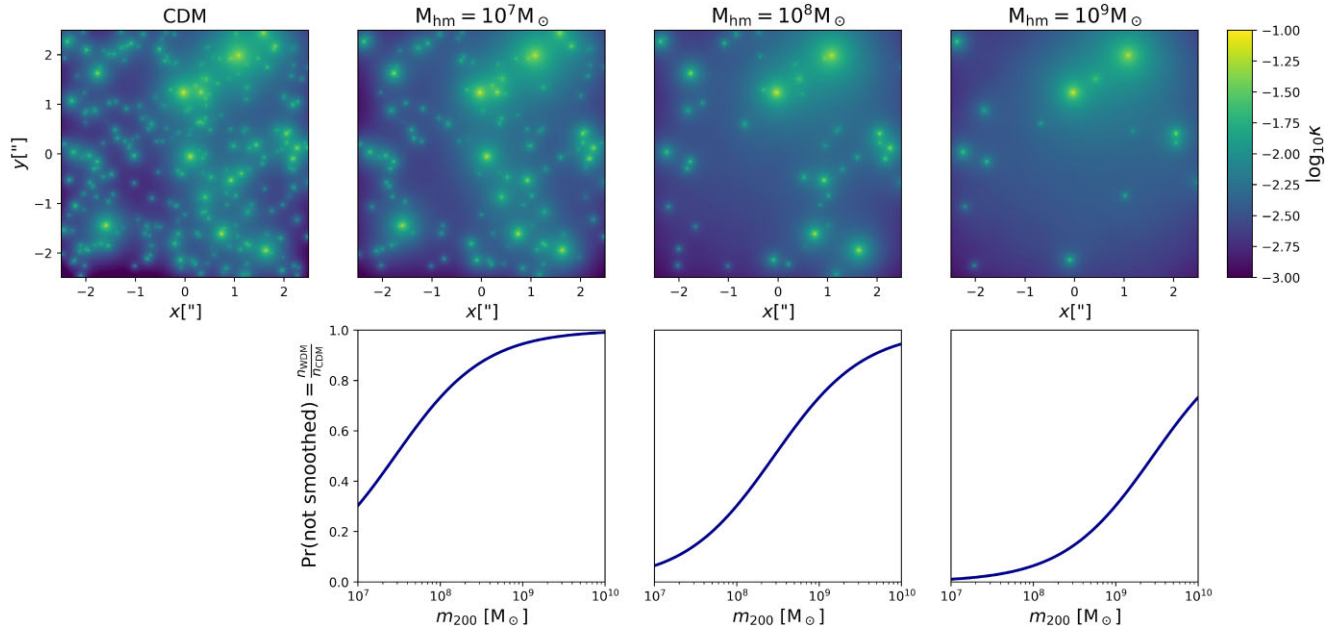

Figure 3. *Top*: Convergence maps for a population of LOS haloes with masses sampled from the CDM HMF in the adopted mass range (Table 1) and projected in the lens plane following the prescription from Şengül et al. (2020). In the second to fourth columns, we imitate the effect of WDM with different cut-off masses (as labelled in the titles) via our smoothing scheme. *Bottom*: We show the probability with which LOS haloes do not get smoothed, equal to the ratio between the WDM and the CDM HMF (equation 14). The smoothing is stochastic, so for each realization of the smoothing different haloes are smoothed.

range from 4 to 10 for spherically symmetric lenses (Cyr-Racine et al. 2019; Gilman et al. 2019b); we fix $\tau = 6$ for simplicity. Compared to the standard NFW form, which has an infinite total mass, the truncated NFW (tNFW) contains an additional truncation term that makes the profile decay as r^{-5} for large radii, resulting in a finite total mass given by

$$m_\tau = 4\pi\rho_s r_s^3 \frac{\tau^2}{(\tau^2+1)^2}[(\tau^2-1)\ln\tau + \tau\pi - (\tau^2+1)]. \quad (8)$$

With a fixed truncation scale, the tNFW profile is fully determined by the same parameters that determine the NFW profile: the virial mass m_{200}[4] and the concentration $c_{200} = r_{200}/r_s$ of the halo. The latter measures how concentrated the mass of a halo is and fixes the density normalization; in principle, it varies from one subhalo to the next and shows dependences on mass and redshift of the main halo. In this paper, instead of adopting a concentration–mass relation, we fix $c_{200} = 15$ in accordance with Richings et al. (2021). We would like to note that accounting for the scatter in the mass–concentration relation might boost the expected lensing signal from a low-mass halo (Amorisco et al. 2021).

LOS haloes are also modelled with a tNFW profile following the prescription by Şengül et al. (2020), which shows how to treat haloes along the LOS as effective subhaloes on the main-lens plane, with a modified scale radius and mass. We give more details on this procedure in Appendix B. For LOS haloes, we adopt the same concentration and truncation scale values used for subhaloes.

The equations for calculating the displacement field of a tNFW halo, given its mass and position, are fully elaborated by Baltz et al. (2009, appendix A).

[4]We parametrize subhaloes by what would be their mass up to the virial radius r_{200} using the untruncated profile, with the same central density ρ_s and scale radius r_s as the truncated one.

2.3.2 *Mass and spatial distributions*

We sample subhalo masses from the CDM mass function of Giocoli et al. (2010):

$$\frac{1}{M}\frac{\mathrm{d}n_{\rm sub}(m_{200}, z)}{\mathrm{d}\log m_{200}} \propto (1+z)^{1/2} m_{200}^{\alpha} \exp\left[-\beta\left(\frac{m_{200}}{M}\right)^3\right], \quad (9)$$

where M is the main halo's mass and m_{200} the subhalo mass.[5] We use the normalization, slope, and exponential cut-off of the subhalo mass function from Despali & Vegetti (2017). The expected number of subhaloes in a given mass interval for the lens halo system can be computed by integrating the mass function over that interval.

For LOS halo masses, we use the Tinker et al. (2008) CDM HMF assuming an overdensity with respect to the critical density of the Universe at the epoch of analysis of $\Delta = 200$. For both subhaloes and LOS haloes, we adopt the following mass range, with $m_{200,\,\rm min} = 10^7\,M_\odot$ and $m_{200,\,\rm max} = 10^{10}\,M_\odot$. The upper limit is chosen based on the assumption that more massive haloes would be visible and could therefore be modelled independently. The lower one is fiducial, and we plan on investigating more the sensitivity of our inference in the future.

The spatial distribution of subhaloes has been shown to follow an Einasto profile (Springel et al. 2008). However, since the virial radius of a typical main lens halo is much larger than its Einstein radius, and

[5]The total mass of the lens galaxy is described by the Einstein radius of the system, a very well-constrained parameter in lensing inference analyses. For the purpose of describing subhaloes, we need to be able to map the measured properties of the lens (the Einstein radius $r_{\rm Ein}$) on to the properties of the host halo (the mass M). For simplicity, we compute the mass of the host halo transforming the Einstein radius distance measure into a mass measure. We would like to point out a similar approach from Brehmer et al. (2019), where they relate the central velocity dispersion of a singular isothermal ellipsoid lens mass distribution profile to the virial mass of the host halo.

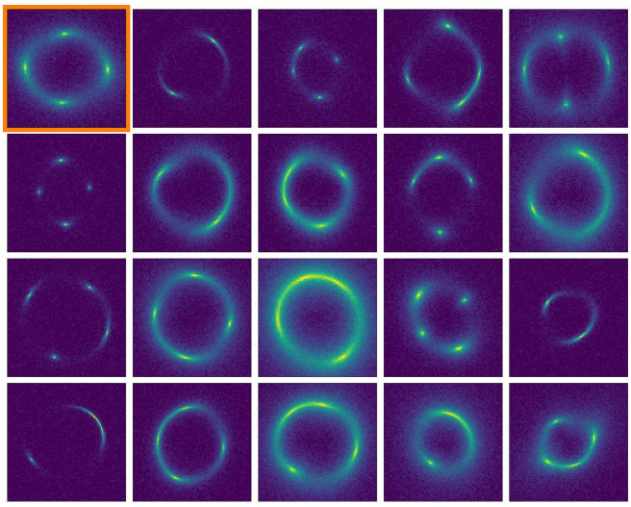

Figure 4. We present a gallery of twenty mock strong lensing images we use as target observations. These mock observations have been generated with arbitrary lens and source parameters drawn from the initial prior in Table 1. Their peak signal-to-noise (SNR) is ∼30, representative of *HST* data. We analyse these images by first constraining their lens and source parameter proposal distribution in Section 4.3. Then, we combine them in order to infer the cut-off mass scale in Section 4.4. For the first one (upper left corner, framed in orange) of these images, we show our results of the first part of the pipeline (Section 4.3) in Figs 6–8.

hence, than the image plane, we approximate the distribution to be uniform in the lensing image area. Still, we derive the total number of expected subhaloes within the image via the Einasto fit of Despali & Vegetti (2017). We find that on average $\bar{n}_{\rm sub} = 4$ subhaloes fall within the lensing image area in our adopted lensing configuration and mass range. When generating a simulated image, we draw the number of subhaloes from Poisson($\bar{n}_{\rm sub}$), we then sample their masses from the subhalo mass function in equation (9) and sample their projected positions uniformly over the lensing image area. In Fig. 1, we show the convergence map for one realization of our subhalo population.

LOS haloes are rendered in a double-cone geometry with the lensing image area as an opening angle, and closing angle such that the cone closes at the source redshift, as described in Şengül et al. (2020, fig. 3). We infer the number of detectable LOS haloes by integrating their mass function in the mass range adopted for the analysis and within the double-cone volume

$$\bar{n}_{\rm los} = \int_0^{z_{\rm src}} \int_{m_{200,\rm min}}^{m_{200,\rm max}} n_{\rm los}(m_{200}, z) {\rm d}m_{200} \frac{{\rm d}V}{{\rm d}z} {\rm d}z. \tag{10}$$

On average, we get $\bar{n}_{\rm los} = 260$ LOS haloes projected in our lens plane. Similarly to what we do with the subhalo population, when generating simulated images, we draw the number of LOS haloes from Poisson($\bar{n}_{\rm los}$), we then sample their masses and redshift from the Tinker et al. (2008) HMF and sample their projected positions uniformly over the lensing image area. In Fig. 2, we show the distribution of LOS haloes in redshift for our lens and source redshifts configuration.

Finally, we label the vector of all substructure parameters with $\boldsymbol{\theta}_h \equiv \{\boldsymbol{m}_{200,\rm sub}, \boldsymbol{p}_{\rm sub}, \boldsymbol{m}_{200,\rm los}, \boldsymbol{p}_{\rm los}, z_{\rm los}\}$, where we use bold letters to denote arrays (e.g. $\boldsymbol{m}_{200,\rm sub}$ is an ordered set of masses, one for each simulated subhalo) and bold letters with an arrow to indicate arrays of vectors (e.g. $\boldsymbol{p}_{\rm sub}$ is an ordered set of positions in the lens plane, one for each simulated subhalo).

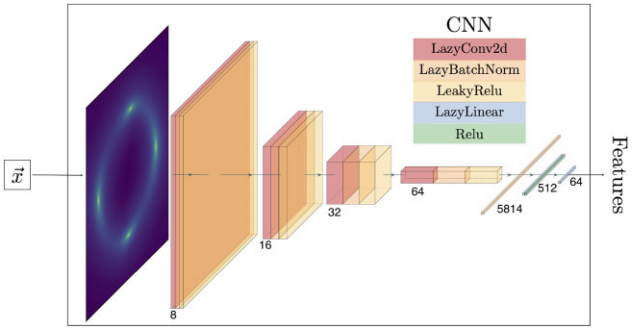

Figure 5. Illustration of the embedding CNN architecture used in the first part of the pipeline to constrain lens and source parameters. The observation $\boldsymbol{x}$ gets compressed into features: estimates of the best possible data summary statistic, by the CNN. In describing the CNN layers we follow `PyTorch` (Paszke et al. 2019) convention. To create the illustration, we have used Iqbal (2018).

2.4 Modelling free-streaming effects in WDM

The free-streaming effects of WDM are well described in terms of the half-mode wavelength $\lambda_{\rm hm}$, which corresponds to the scale at which the DM transfer function falls to half the CDM transfer function. We can define the half-mode mass as the mass contained within a radius of the half-mode wavelength:

$$M_{\rm hm} = \frac{4\pi\,\Omega_{\rm m}\rho_{\rm crit}}{3}\left(\frac{\lambda_{\rm hm}}{2}\right)^3, \tag{11}$$

where $\Omega_{\rm m}$ is the matter density parameter and $\rho_{\rm crit}$ is the critical density of the Universe. Following Schneider et al. (2012), the half-mode wavelength,

$$\lambda_{\rm hm} = 2\pi\alpha_{\rm hm}\left(2^{\nu/5} - 1\right)^{-1/(2\nu)}, \tag{12}$$

is the scale below which the initial density perturbations are completely erased, with $\nu = 1.12$ and, assuming that all DM is warm,

$$\alpha_{\rm hm} = 0.049\left(\frac{m_{\rm WDM}}{\rm keV}\right)^{-1.11}\left(\frac{\Omega_{\rm DM}}{0.025}\right)^{0.11}\left(\frac{h}{0.7}\right)^{1.22} h^{-1}\,{\rm Mpc}. \tag{13}$$

We then have a one-to-one mapping between the mass of the WDM particle and the half-mode mass. For strong lensing, the half-mode mass can be thought of as an effective cut-off mass below which the DM mass function is strongly suppressed. To model this suppression in the WDM mass function, we adopt for both subhaloes and LOS haloes the functional form from Lovell (2020):

$$\frac{n_{\rm WDM}}{n_{\rm CDM}} = \left(1 + \left(\alpha\frac{\rm M_{hm}}{m_{200}}\right)^{\beta}\right)^{\gamma}, \tag{14}$$

with best-fitting parameters $\alpha = 4.2$, $\beta = 2.5$, and $\gamma = 0.2$ for subhaloes, $\alpha = 2.3$, $\beta = 0.8$, and $\gamma = 1$ for central haloes.

2.4.1 Smoothing substructures

The observational signature of WDM is, thus, the absence of small-scale structures. However, in the current parametrization, this is accompanied by the removal of the corresponding mass enclosed in them, whereas in reality the mass will still be present but will be diffused throughout the smooth main halo. This effect is manifested in a correlation between the half-mode mass and the main-halo Einstein radius: suppressing more substructure leads to an increase

in the inferred Einstein radius since the total mass of the system (within the image) is tightly constrained by the size of the observed ring (or arcs).

We introduce a prescription for dealing with this degeneracy, which well captures the physical reality of structure suppression due to free streaming. Haloes that should be suppressed are not present because the DM particles that should make them up are freely streaming, and their mass is therefore more diluted throughout the main halo. Therefore, rather than removing or adding substructures as a response to a changing cut-off, we still sample substructures from the CDM mass function, but we smooth the displacement field generated by haloes that should be suppressed based on the aforementioned prescription by Lovell (2020) to hide their lensing signature. In other words, each sampled small-scale halo has a probability equal to the ratio between the WDM and the CDM HMF (equation 14) of not being smoothed.

We then effect the smoothing by convolving the deflection field of each individual sub-/LOS halo with a radially symmetric filter

$$f \propto 1 - \exp\left(-\left(\frac{r}{r_{\rm smooth}}\right)^{n_{\rm smooth}}\right). \tag{15}$$

This filtering preserves the far-field lensing signature of the halo, which is only determined by its total mass. By default, we choose the smoothing scale to be equal to the halo virial radius: $r_{\rm smooth} = r_{200}$, and the smoothing exponent $n_{\rm smooth} = 2$.

In the top row of Fig. 3, we visualize the convergence maps in the lens plane for the same realization of LOS haloes drawn from CDM distributions (panel 1), and with different cut-off masses implemented with our smoothing scheme (panels 2–4). In the bottom row, we show how we decide to smooth the lensing signature of certain haloes based on the ratio between the WDM and the CDM HMF (equation 14).

3 STATISTICAL ANALYSIS

Constraining the fundamental properties of DM by characterizing the population of DM haloes in a strong lensing image is an extremely difficult problem since the signal we are interested in has a sub-per cent level influence on images dominated by statistical noise. The problem is further complicated by the large differences between images of different lensing systems.

Our ultimate goal is to compute the marginal posterior $p(\vartheta|\boldsymbol{x})$ for a single parameter of interest $\vartheta = M_{\rm hm}$, the half-mode mass, given an observation $\boldsymbol{x}$, for which we have the generative model

$$p(\boldsymbol{x}, \boldsymbol{\theta}_s, \boldsymbol{\theta}_l, \boldsymbol{\theta}_h, \vartheta) = p(\boldsymbol{x}|\boldsymbol{\theta}_s, \boldsymbol{\theta}_l, \boldsymbol{\theta}_h, \vartheta)p(\boldsymbol{\theta}_s)p(\boldsymbol{\theta}_h|\boldsymbol{\theta}_l, \vartheta)p(\boldsymbol{\theta}_l)p(\vartheta). \tag{16}$$

The first factor on the right-hand side is the simulator, while the other factors denote the priors on the various source, lens, and DM substructure parameters as listed in Table 1.

Therefore, in order to derive $p(\vartheta|\boldsymbol{x})$, we need to marginalize over all the nuisance parameters $\boldsymbol{\eta} \equiv \{\boldsymbol{\theta}_l, \boldsymbol{\theta}_s, \boldsymbol{\theta}_h\}$:

$$p(\vartheta|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\vartheta)}{p(\boldsymbol{x})}p(\vartheta) = \frac{\int {\rm d}\boldsymbol{\eta}\, p(\boldsymbol{\eta})p(\boldsymbol{x}|\vartheta, \boldsymbol{\eta})}{p(\boldsymbol{x})}p(\vartheta). \tag{17}$$

This is a very high-dimensional and multimodal integral, even for simple analytical lens and source models, due to the large population of interchangeable substructures. Therefore, it is intractable, which renders likelihood-based inference infeasible in this case.

Instead, we approximate $p(\vartheta|\boldsymbol{x})$ using simulation-based inference with amortized approximate ratio estimators (Hermans et al. 2020). In particular, we employ the TMNRE algorithm developed by Miller et al. (2020, 2022) and implemented in the package SWYFT.[6]

3.1 Marginal neural ratio estimation

MNRE (Miller et al. 2020) sets up a classification problem which produces an estimate $\hat{r}(\boldsymbol{x}, \vartheta)$ of the marginal likelihood-to-evidence ratio:

$$r(\boldsymbol{x}, \vartheta) \equiv \frac{p(\vartheta|\boldsymbol{x})}{p(\vartheta)} = \frac{p(\boldsymbol{x}|\vartheta)}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}, \vartheta)}{p(\boldsymbol{x})p(\vartheta)}. \tag{18}$$

Given the prior $p(\vartheta)$, the ratio estimator $\hat{r}(\boldsymbol{x}, \vartheta)$ may then be used as a surrogate model to draw samples from an approximate posterior $\hat{p}(\vartheta|\boldsymbol{x}) = \hat{r}(\boldsymbol{x}, \vartheta)p(\vartheta)$. In order to estimate the ratio, the strategy is to train a neural network $d_\phi(\boldsymbol{x}, \vartheta)$, where ϕ are the network weights, via stochastic gradient descent. The network is parametrized as a binary classifier to discriminate between two hypotheses labelled by the binary variable C. In the first one, with class label $C = 1$, the observation $\boldsymbol{x}$ and the parameter of interest ϑ are drawn jointly from the parameter prior and model: $\boldsymbol{x}, \vartheta \sim p(\boldsymbol{x}, \vartheta)$. In the second one, with class label $C = 0$, they are sampled marginally: $\boldsymbol{x}, \vartheta \sim p(\boldsymbol{x})p(\vartheta)$. We sample from the two classes with equal probability, enforcing the outcome of the binary variable C to be random. To train the network, we use the binary-cross entropy loss function:

$$\ell[d_\phi(\boldsymbol{x}, \vartheta)] = -\int {\rm d}\boldsymbol{x}{\rm d}\vartheta \left\{p(\boldsymbol{x}, \vartheta)\log d_\phi(\boldsymbol{x}, \vartheta) + p(\boldsymbol{x})p(\vartheta)\log\left[1 - d_\phi(\boldsymbol{x}, \vartheta)\right]\right\}. \tag{19}$$

The loss functional defined in equation (19) is minimized when the output of the neural network $d_\phi(\boldsymbol{x}, \vartheta)$ corresponds to the probability of the class with label $C = 1$:

$$d_\phi(\boldsymbol{x}, \vartheta) = p(C = 1|\boldsymbol{x}, \vartheta) = \frac{p(\boldsymbol{x}, \vartheta)}{p(\boldsymbol{x}, \vartheta) + p(\boldsymbol{x})p(\vartheta)} \equiv \sigma[\log \hat{r}(\boldsymbol{x}, \vartheta)], \tag{20}$$

so we can express the ratio estimator $\hat{r}(\boldsymbol{x}, \vartheta)$ in terms of the binary classifier $d_\phi(\boldsymbol{x}, \vartheta)$ using the sigmoid function $\sigma(y) \equiv 1/(1 + e^{-y})$.

Marginalization over nuisance variables $\boldsymbol{\eta}$ is done implicitly since the data will incorporate the variance from the nuisance parameters, but the inference procedure estimates only the marginal likelihood-to-evidence ratio. In other words, parameters to be marginalized over are sampled during training data generation, but not shown to the binary classifier $d_\phi(\boldsymbol{x}, \vartheta)$. As a result, the trained network effectively learns an estimate of the marginal likelihood-to-evidence ratios $\hat{r}(\boldsymbol{x}, \vartheta)$, which we can use to evaluate the marginal posterior for the parameter of interest directly (if the prior probability density function (PDF) is known) or obtain samples otherwise.

3.2 Truncated marginal neural ratio estimation

Formally, inferring the marginal posterior for the substructure population parameter of interest would require marginalizing over all the source, lens, and substructure realizations compatible with all possible strong lensing images. However, sampling lens and source parameters from their priors would require a very large amount of training data and a more complex network architecture when using neural ratio estimation. This has been attempted only in Brehmer et al. (2019) to infer the slope and normalization of the HMF. In order to reduce the complexity of the problem and fully exploit available

[6] https://github.com/undark-lab/swyft

information in the data with limited computational resources, we propose to target one image at a time, focusing simulations and the network training on a specific observation of interest. Thanks to SWYFT, we implement this with a truncation scheme.

TMNRE generates a sequence of likelihood-to-evidence ratio estimators on both nuisance and parameters of interest for a specific observation $\boldsymbol{x}$. In multiple inference rounds, the proposal distribution for nuisance parameters is updated and constrained, based on these ratio estimators, in order for the training data to match each round more closely the observation of interest $\boldsymbol{x}$ (this can be visually appreciated in Fig. 7, which will be discussed in more details in Section 4.3).

The procedure for the truncation scheme is the following. In the first inference round, we generate training data sampling the nuisance parameters from the initial prior $p(\boldsymbol{\eta})$. Then, in each round, we constrain the proposal distribution $p_\Gamma(\boldsymbol{\eta})$ for the parameters we want to marginalize over to a region Γ where the nuisance parameters are more likely to have generated $\boldsymbol{x}$ based on the ratio estimator trained in that round. In particular, we estimate the new region Γ by very conservatively truncating the previous proposal distribution $p_\Gamma(\boldsymbol{\eta})$ in the region where the ratio estimator exceeds a predetermined threshold. We set the threshold hyperparameter to $\epsilon = 10^{-5}$, which, in case of a Gaussian posterior, corresponds to truncating at $\sim 4.78\sigma$ (Miller et al. 2022). We obtain the final proposal distribution for our nuisance parameters when the region Γ does not change significantly anymore between rounds.

In this work, we target with TMNRE a restricted set of the nuisance parameters $\boldsymbol{\eta}$: namely, those of the analytic smooth lens and source models, $\boldsymbol{\theta}_l$ and $\boldsymbol{\theta}_s$, while leaving halo parameters, $\boldsymbol{\theta}_h$ unconstrained.

To summarize, thanks to TMNRE, the overall analysis strategy splits into the following steps:

(1) Train an inference network on an image $\boldsymbol{x}$ to constrain the source and lens parameters, $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_l$, within ranges consistent with the observation. We then generate targeted training data based on this constrained model.

(2) Train an inference network to learn the marginal likelihood-to-evidence ratio for our parameter of interest, the half-mode mass $M_{\rm hm}$, on the targeted training data.

Similarly to the reasoning behind the approximate Bayesian computation rejection algorithm, which discards sampled parameters values if the generated data are too different from the observed data, we justify this approach by noting that parameters that do not produce observations similar to $\boldsymbol{x}$ will not contribute to the integral in equation (17). Restricting the input parameters in this way immensely reduces the variability of simulated data, which allows us to use simpler network architectures and fewer training examples in the next step. As a result, the inference is now *targeted* to the specific observation at hand rather than amortized over all the possible lens/source combinations from the full prior. We would like to point out that the inference is still locally amortized in the constrained proposal distribution region, and this enables empirical test of the inference result (see Section 4.5).

4 RESULTS

In this section, we show our results. First, we describe the simulated data in Section 4.1 and the inference network architectures in Section 4.2. We then show how we constrain the lens and source parameters in Section 4.3. Next, we show our results for the cut-off mass and describe how we can combine the information from different strong lensing images in Section 4.4. In the same subsection, we show our results on the DM mass. Finally, we directly assess the statistical behaviour of the trained neural networks in Section 4.5.

4.1 Mock data generation

We want our simulations to be representative of *HST* data, in order to demonstrate that our pipeline is in principle able to extract the signal of interest from them. We adopt a pixel scale of 0.05 arcsec, being slightly larger than the expected 0.04 arcsec, which allows us to disregard its point spread function (PSF; Gennaro 2018) for simplicity. The size of the images is 100×100 pixels, so they cover an area of 5 arcsec $\times$ 5 arcsec on the sky. Initially, we generate the mock data with a resolution 10-times higher and then downsample it to the adopted resolution by local averaging, effectively simulating integration of the light across the pixel areas.

We model the instrumental effects by simply assuming a Gaussian and uncorrelated pixel noise. The noise level σ is set so that the peak SNR ratio of the image is $\sim$30 (after downsampling), representative of *HST* data. Then, given a modelled flux, our simulator is given by

$$p(\boldsymbol{x}|\boldsymbol{\theta}_s, \boldsymbol{\theta}_l, \boldsymbol{\theta}_h, \vartheta) = \mathcal{N}(\boldsymbol{x}|\mathrm{obs}(\boldsymbol{\theta}_s, \boldsymbol{\theta}_l, \boldsymbol{\theta}_h, \vartheta), \sigma^2). \tag{21}$$

We leave to future works to account for the correct modelling of the PSF and correlated pixel noise, which are fundamental in order to correctly conduct substructure studies in strong gravitational lensing images.

In Fig. 4, we show a gallery of 20 mock strong lensing images we use as target observations. These mock observations have been generated with arbitrary lens and source parameters drawn from the initial prior in Table 1. Their peak SNR is $\sim$30, representative of *HST* data.

4.2 Inference network architecture

The inference neural network used to perform TMNRE is split into two different components: an embedding network $C_\phi(\boldsymbol{x})$ and a binary classification network. The embedding network compresses data into a low-dimensional feature vector, estimating the best possible summary statistics from the full input image. The binary classification network is the marginal classifier that performs the actual ratio estimation. It passes the featurized observational data concatenated with the parameter of interest into a multilayers perceptron (MLP) to estimate the marginal likelihood-to-evidence ratios. The network architecture can be expressed as

$$d_\phi(\boldsymbol{x}, \vartheta) = \mathrm{MLP}_\phi(\mathrm{features} = C_\phi(\boldsymbol{x}), \vartheta) = \sigma[\log \hat{r}(\boldsymbol{x}, \vartheta)]. \tag{22}$$

For the embedding network, in both steps of the pipeline, we adopt a simple convolutional neural network (CNN). In Fig. 5, we show the CNN architecture used to constrain lens and source parameters. The one used to estimate the cut-off mass has a similar structure.

4.3 Constraining lens and source parameters

We constrain lens and source parameter regions with TMNRE, as described in Section 3.2, with multiple sampling and training rounds.

In total, we perform six sampling and training rounds. In each round, we simulate 10^5 observations, of which 90 per cent are used as the training data set, and the remaining 10 per cent as the validation data set. Evaluations of the network on the mock target image are used to truncate the training data proposal distribution after each round, so that the region for lens and source parameters is targeted. The first training round is performed on the data set generated from the initial source and lens parameters priors, shown in Table 1. In Fig. 6,

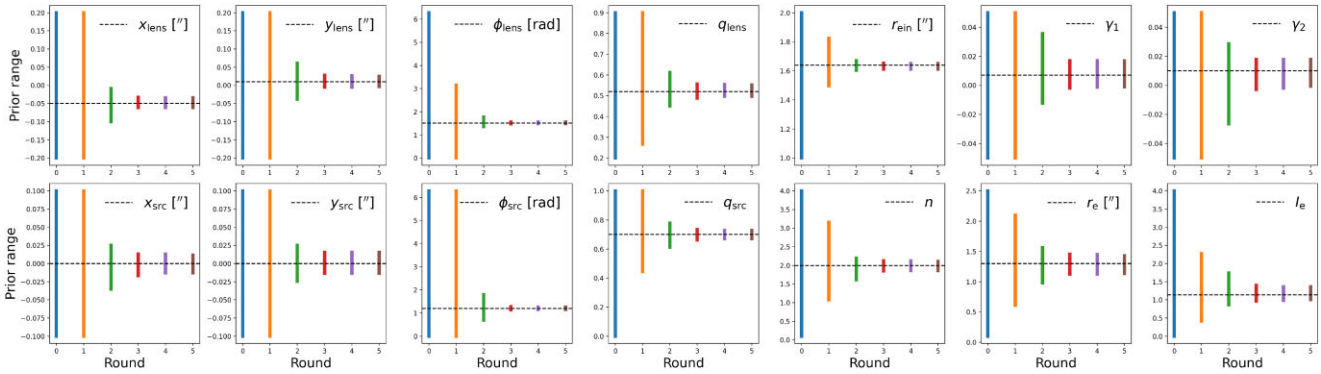

Figure 6. Constrained proposal distribution. Visualization of the sequential truncation of the lens and source proposal distributions over the six rounds of training. The particular target is the first mock image (framed in orange in Fig. 4), whose parameters are depicted as black dashed horizontal lines.

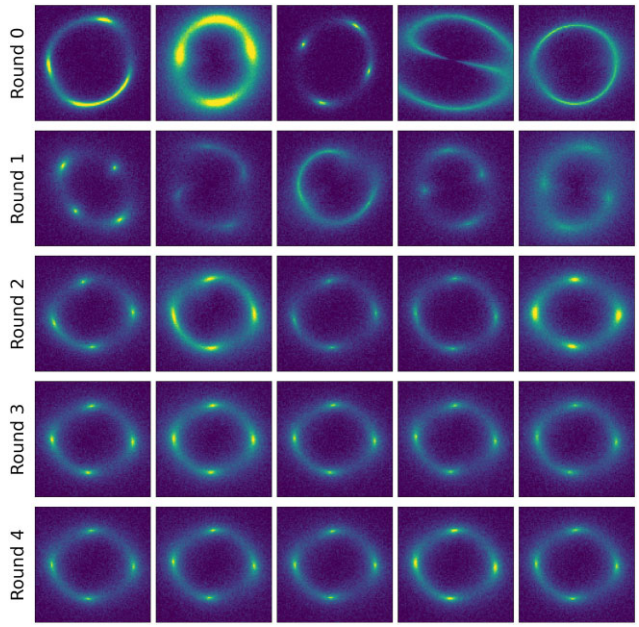

Figure 7. Training data targeting the first mock observation (framed in orange in Fig. 4). In each row, we show five examples of training data for the first five rounds. In the first round, we sample our data from the initial prior shown in Table 1. For the following rounds, the lens and source parameters are sampled from the constrained proposal distributions, obtained by evaluating the network trained with the previous round data set on our target observation (see Section 4.3). It is evident that with each round the training data more closely resembles the target image $\boldsymbol{x}$.

we show the initial prior and the following constrained proposal distributions. It can be seen that after the first round just a few of the parameters proposal distributions get truncated, e.g. the Einstein radius. By having truncated these initial parameters, in the following rounds the other parameters can be better learned by the network and so constrained. In Fig. 7, we show samples from the first five training data sets, which demonstrate that the constrained regions are indeed the ones that are likely to produce data similar to the targeted image $\boldsymbol{x}$. After the sixth round of training, it is not possible anymore to truncate the proposal distribution region based on the predetermined threshold, as seen in Fig. 8. The truncation scheme has then efficiently identified the constrained region for lens and source parameters consistent with the targeted observation.

Using the last constrained data set, it is then easier in the second step of the pipeline to train a marginal neural ratio estimator to perform the final inference on the cut-off mass, as explained in Section 3.2.

We would like to stress that these constrained proposal distributions correctly account for lens and source parameter uncertainties. In all our simulated data, the substructure parameters $\boldsymbol{\theta}_h$ are randomly sampled from their prior, in order to account for the presence of substructure. This has the desirable outcome of approximately accounting for the average effect that an additional mass component has on the main lens parameters (e.g. inferring an unbiased Einstein radius) and contributes to the source and lens uncertainties.

4.4 Dark matter inference

For the second step of the pipeline, we train an inference network to learn the cut-off mass on the last constrained data set.

From initial tests, we have found that features from a single image are very hard to learn for the classifier, resulting in a very noisy ratio estimator. In order to reduce the estimator uncertainty, we then train the cut-off mass classifier on a data set $X^N = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ of N different observations. For each observation, first, we constrain its lens and source parameters as explained in Section 4.3. Then, we train the cut-off classifier on the concatenation of the features coming from their embedding networks, effectively learning $r(X^N, \vartheta)$. Note that the images in one data set are sampled with the same cut-off mass $M_{\rm hm}$, but different lens, source, and substructures realizations. In fact, our final goal is to apply the full pipeline to real data, which will all have different source, lens, and substructures configurations, but will have encoded the same DM properties.

In the first row of Fig. 9, we show the results from the inference network on 10 test sets of lenses generated with a $M_{\rm hm}$ value of 10^7, 10^8, 10^9, and $10^{10}\,{\rm M}_\odot$. Each curve is the posterior obtained for a set of $N = 20$ lenses. Each of the mock observations has lens and source parameters sampled from their own final constrained proposal distribution, and different substructure population.

Now that we have reduced the estimator noise, it is straightforward to perform inference on a group of sets of images by combining their ratios. Given a data set $X^N = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ of images, the combined ratio for multiple M data sets is simply given by $r(X_M^N, \vartheta) \propto \prod_{i=1}^{M} r(X_i^N, \vartheta)$, where the proportionality is a ratio of pieces of evidence, independent of the parameter value, so it only accounts for a proper normalization (Brehmer et al. 2019; Hermans et al. 2020). In the second row of Fig. 9, we show the results for the

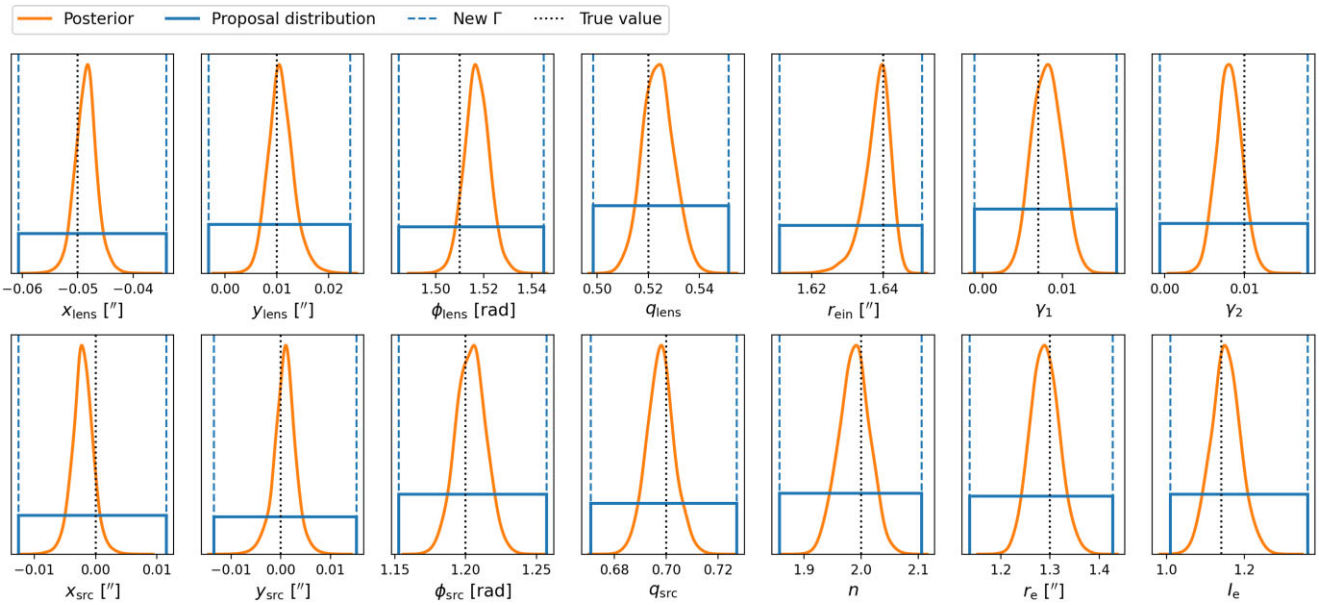

Figure 8. Lens and source parameter posteriors. In solid blue, we show the last round of constrained proposal distributions for the first (upper left corner, framed in orange) target image in Fig. 4. The dotted black lines correspond to the true lens and source parameter values with which we have generated the target image. In orange, we show the estimated posteriors for lens and source parameters in the last training round. Based on the predetermined threshold, the new bounding limits Γ (dashed blue) do not change significantly from the previous constrained proposal distribution region, so it is not possible to constrain the proposal distribution more and we stop the truncation procedure.

combination of the $M = 10$ different posteriors shown in the first column.

In the third row, we show a combined posterior for the WDM mass function from 200 images ($M = 10$ sets of $N = 20$ images). These plots show the uncertainty in the subhalo mass function under the assumption that it has the functional form in equation (14) with parameters from Lovell (2020).

These first results show that our method is sensitive to the low-mass end of the HMF, and that we have unbiased results from combining just 10 sets of 20 observations, given that in the second panel of Fig. 9 the true input value for the half-mode mass M_{hm} is consistently contained within the estimated posterior. In Section 4.5, we will show a more sophisticated method to assess the statistical behaviour of our inference results.

Furthermore, we can translate the constraints we obtain on the cut-off mass to constraints on the WDM mass given the mapping between those two quantities defined in Section 2.4. In Fig. 10, we show our results for the WDM mass. Each column corresponds to a different cut-off mass input value, so a different WDM mass. In the first row, we plot five examples of the combined posterior density for $\log_{10} M_{\mathrm{hm}}$ of $M = 10$ sets of $N = 20$ observations. In the second row, we show the corresponding colour coded five examples for m_{WDM}. In this case, we just transform the posterior from the first row using the parametrization shown in Section 2.4, so we assume a flat prior on $\log_{10} M_{\mathrm{hm}}$. Finally, in the last row, we show the WDM mass posterior densities assuming a flat prior on the latter. The posteriors in the second and third row are not actually the same because a flat prior $\log_{10} M_{\mathrm{hm}}$ is different from a flat prior on m_{WDM}.

4.5 Credible interval testing

We would like to directly test and validate the statistical behaviour of our inference results by determining the expected coverage of the ratio estimator produced by the network. This can be easily done in SWYFT thanks to local amortization (Miller et al. 2022). The goal is to compare the nominal and empirical expected coverage probabilities of estimated Bayesian credible intervals, which should coincide for a well-calibrated estimator. For the statistical formalism and definition of credible region and expected coverage probability, we refer the reader to Hermans et al. (2021). In brief, an ideal estimator has matching empirical and nominal expected coverage, a conservative one predicts lower credibility than empirically obtained, and an overconfident one has higher nominal than empirical credibility. In plots like Fig. 11, the line for an ideal ratio estimator should perfectly align with the diagonal, whereas for a conservative (overconfident) estimator, it will lie above (below) the diagonal. In combination with visually checking the posteriors, this test supports the accuracy of the posterior estimator and is also particularly useful when one does not have access to the ground truth against which to compare the results. In Fig. 11, we show the empirical versus nominal expected coverage probabilities for the cut-off mass inference network. We can see that the inference network for the half-mode mass has converged with good expected coverage.

5 DISCUSSION

In this section, we discuss the improvements to the model and inference question that need to be addressed before we can safely apply our pipeline to the analysis of real data.

First, we have neglected effects such as inadequate lens light subtraction and assumed the lens light to be known. Regarding the noise model, we did not account for correlated pixel noise due to instrumental effects including the telescope's PSF (see e.g. Wagner-Carena et al. 2022).

In this work, we have employed an analytical parametrization (the Sérsic profile) as a lensed source light distribution model, which is adequate to analyse low-resolution images. However, to accurately model higher fidelity lensing observations, such as those from

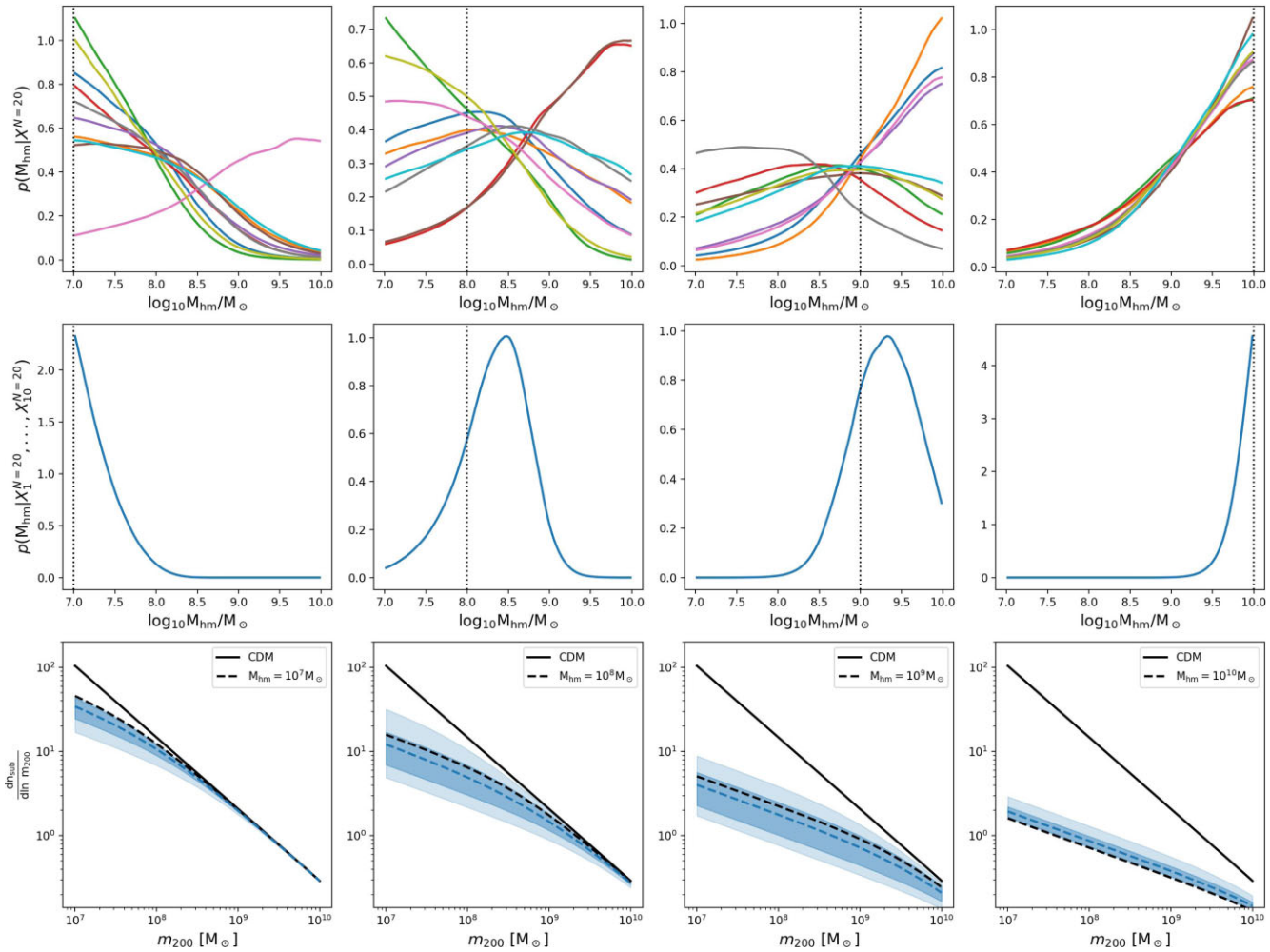

Figure 9. *Top:* Approximate posteriors for the half-mode mass derived from 10 different sets of 20 images. The dotted black line represents the true value of the half-mode mass with which we have generated the images (10^7, 10^8, 10^9, 10^{10} $M_\odot$). *Middle:* We show the approximate posterior resulting from the combination of the $M = 10$ different posteriors shown in the first column, as explained in the text (Section 4.4). *Bottom:* Subhalo mass function constraints derived from the cut-off mass posterior shown in the second column. The black solid line shows the CDM subhalo mass function according to equation (9), whereas the black dashed one shows the WDM subhalo mass function according to equation (14), given the true cut-off mass shown in the label. The blue dashed line shows the mean of the WDM subhalo mass function obtained by sampling 1000 samples from the cut-off mass posterior shown in the second panel and using this value in equation (14). We also show the central 68 and 95 percentiles as shaded bands. These plots show how uncertain the subhalo mass function is under the assumption that it has the functional form in equation (14) with parameters from Lovell (2020).

ongoing (e.g. *HST*) and future (e.g. *JWST*, *ELT*, *SKA*) telescopes, more complex source models need to be employed. Existing models, in order of complexity, are regularized pixellation of the source plane (see e.g. Suyu et al. 2006; Vegetti & Koopmans 2009a; Karchev et al. 2021), source modelling through basis functions [e.g. shapelets, (Birrer & Amara 2018) or wavelets (Galan et al. 2021)] attached to the source plane, and deep learning approaches (see e.g. Morningstar et al. 2019; Adam, Perreault-Levasseur & Hezaveh 2022). The ability to accurately and precisely reconstruct the complex morphology of strong lensing sources is of the utmost importance, as to disentangle the source surface brightness inhomogeneities from the per cent-level fluctuations introduced by substructures in the lens. We anticipate that using sources with more complex morphologies will result in higher sensitivity to the DM cut-off mass, provided that it is possible to model these sources. In fact, the residuals between the image of an extended source lensed by the total lens potential (accounting for substructures) and that of the same source lensed only by the main lens component are proportional to the gradient of that source evaluated in the image plane (equation 16; Cyr-Racine et al. 2019).

Regarding DM modelling, validation of our smoothing scheme (Section 2.4.1) is required to accurately account for DM free-streaming effects. Moreover, we should account for uncertainties due to the assumed halo density profile by considering different DM distributions around galaxies (see e.g. Salucci 2019, for a review).

Finally, we would like to draw the reader's attention on the fact that in our modelling we assume that the halo mass of the lens is known exactly from its Einstein radius (see Section 2.3.2). This is a strong assumption that has as a consequence the separation of substructure parameters $\boldsymbol{\theta}_h$ and lens parameter $\boldsymbol{\theta}_l$ once we marginalize the posterior probability over the halo mass in equation (16). The inference question we have addressed in this work, constraining the cut-off mass of the subhalo mass distribution, is then a simplified version of the real one, which is to simultaneously determine the halo mass and subhalo mass distribution of the lenses from real data (see e.g. Birrer et al. 2017).

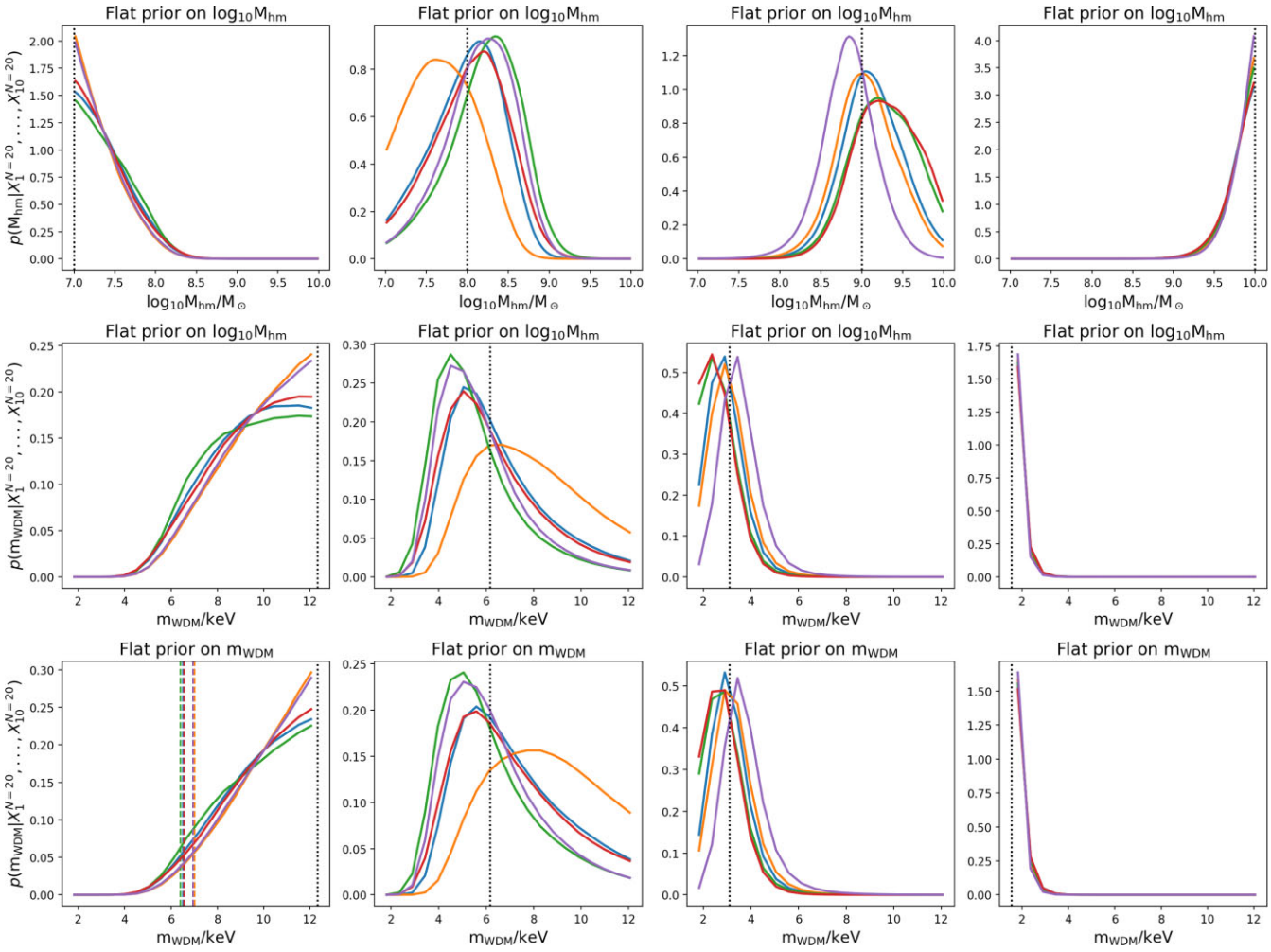

Figure 10. *Top:* We show five examples of combined posterior of $M = 10$ sets of $N = 20$ observations in terms of the cut-off mass (as the second row in Fig. 9). The dotted black line represents the true input value of the half-mode mass with which we have generated the analysed mock observations (10^7, 10^8, 10^9, 10^{10} $M_\odot$). *Middle:* Same results as shown in the first column but for the WDM mass. The dotted black line represents the true value of the WDM mass with which we have generated the analysed mock observations, given the mapping between DM cut-off and DM mass in Section 2.4. The WDM mass posteriors assume a flat prior on the cut-off mass. *Bottom:* Same results as shown in the first column but for the WDM mass and assuming a flat prior on the latter. In the first plot of the row, we show for the five examples the expected 95 per cent credible lower limit on the WDM mass for the highest value of our prior distribution.

We believe there are no major obstacles in incorporating all of these modelling components in our framework without fundamentally altering the inference procedure.

6 CONCLUSIONS

Strong gravitational lensing as a probe of the particle nature of DM has sparked much interest over the last few years. Moreover, the development of fast and accurate techniques to extract information from strong lensing images is well motivated by the wealth of new high-resolution strong lensing observations that will become available in the near future.

In this work, we have presented the first step towards a new neural simulation-based inference pipeline (see Section 3) to analyse present and future strong gravitational lensing systems in order to constrain the cut-off in the DM HMF, and so the DM mass. To this end, we have used a recent machine learning development, TMNRE, that makes it possible to *target* the analysis to a specific observation rather than amortize over all possible variations in lensing systems, making inference more efficient and precise. Thanks to TMNRE, we overcome the computational challenges of traditional MCMC, nested sampling, and transdimensional MCMC methods, by directly learning the marginal posterior for the parameter of interest from the observation. TMNRE leverages neural networks to directly learn the best summary statistic possible from the full input data, without having to compress the observation into hand-crafted summary statistics. This work is then a step forward towards making the analysis of strong lensing images for DM science faster, more efficient, and more accurate. In addition, our inference results can be validated with expected coverage tests (see Section 4.5).

Our key results can be summarized as follows:

(i) Thanks to our targeted approach, we are able to correctly estimate the lens and source parameter uncertainties, accounting for the presence of substructures in the mass range $[10^7, 10^{10}]\,M_\odot$. We use the final lens and source parameters' truncated proposal distributions (see Section 4.3) to generate a targeted training data set in order to infer the DM cut-off.

(ii) In the case that DM is warm, we are able to infer the location of the cut-off in the HMF in the $[10^7, 10^{10}]\,M_\odot$ mass range by combining

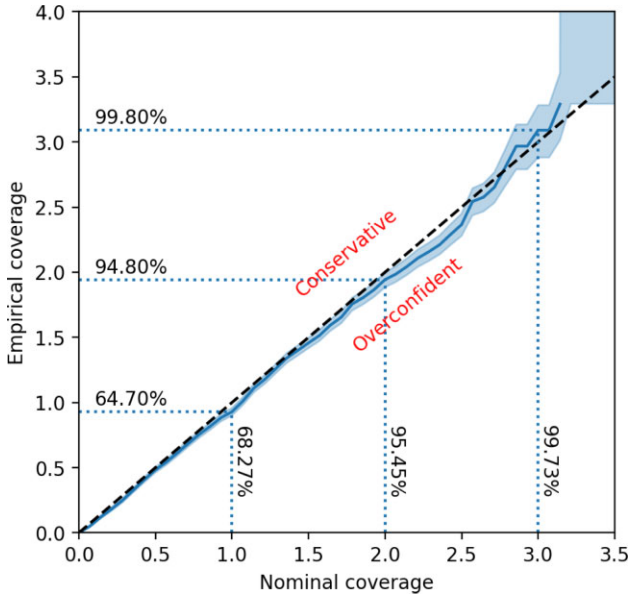

Figure 11. Empirical versus nominal expected coverage probabilities for the cut-off mass inference network. In case the line lies above (below) the black dashed diagonal line, the credible intervals are conservative (overconfident) and contain the true value with a frequency higher (lower) than nominally expected. We show the empirical (nominal) probabilities as horizontal (vertical) text.

up to 200 observations (see Section 4.4). We show our results in Fig. 9. By construction, these results are correctly marginalized over model uncertainties and have proper expected coverage (see Section 4.5).

(iii) A cut-off mass posterior translates into a posterior on the WDM mass, given the mapping in Section 2.4. We show our results in Fig. 10 for a flat prior on the cut-off mass and a flat prior on the WDM mass. We obtain an expected 95 per cent credible lower limits around 6.5 keV in the case of the scenario closest to CDM (see the bottom left panel in Fig. 10), given the adopted prior and the various assumptions of our simulation model that will be discussed below.

Throughout this study, we have made a number of simplifying assumptions for the halo mass of the lens, source light profile, and substructure models. We have also neglected effects such as inadequate lens light subtraction, realistic PSF modelling, and correlated pixel noise due to effects including the telescope's PSF. Before this analysis pipeline can be safely extended to real observations, these assumptions need to be correctly addressed, as discussed in Section 5.

In this work, we have demonstrated that, in principle, the DM cut-off mass signal can be statistically extracted from a population of small-scale DM haloes by a neural network using TMNRE. In future works, we plan on studying the correlation between different subhalo mass function parameters (e.g. its normalization, slope, and cut-off mass), and the one between the halo mass and subhalo mass distribution of the lenses, using, on one hand, more advanced modelling techniques (as specified above) on multiband observations and, on the other hand, better neural network architectures to target low SNR scenarios.

Finally, we note that, thanks to its flexibility, our pipeline can incorporate any arbitrary DM model, as long as it specifies the form of the HMF and the density profiles of individual substructures. We are optimistic that the presented Bayesian inference pipeline will be able to constrain the amount of substructures, pinning down DM nature, using both the strong lensing images that exist today and the wealth of new strong lensing data coming from near-future observatories.

ACKNOWLEDGEMENTS

We thank Benjamin Kurt Miller and Elias Dubbeldam for helpful discussions. We thank the anonymous referee for a careful reading and helpful comments. This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 864035 – UnDark). AC received funding from the Netherlands eScience Center (grant number ETEC.2019.018) and the Schmidt Futures Foundation. CC acknowledges the support of the Dutch Research Council (NWO Veni 192.020). This work was carried out on the Lisa Compute Cluster at SURFsara. We acknowledge the use of the PYTHON (Van Rossum & Drake 1995) modules, MATPLOTLIB (Hunter 2007), SEABORN (Waskom 2021), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), ASTROPY (Astropy Collaboration 2018), PYTORCH (Paszke et al. 2019), PYRO (Bingham et al. 2019), TQDM (Costa-Luis et al. 2021), and JUPYTER (Kluyver et al. 2016).

DATA AVAILABILITY

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

REFERENCES

Abolfathi B. et al., 2021, ApJS, 253, 31
Adam A., Perreault-Levasseur L., Hezaveh Y., 2022, ICML 2022 Workshop on Machine Learning for Astrophysics, preprint (arXiv:2207.01073)
Ade P. A. R. et al., 2016, A&A, 594, A13
Alexander S., Gleyzer S., McDonough E., Toomey M. W., Usai E., 2020, ApJ, 893, 15
Amorisco N. C. et al., 2021, MNRAS, 510, 2464
Astropy Collaboration, 2018, AJ, 156, 123
Baltz E. A., Marshall P., Oguri M., 2009, J. Cosmol. Astropart. Phys., 2009, 015
Bayer D., Chatterjee S., Koopmans L. V. E., Vegetti S., McKean J. P., Treu T., Fassnacht C. D., 2018, preprint (arXiv:1803.05952)
Bingham E. et al., 2019, J. Mach. Learn. Res., 20, 28
Birrer S., Amara A., 2018, Phys. Dark Universe, 22, 189
Birrer S., Amara A., Refregier A., 2017, J. Cosmol. Astropart. Phys., 2017, 037
Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, ApJ, 638, 703
Bond J. R., Szalay A. S., Turner M. S., 1982, Phys. Rev. Lett., 48, 1636
Boyarsky A., Drewes M., Lasserre T., Mertens S., Ruchayskiy O., 2019, Prog. Part. Nucl. Phys., 104, 1
Brehmer J., Mishra-Sharma S., Hermans J., Louppe G., Cranmer K., 2019, ApJ, 886, 49
Brennan S., Benson A. J., Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., Pullen A. R., 2019, MNRAS, 488, 5085
Brewer B. J., Huijser D., Lewis G. F., 2016, MNRAS, 455, 1819
Brownstein J. R. et al., 2011, ApJ, 744, 41
Bullock J. S., 2010, preprint (arXiv:1009.4505)
Chatterjee S., Koopmans L. V. E., 2017, MNRAS, 474, 1762
Ciotti L., Bertin G., 1999, A&A, 352, 447
Cole A., Miller B. K., Witte S. J., Cai M. X., Grootes M. W., Nattino F., Weniger C., 2022, J. Cosmol. Astropart. Phys., 2022, 004
Colin P., Avila-Reese V., Valenzuela O., 2000, ApJ, 542, 622

Coogan A., Karchev K., Weniger C., 2020, Accepted for the NeurIPS 2020 workshop Machine Learning and the Physical Sciences, preprint (arXiv:2010.07032)
Coogan A., Anau Montel N., Karchev K., Grootes M. W., Nattino F., Weniger C., 2022 (arXiv:2209.09918)
da Costa-Luis C., 2021, tqdm: A fast, Extensible Progress Bar for Python and CLI, https://joss.theoj.org/papers/c44313ada36f12eebbaff10eb0888071.pdf
Cranmer K., Brehmer J., Louppe G., 2020, Proc. Natl. Acad. Sci., 117, 30055
Cyr-Racine F.-Y., Keeton C. R., Moustakas L. A., 2019, Phys. Rev. D, 100
Dalal N., Kochanek C. S., 2002, ApJ, 572, 25
Daylan T., Cyr-Racine F.-Y., Rivero A. D., Dvorkin C., Finkbeiner D. P., 2018, ApJ, 854, 141
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
Diaz Rivero A., Cyr-Racine F.-Y., Dvorkin C., 2018a, Phys. Rev. D, 97
Díaz Rivero A., Dvorkin C., Cyr-Racine F.-Y., Zavala J., Vogelsberger M., 2018b, Phys. Rev. D, 98
Drlica-Wagner A. et al., 2019, preprint (arXiv:1902.01055)
Galan A., Peel A., Joseph R., Courbin F., Starck J.-L., 2021, A&A, 647, A176
Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485
Gennaro M., 2018, WFC3 Data Handbook, Vol. 4, Baltimore: STScI, p. 4
Gilman D., Birrer S., Treu T., Keeton C. R., Nierenberg A., 2018, MNRAS, 481, 819
Gilman D., Birrer S., Treu T., Nierenberg A., Benson A., 2019a, MNRAS, 487, 5721
Gilman D., Birrer S., Nierenberg A., Treu T., Du X., Benson A., 2019b, MNRAS, 491, 6077
Giocoli C., Tormen G., Sheth R. K., van den Bosch F. C., 2010, MNRAS, 404, 502
Grazian C., Fan Y., 2019, preprint (arXiv:1909.02736)
Harris C. R. et al., 2020, Nature, 585, 357
He Q. et al., 2022a, MNRAS, 511, 3046
He Q. et al., 2022b, MNRAS, 512, 5862
Hermans J., Begy V., Louppe G., 2020, presented at ICML 2020, preprint (arXiv:1903.04057)
Hermans J., Delaunoy A., Rozet F., Wehenkel A., Louppe G., 2021, preprint (arXiv:2110.06581)
Hezaveh Y. D. et al., 2016a, ApJ, 823, 37
Hezaveh Y., Dalal N., Holder G., Kisner T., Kuhlen M., Levasseur L. P., 2016b, J. Cosmol. Astropart. Phys., 2016, 048
Hezaveh Y. D., Levasseur L. P., Marshall P. J., 2017, Nature, 548, 555
Hsueh J.-W., Fassnacht C. D., Vegetti S., McKean J. P., Spingola C., Auger M. W., Koopmans L. V. E., Lagattuta D. J., 2016, MNRAS, 463, L51
Hsueh J.-W. et al., 2017, MNRAS, 469, 3713
Hsueh J.-W., Enzi W., Vegetti S., Auger M. W., Fassnacht C. D., Despali G., Koopmans L. V. E., McKean J. P., 2019, MNRAS, 492, 3047
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Iqbal H., 2018, /HarisIqbal88/PlotNeuralNet
Karchev K., Coogan A., Weniger C., 2022, MNRAS, 512, 661
Kluyver T. et al., 2016, in Loizides F., Schmidt B., eds, Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, Amsterdam, Berlin, Washington, DC, p. 87
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82
Kochanek C. S., 2004, in Meylan G., Jetzer P., North P., eds, Proceedings of the 33rd Saas-Fee Advanced Course, The Saas Fee Lectures on Strong Gravitational Lensing. Springer-Verlag, Berline
Koopmans L., 2005, EAS Publ. Ser. Vol. 20, Gravitational Lensing & Stellar Dynamics. Cambridge Univ. Press, Cambridge, p. 161
Koopmans L., Browne I., Jackson N., 2004, New Astron. Rev., 48, 1085
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lovell M. R., 2020, ApJ, 897, 147
Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., Theuns T., 2014, MNRAS, 439, 300
McKean J. P. et al., 2015, Proc. Sci. Strong Gravitational Lensing with the SKA. SISSA, Trieste, PoS#84
Mao S., Schneider P., 1998, MNRAS, 295, 587
Meneghetti M., 2016, Introduction to Gravitational Lensing, Springer Nature, Switzerland
Miller B. K., Cole A., Louppe G., Weniger C., 2020, preprint (arXiv:2011.13951)
Miller B. K., Cole A., Forrè P., Louppe G., Weniger C., 2022, J. Open Source Softw., 7, 4205
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Morningstar W. R. et al., 2019, ApJ, 883, 14
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
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
Paszke A. et al., 2019, in Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E., Garnett R., eds, Advances in Neural Information Processing Systems 32. Curran Associates, Inc., Montreal, Canada, p. 8024
Peebles P. J. E., 1982, ApJ, 263, L1
Perreault Levasseur L., Hezaveh Y. D., Wechsler R. H., 2017, ApJ, 850, L7
Refregier A., Amara A., Kitching T. D., Rassat A., Scaramella R., Weller J., 2010, preprint (arXiv:1001.0061)
Richings J., Frenk C., Jenkins A., Robertson A., Schaller M., 2021, MNRAS, 501, 4657
Ritondale E., Vegetti S., Despali G., Auger M. W., Koopmans L. V. E., McKean J. P., 2019, MNRAS, 485, 2179
Rubin V. C., Ford W. K. J., Thonnard N., 1980, ApJ, 238, 471
Salucci P., 2019, A&AR, 27
Schneider A., Smith R. E., Macciò A. V., Moore B., 2012, MNRAS, 424, 684
Şengül A. Ç., Tsang A., Diaz Rivero A., Dvorkin C., Zhu H.-M., Seljak U., 2020, Phys. Rev. D, 102
Sérsic J. L., 1963, Bol. Asoc. Argentina Astron. Argentina, 6, 41
Simon J. D. et al., 2019, BAAS, 51, 153
Springel V. et al., 2008, MNRAS, 391, 1685
Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983
Suyu S. H., Marshall P. J., Blandford R. D., Fassnacht C. D., Koopmans L. V. E., McKean J. P., Treu T., 2009, ApJ, 691, 277
Taylor A. N., Dye S., Broadhurst T. J., Benitez N., van Kampen E., 1998, ApJ, 501, 539
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottl'ober S., Holz D. E., 2008, ApJ, 688, 709
Van Rossum G., Drake F. L. Jr, 1995, Python Reference Manual. Centrum voor Wiskunde en Informatica, Amsterdam
Vegetti S., Koopmans L. V. E., 2009a, MNRAS, 392, 945
Vegetti S., Koopmans L. V. E., 2009b, MNRAS, 400, 1583
Vegetti S., Czoske O., Koopmans L. V. E., 2010a, MNRAS, 407, 225
Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2010b, 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
Vernardos G., Tsagkatakis G., Pantazis Y., 2020, MNRAS, 499, 5641
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wagner-Carena S., Aalbers J., Birrer S., Nadler E. O., Darragh-Ford E., Marshall P. J., Wechsler R. H., 2022, preprint (arXiv:2203.00690)
Waskom M. L., 2021, J. Open Source Softw., 6, 3021
Zwicky F., 1933, Helv. Phys. Acta, 6, 110

APPENDIX A: SÉRSIC SOURCE

Here, we describe the elliptical coordinates (r_x, r_y), the normalization k_n, and the index n that enter in the modelling of the Sérsic profile in equation (6).

The transformation from Cartesian (x, y) to elliptical coordinates (r_x, r_y) is given by

$$\begin{pmatrix} r_x \\ r_y \end{pmatrix} = \begin{pmatrix} \sqrt{q} & 0 \\ 0 & 1/\sqrt{q} \end{pmatrix} \begin{pmatrix} \cos\phi_s & \sin\phi_s \\ -\sin\phi_s & \cos\phi_s \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}, \tag{A1}$$

where ϕ_s is the rotation angle, q_s is the axial ratio, and (x_0, y_0) is the centre of light position.

The normalization k_n is related to the index n by an implicit transcendental equation in terms of the complete and lower incomplete gamma functions $2\gamma(2n, k_n) = \Gamma(2n)$. We use the expansion in series from Ciotti & Bertin (1999), valid over a wide range of indices n, stopping at order $\mathcal{O}(n^{-3})$.

APPENDIX B: LINE-OF-SIGHT HALOES AS EFFECTIVE SUBHALOES

Following Şengül et al. (2020), LOS haloes at comoving distance χ can be treated as subhaloes on the main-lens plane with an effective projected mass density given by

$$\Sigma_{\chi,\mathrm{eff}}(D_l\vec{x}; m_{200}, r_s, \tau) = \Sigma(D_l\vec{x}; m_{200,\mathrm{eff}}, r_{s,\mathrm{eff}}, \tau). \tag{B1}$$

The effective scale radius $r_{s,\,\mathrm{eff}}$ and mass $m_{200,\,\mathrm{eff}}$ are respectively

$$r_{s,\mathrm{eff}} = \frac{D_l}{g(\chi)D_\chi} r_s, \tag{B2}$$

and

$$m_{200,\mathrm{eff}} = f(\chi)\frac{\Sigma_{\mathrm{cr},l}}{\Sigma_{\mathrm{cr},\chi}}\left(\frac{D_l}{g(\chi)D_\chi}\right)^2 m_{200}. \tag{B3}$$

The piecewise functions $f(\chi)$ and $g(\chi)$ are

$$f(\chi) = \begin{cases} 1 - \beta_{\chi l} & \chi \le \chi_l \\ 1 - \beta_{l\chi} & \chi > \chi_l \end{cases}, \tag{B4}$$

and

$$g(\chi) = \begin{cases} 1 & \chi \le \chi_l \\ 1 - \beta_{l\chi} & \chi > \chi_l \end{cases}, \tag{B5}$$

with $\beta_{ij} = \dfrac{D_{ij}D_s}{D_j D_{is}}$, where D_i is the angular diameter distance from the observer to plane i, and D_{ij} is the angular diameter distance from lens plane i to lens plane j, and χ_l is the comoving distance to the main-lens plane. We have also introduced the critical surface density at plane i

$$\Sigma_{\mathrm{cr},i} \equiv \frac{c^2 D_s}{4\pi G D_i D_{is}}. \tag{B6}$$

This paper has been typeset from a TEX/LATEX file prepared by the author.