

Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning

JOHANN BREHMER,^{1,2} SIDDHARTH MISHRA-SHARMA,¹ JOERI HERMANS,³ GILLES LOUPPE,³ AND KYLE CRANMER^{1,2}

¹*Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10003, USA*

²*Center for Data Science, New York University, 60 Fifth Ave, New York, NY 10011, USA*

³*University of Liège, Belgium*

ABSTRACT

The subtle and unique imprint of dark matter substructure on extended arcs in galaxy-galaxy strong lenses contains a wealth of information about the properties and distribution of dark matter on small scales and, consequently, about the underlying particle physics. Teasing out this effect poses a significant challenge however due to the high dimensionality of the underlying latent space associated with a large number of dark matter subhalos. We apply recently-developed simulation-based techniques to the problem of substructure inference in galaxy-galaxy strong lenses. By leveraging additional information extracted from the simulator, these methods can be used to train neural networks to estimate likelihood ratios associated with population-level parameters characterizing the distribution of substructure. We show through proof-of-principle application to simulated data how these methods can provide an efficient and principled way to infer substructure properties by simultaneously analyzing the large lens samples deliverable by upcoming surveys such as LSST and *Euclid*.

Keywords: strong gravitational lensing (1643) — gravitational lensing (670) — nonparametric inference (1903) — astrostatistics techniques (1886) — cosmology (343) — dark matter (353)

Contents		4. Results	8
1. Introduction	1	5. Extensions	8
2. Strong lensing formalism and simulation set-up	3	6. Conclusions	8
2.1. Strong lensing formalism	3		
2.2. Background source	3		
2.3. Lensing host galaxy	4		
2.4. Lensing substructure	4		
2.5. Observational considerations	5		
2.6. Population statistics of the lens and source samples	6		
3. Statistical formalism and simulation-based inference	6		
3.1. Extracting additional information from the simulator	7		
3.2. Machine learning	8		
3.3. Calibration	8		
3.4. Inference	8		

1. INTRODUCTION

Dark matter (DM) accounts for nearly a quarter of the energy budget of the Universe, and pinning down its fundamental nature and interactions is one of the most pressing problems in cosmology and particle physics today. Despite an organized effort to do so through terrestrial (Aprile et al. 2018; Cui et al. 2017; Akerib et al. 2017), astrophysical (Albert et al. 2017; Chang et al. 2018; Lisanti et al. 2018), and collider searches (Aaboud et al. 2019; Sirunyan et al. 2017), no conclusive evidence of interactions between the Standard Model (SM) and dark matter exists to-date.

An alternative and complementary approach involves studying dark matter directly through its irreducible gravitational interactions. The concordance Cold Dark Matter (CDM) framework of non-relativistic, collisionless dark matter particles provides an excellent description of the observed distribution of matter on large

scales. However, many well-motivated models predict deviations from CDM on smaller scales. Fundamental dark matter microphysical properties, such as its particle mass and self-interaction cross-section, can imprint themselves onto its macroscopic distribution in ways that can be probed by current and future experiments (Drlica-Wagner et al. 2019). As motivating examples, theories where dark matter has a significant free-streaming length would lead to a dearth of subhalos at lower masses ($\lesssim 10^9 M_\odot$) (Bond & Szalay 1983; Bode et al. 2001; Dalcanton & Hogan 2001; Boyanovsky et al. 2008; Boyanovsky & Wu 2011), and self-interactions (Kaplinghat et al. 2016, 2014; Zavala et al. 2013; Peter et al. 2013; Vogelsberger et al. 2012, 2019; Kahlhoefer et al. 2019) or dissipative dynamics (Agrawal & Randall 2017; Agrawal et al. 2017; Buckley & DiFranzo 2018; Fan et al. 2013; Vogelsberger et al. 2016) in the dark sector would modify the structure of the inner core of subhalos as compared to CDM predictions.

There exist several avenues for probing the structure of dark matter on small scales. While the detection of ultrafaint dwarf galaxies through the study of stellar overdensities and kinematics (Bechtol et al. 2015; Koposov et al. 2008, 2015) can be used to make statements about the underlying dark matter properties, theoretical uncertainties in the connection between stellar and halo masses (Wechsler & Tinker 2018) and the effect of baryons on the satellite galaxy population (Brooks 2018; Fitts et al. 2018; Garrison-Kimmel et al. 2017; Errani et al. 2017) pose a challenge. Furthermore, suppressed star-formation in smaller halos means that there exists a threshold ($\lesssim 10^8 M_\odot$) below which subhalos are expected to be mostly dark and devoid of stars (Efsthathiou 1992; Fitts et al. 2017; Read et al. 2017). This makes studying the imprints of gravitational interaction the *only* viable avenue for probing substructure at smaller scales. In this spirit, the study of perturbations to the stellar phase-space distribution in cold stellar streams (Bonaca & Hogg 2018; Johnston et al. 1999; Carlberg 2012; Carlberg & Grillmair 2013; Bonaca et al. 2019), and in stellar fields in the disk and halo (Buschmann et al. 2018) have been proposed as methods to look for low-mass subhalos through their gravitational interactions in the Milky Way.

Complementary to the study of locally-induced gravitational effects, gravitational lensing has emerged as an important tool for studying the distribution of matter over a large range of scales. Locally, the use of time-domain astrometry has been proposed as a promising method to measure the distribution of local substructure through correlated, lens-induced motions on back-

ground celestial objects (Van Tilburg et al. 2018). In the extragalactic regime, galaxy-scale strong lensing systems are a laboratory for studying substructure. The presence of flux-ratio anomalies in multiply-imaged quasar lenses has been used to infer the typical abundance of substructure within galaxy-scale lenses (Dalal & Kochanek 2002; Hsueh et al. 2019; Dalal & Kochanek 2002). Lensed images of extended sources have been used to find evidence for a handful of subhalos with masses $\gtrsim 10^8 M_\odot$ (Hezaveh et al. 2016b; Vegetti et al. 2010, 2012).

A complementary approach relies on probing the collective effect of sub-threshold (*i.e.*, not individually resolvable) subhalos on extended arcs in strongly lensed systems. A particular challenge here is the high dimensionality of the latent parameter space associated with the large number of subhalos and their (potentially covariant) individual as well as population properties, a consequence of which is the intractability of the likelihood of high-level substructure parameters conditional on the data. Methods based on summary statistics (Birrer et al. 2017) and studying the amplitude of spatial fluctuations on different scales through power spectra (Hezaveh et al. 2016a; Díaz Rivero et al. 2018; Díaz Rivero et al. 2018; Cyr-Racine et al. 2019; Brennan et al. 2019; Chatterjee & Koopmans 2018; Cyr-Racine et al. 2016) have been proposed as ways to reduce the dimensionality of the problem and enable substructure inference in a tractable way. Trans-dimensional techniques may also be able to efficiently map out the parameter space associated with multiple sub-threshold subhalos in these systems (Brewer et al. 2016; Daylan et al. 2018). This class of methods is well-suited to studying dark matter substructure since they can be sensitive to the *population* properties of low-mass subhalos in strongly lensed galaxies which are directly correlated with the underlying dark matter particle physics. Furthermore, near-future observatories like LSST (LSST Science Collaboration et al. 2009; Drlica-Wagner et al. 2019; Verma et al. 2019) and *Euclid* (Refregier et al. 2010) are expected to find tens of thousands of galaxy-galaxy strong lenses (Collett 2015), making substructure inference in these systems (and high-resolution followups on a subset) one of the key ways to investigate dark matter substructure and stress-test the Cold Dark Matter paradigm in the near future. This calls for methods that can efficiently analyze large samples of lensed images to infer the underlying substructure properties with minimal loss of information stemming from dimensional reduction.

In this paper we apply a powerful recently-developed class of techniques for simulation-based inference (Brehmer et al. 2018a,b,c) to the problem of ex-

tracting high-level substructure properties from an ensemble of galaxy-galaxy strong lensing images. In contrast to traditional simulation-based (or “likelihood-free”) approaches, these methods do not rely on summary statistics and instead leverage additional information extracted from the simulator in order to train neural networks that can be used to efficiently estimate the likelihood ratio. This provides an elegant bridge between machine learning and the ubiquitous likelihood ratio, which is provably the most powerful statistic for hypothesis testing (Neyman & Pearson 1933).

This paper is organized as follows. In Sec. 2 we briefly review the formalism of gravitational strong lensing and describe our simulation setup, including the assumptions we make about the population of lensed sources and host galaxies, the substructure population and observational parameters. In Sec. 3 we describe the simulation-based analysis technique used and its particular application to the problem of mining substructure properties from an ensemble of extended lensed arcs. We show a proof-of-principle application of this method to simulated data in Sec. 4 and comment on how these methods can be extended to more “realistic” scenarios in Sec. 5. We conclude in Sec. 6.

2. STRONG LENSING FORMALISM AND SIMULATION SET-UP

In strong lensing systems, the background light emission source can in general be a point-like quasar or supernova, or a faint, extended “blue” galaxy. The former results in multiple localized images on the lens plane rather than extended arc-like images, providing the ability to probe substructure over a limited region on the lens plane. For this reason, we focus our method towards galaxy-galaxy lenses — systems producing images with extended arcs — since we aim to disentangle the collective effect of a population of subhalo perturbers over multiple images. Young, blue galaxies are ubiquitous in the redshift regime $z \gtrsim 1$ and dominate the faint end of the galaxy luminosity function, resulting in a much larger deliverable sample of galaxy-galaxy strong lenses compared to quadruply- and doubly-imaged quasars/supernovae.

We now briefly review the basic strong lensing formalism before describing in turn the models for the background source, lensing galaxy and population parameters of the lens systems used in this study.

2.1. Strong lensing formalism

We briefly review here the mathematical formalism behind strong lensing. For more details see, *e.g.*, Keeton (2001); Schneider et al. (1992). For a mass distribution with dimensionless projected surface mass density

$\kappa(\mathbf{x}) = \Sigma(\mathbf{x})/\Sigma_{\text{cr}}$, where $\Sigma_{\text{cr}} \equiv \frac{1}{4\pi G_N} \frac{D_s}{D_{ls}D_l}$ the critical lensing surface density, the two-dimensional lensing potential is given by

$$\psi(\mathbf{x}) = \frac{1}{\pi} \int \ln |\mathbf{x} - \mathbf{y}| \kappa(\mathbf{y}) d\mathbf{y}. \quad (1)$$

The lensed position of the source can be determined through the lens equation,

$$\mathbf{u} = \mathbf{x} - \nabla\psi(\mathbf{x}) \quad (2)$$

where \mathbf{u} is the position of the source and $\nabla\psi$ is typically referred to as the deflection, which we will denote as ϕ for brevity. For an extended source brightness profile f_{src} , the final lensed image can be obtained as the source profile evaluated on the image plane,

$$f'_{\text{src}}(\mathbf{x}) = f_{\text{src}}(\mathbf{x} - \nabla\psi(\mathbf{x})). \quad (3)$$

For a spherically symmetric, the radial deflection field is given by

$$\phi_r(r) = \frac{2}{r} \int_0^r du u \kappa(u) = \frac{1}{\pi \Sigma_{\text{cr}}} \frac{M_{\text{cyl}}(r)}{r} \quad (4)$$

where $M_{\text{cyl}}(r)$ is the mass enclosed within a cylinder or radius r . Extension to the slightly more general case of elliptical symmetry is straightforward (see, *e.g.*, Keeton (2001)).

2.2. Background source

We model the emission from background source galaxies using a Sérsic profile, with the surface brightness given by

$$f_{\text{src}}(r) = f_e \exp \left\{ -b_n \left[\left(\frac{r}{r_e} \right)^{1/n} - 1 \right] \right\}, \quad (5)$$

where r_e is the effective circular half-light radius, n is the Sérsic index, and b_n is a factor depending on n that ensures that r_e contains half the total intensity from the source galaxy, given by (Ciotti & Bertin 1999)

$$b_n \approx 2n - \frac{1}{3} + \frac{4}{405n} + \frac{46}{25515n^2} + \frac{131}{1148175n^3} - \frac{2194697}{30690717750n^4}.$$

We assume $n = 1$ for the source galaxies, corresponding to a flattened exponential profile and consistent with expectation for blue-type galaxies at the relevant redshifts. f_e encodes the flux at half-light radius, which can be mapped onto the total flux (or magnitude) associated with a given galaxy.

The total unlensed magnitude M (in a given band) of a galaxy can be mapped on to f_e as follows. For

a detector with zero-point magnitude M_0 , which specifies the magnitude of a source giving 1 counts s^{-1} in expectation, by definition the total counts are given by $S_{\text{tot}} = 10^{0.4(M-M_0)}$. Requiring the half-light radius to contain half the expected counts, for $n = 1$ we have the relation $f_e \approx 0.526 t_{\text{exp}} S_{\text{tot}} / (2\pi r_e^2)$ in $\text{counts arcsec}^{-2}$, where t_{exp} is the exposure length.

Treatment of the other Sérsic parameters, in particular the total emission and half-light radius, in the context of population studies is described in Secs. 2.5 and 2.6 below.

2.3. Lensing host galaxy

Cosmological N -body simulations suggest that the dark matter distribution in structures at galactic scales can be well-described by a universal, spherically symmetric NFW profile. However, strong lensing probes a region of the host galaxy much smaller than the typical virial radii of galaxy-scale dark matter halo, and the mass budget here is dominated by the baryonic bulge component of the galaxy. Taking this into account, the total mass budget of the lensing host galaxy, being early-type, can be well describe by a singular isothermal ellipsoid (SIE) profile, known as the bulge-halo conspiracy since neither the dark matter nor the baryonic components are individually isothermal. The host profile is thus described as

$$\rho(x, y) = \frac{\sigma_v^2}{2\pi G (x^2/q + qy^2)} \quad (6)$$

where σ_v is the central 1-D velocity dispersion of the lens galaxy and q is the ellipsoid axis ratio, with $q = 1$ corresponding to a spherical profile. The Einstein radius for this profile, giving the characteristic lensing scale, is given by

$$\theta_E = 4\pi \left(\frac{\sigma_v}{c} \right)^2 \frac{D_{ls}(z_l, z_s)}{D_s(z_s)} \quad (7)$$

where D_{ls} and D_s are respectively the angular diameter distances from the source to the lens planes and from the source plane to the observer respectively.

The deflection field for the SIE profile is given by (Keeton 2001)

$$\phi_x = \frac{\theta_E q}{\sqrt{1-q^2}} \tan^{-1} \left[\frac{\sqrt{1-q^2} x}{\psi} \right] \quad (8)$$

$$\phi_y = \frac{\theta_E q}{\sqrt{1-q^2}} \tanh^{-1} \left[\frac{\sqrt{1-q^2} y}{\psi + q^2} \right] \quad (9)$$

with $\psi \equiv \sqrt{x^2 q^2 + y^2}$.

Although the total galaxy mass (baryons + dark matter) describe the macro lensing field, for the purposes

of describing substructure we require being able to map the measure properties of an SIE lens onto the properties of the host dark matter halo. To do this, we relate the central stellar velocity dispersion σ_v to the mass M_{200} of the host dark matter halo. Zahid et al. (2018) derived a tight correlation between σ_v and M_{200} , modeled as

$$\log \left(\frac{M_{200}}{10^{12} M_\odot} \right) = \alpha + \beta \left(\frac{\sigma_v}{100 \text{ km s}^{-1}} \right) \quad (10)$$

with $\alpha = 0.09$ and $\beta = 3.48$. We model the host dark matter halo with a Navarro-Frenk-White (NFW) profile (Navarro et al. 1996, 1997)

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2} \quad (11)$$

where ρ_s and r_s are the scale density and scale radius, respectively. The halo virial mass M_{200} describes the total mass contained within the virial radius r_{200} , defined as the radius within which the mean density is 200 times the critical density of the universe and related to the scale radius through the concentration parameter $c_{200} \equiv r_{200}/r_s$. Thus, an NFW halo is completely described by the parameters $\{M_{200}, c_{200}\}$. We use the concentration-mass relation from Sánchez-Conde & Prada (2014) assuming a log-normal distribution for c_{200} around the median inferred value given by the relation with scatter 0.15 dex.

The spherically-symmetric deflection for an NFW perturber is given by (Keeton 2001)

$$\phi_r = 4\kappa_s r_s \frac{\ln(x/2) + \mathcal{F}(x)}{x} \quad (12)$$

where $x = r/r_s$, $\kappa_s = \rho_s r_s / \Sigma_{\text{cr}}$ with Σ_{cr} the critical surface density, and

$$\mathcal{F}(x) = \begin{cases} \frac{1}{\sqrt{x^2-1}} \tan^{-1} \sqrt{x^2-1} & (x > 1) \\ \frac{1}{\sqrt{1-x^2}} \tanh^{-1} \sqrt{1-x^2} & (x < 1) \\ 1 & (x = 1). \end{cases} \quad (13)$$

We described the population parameters we use to model the host velocity dispersion (and thus its Einstein radius and dark matter halo mass) in Secs. 2.5 and 2.6 below.

2.4. Lensing substructure

The ultimate goal of our method is to characterize the substructure population in strong lenses. Here we describe our procedure to model the substructure contribution to the lensing signal. Understanding the expected abundance of substructure in galaxies over a large range of epochs is complex undertaking and an active area of research. Properties of individual subhalos (such as their

density profiles) as well as those that describe their population (such as the mass and spatial distribution) are strongly affected by their host environment, and accurately modeling all aspects of subhalo evolution and environment is beyond the scope of this paper. Instead, we use simple physically justifiable assumptions to model the substructure contributions in order to highlight the broad methodological points associated with the application of our method.

Λ CDM, often called the standard model of cosmology, predicts a scale-invariant power spectrum of primordial fluctuations and the existence of substructure over a broad range of masses with equal contribution per logarithmic mass interval. We parameterize the distribution of subhalo masses m_{200} in a given host halo of mass M_{200} — the subhalo mass function — as power law distribution:

$$\frac{M_{200,0}}{M_{200}} \frac{dn}{dm_{200}} = \alpha \left(\frac{m_{200}}{m_{200,0}} \right)^\beta \quad (14)$$

where α encodes the overall substructure abundance, with larger α corresponding to more substructure, and the slope β encodes the relative contribution of subhalos at different masses, with more negative β corresponding to a steeper slope with more low-mass subhalos. The normalization factors $m_{200,0}$ and $M_{200,0}$ are arbitrarily set to $10^9 M_\odot$ and the Milky Way mass $M_{\text{MW}} \simeq 1.1 \times 10^{12} M_\odot$, respectively.

Theory and simulations within the framework of Λ CDM predict a slope $\beta \sim -0.9$, giving a nearly scale-invariant spectrum of subhalos, which we assume in our fiducial setup.

We parameterize the overall subhalo abundance through the mass fraction contained in subhalos, f_{sub} , defined as the fraction of the total dark matter halo mass contained in bound substructure in a given mass range. We have

$$f_{\text{sub}} = \frac{\int_{m_{200,\text{min}}}^{m_{200,\text{max}}} dm_{200} m_{200} \frac{dn}{dm_{200}}}{M_{200,\text{host}}} \quad (15)$$

For a given f_{sub} , β and host halo mass $M_{200,\text{host}}$, this can be used to determine α in Eq. 14. The linear scaling of the subhalo mass function with the host halo mass M_{200} in Eq. 14 is additionally described in Han et al. (2016); Despali & Vegetti (2017). In our fiducial setups, we take the minimum mass $m_{200,\text{min}} = 10^7 M_\odot$ and $m_{200,\text{max}} = 0.01 M_{200,\text{host}}$ (Despali & Vegetti 2017; Hiroshima et al. 2018), and corresponding fiducial substructure fraction in this range of 5%, roughly consistent with Hiroshima et al. (2018); Hsueh et al. (2019); Dalal & Kochanek (2002) within our considered mass range.

With all parameters of the subhalo mass function specified, the total number n_{tot} of subhalos expected

within the virial radius R_{200} of the host halo can be inferred as $\int_{m_{200,\text{min}}}^{m_{200,\text{max}}} dm_{200} \frac{dn}{dm_{200}}$. Strong lensing probes a region much smaller this scale — the typical Einstein radii for the host deflector are much smaller than the virial radius of the host dark matter halos. In order to obtain the expected number of subhalos within the lensing observations region of interest, we scale the total number of subhalos obtained from the above procedure by the ratio of projected mass within our region of interest θ_{ROI} and the host halo mass M_{200} as follows. We assume the subhalos to be distributed in number density following the host NFW dark matter profile. In this case, the NFW enclosed mass function is $M_{\text{enc}}(x) = M_{200}(\ln(x/2) + \mathcal{F}(x))$ (Keeton 2001), where x is the angular radius in units of the virial radius, $x \equiv \theta/\theta_{200}$ and \mathcal{F} is given by Eq. 13 above. The number of subhalos within our ROI is thus obtained as $n_{\text{ROI}} = n_{\text{tot}}(\ln(x_{\text{ROI}}/2) + \mathcal{F}(x_{\text{ROI}}))$. We conservatively take the lensing ROI to enclose a region of angular size twice the Einstein radius of the host halo, $\theta_{\text{ROI}} = 2\theta_E$.

Since strong lensing probes the line-of-sight distribution of subhalos within the host, their projected spatial distribution is approximately uniform within the lensing ROI (Despali & Vegetti 2017). We thus distribute subhalos uniformly within our ROI. The density profile of subhalos is assumed to be NFW and given by Eq. 11, with associated lensing properties as described and the concentration inferred from the relation modeled in Sánchez-Conde & Prada (2014).

We finally emphasize that we do not intent to capture all of the intricacies of the subhalo distribution, such as the effects of baryonic physics, tidal disruption of subhalos in proximity to the center of the host and redshift evolution of host as well as substructure properties. Although our description can be extended to take these into account, their precise characterization and effect is still subject to large uncertainties, and our simple model above captures the essential physics for demonstration purposes.

2.5. Observational considerations

As noted above, our method is best-suited to analyzing a statistical sample of strong lenses to search for substructure, such as those that are expected to be obtained in the near future with optical telescopes like *Euclid* and LSST. Given the challenges associated with the precise characterization of such a sample at the present time, we describe here the observational characteristics we assume in order to build up training and testing samples to validate our inference techniques.

We largely follow the description in Collett (2015), and use the associated *LensPop* package, to characterize our

mock observations. In particular, we use the detector configuration for *Euclid*, assuming a zero-point magnitude $m_{\text{AB}} = 25.5$ in the single optical VIS passband, pixel size 0.1 arcsec, a Gaussian point spread function (PSF) with FWHM 0.18 arcsec, individual exposures with exposure time 1610 s, an isotropic sky background with magnitude 22.8 arcsec $^{-2}$ in the detector passband.

These properties, in particular the exposure, sky background and PSF shape are expected to vary somewhat across the lens sample. Additionally, a given region may be imaged by multiple exposures over a range of color bands. Although these variations can easily be incorporated into our analysis, modeling this is beyond the scope of this study. We briefly comment on how this information can be taken into account later.

2.6. Population statistics of the lens and source samples

The fact that the strong lens population is expected to be dominated by higher-redshift ($z \gtrsim 1$) blue source galaxies lensed by intermediate-redshift ($z \sim 0.5$ –1) elliptical galaxies presents significant challenges for quantifying the lens population obtainable with future observations. Specifically, planned ground-based surveys like LSST and space telescopes like *Euclid* present complementary challenges for delivering images of strong lensing systems suitable for substructure studies. LSST is expected to image in six bands, allowing efficient source selection and distinguishing source and lens emission, but at the cost of lower resolution by virtue of being a ground-based instrument. *Euclid* imaging is expected to be much higher in resolution but with a single optical passband (VIS). Near-IR imaging from WFIRST may deliver a high-resolution, multi-wavelength dataset that is more suitable for substructure studies, although the lens and source populations may differ from those probed by optical telescopes.

In light of these uncertainties, we confine ourselves to a setting where the main methodological points can be made without detailed modeling of the detector capabilities and the deliverable lensing dataset, which is outside of the scope of the current paper. For concreteness, we simulate a sample of lenses with a simplified subset of host galaxy properties consistent with those deliverable by *Euclid* as modeled by Collett (2015). In particular, we assume spherical lenses, with ellipticity parameter $q = 1$ in Eq. 6. We draw the central 1-D velocity dispersions σ_v of host galaxies from a normal distribution with mean 225 km s $^{-1}$ and standard deviation 50 km s $^{-1}$. Equation 10 the results of Zahid et al. (2018) are used to map the drawn σ_v to a dark matter halo mass M_{200} ,

and the host Einstein radius is analytically inferred with Eq. 7.

We draw the lens redshifts z_l from a log-normal distribution with mean 0.56 and scatter 0.25 dex, discarding lenses with $z_l > 1$ as these tend to have a small angular size over which substructure perturbations are relevant. The source offsets θ_x and θ_y are drawn from a normal distribution with zero mean and standard deviation 0.2. These are consistent with the lens sample generated from the *LensPop* code packaged with Collett (2015).

3. STATISTICAL FORMALISM AND SIMULATION-BASED INFERENCE

Our goal is to infer the subhalo mass function parameters from a catalog of images of observed lenses. In this section we will describe the challenges of this inference problem and our approach of simulation-based inference. For simplicity, we will use a more abstract notation, distinguishing between three sets of quantities in the lensing system:

Parameters of interest θ : The vector $\theta = (f_{\text{sub}}, \beta)^T$ parameterizes the subhalo mass function given, our goal is to infer their values.

Latent variables z : A vector of all other unobservable random variables in the simulator. These include the mass and offset of the host galaxy, the number of subhalos in the region of interest, the position and mass of each subhalo, and the random variables related to the point spread function and Poisson fluctuations.

Observables x : The observed lens images.

As described in the previous section, we have implemented a simulator for the lensing process in the “forward” direction: for given parameters θ , the simulator samples latent variables z and observed images $x, z \sim p(x, z|\theta)$. The joint likelihood of observables and latents $p(x, z|\theta)$ is schematically given by

$$p(x, z|\theta) = p_{\text{host}}(z_{\text{host}}) \text{Pois}(n|\bar{n}_{\text{ROI}}(\theta)) \prod_i^n [p_m(m_{200,i}|\theta) \text{Uniform}(r_i)] \times p_{\text{obs}}(x|f(z_{\text{host}}, \{(m_{200,i}, r_i)\})) \quad (16)$$

where p_{host} summarizes the distribution of the host halo parameters z_{host} such as its mass, offset and redshift; n is the actual number of subhalos in the region of interest, $\bar{n}_{\text{ROI}}(f_{\text{sub}}, \beta)$ is the mean number of subhalos in the region of interest as a function of the parameters

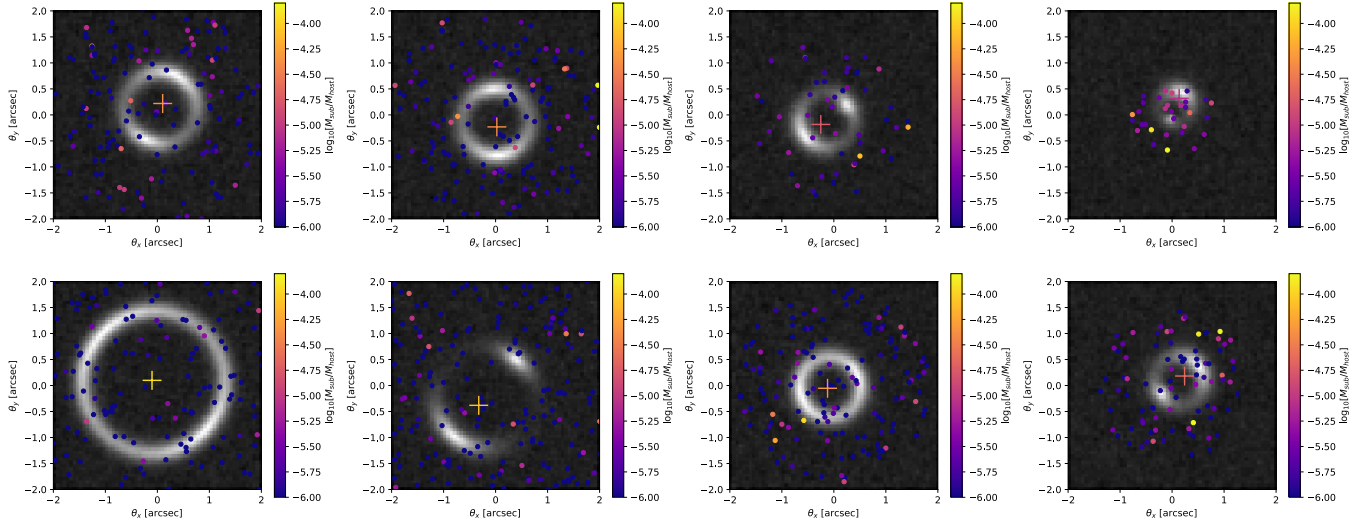


Figure 1.

$\theta = (f_{\text{sub}}, \beta)^T$; $m_{200,i}$ and r_i are the subhalo masses and positions; $p_m(m) = 1/n \, dn/dm_{200}$ is the normalized subhalo mass function given in Eq. (14); and in the last line p_{obs} is the probability of observing an image x based on the true lensed image $f(z_{\text{host}}, \{(m_{200,i}, r_i)\})$ due to the point spread function and Poisson fluctuations.

Inference with Markov Chain Monte Carlo (MCMC) methods based directly on this joint likelihood function $p(x, z|\theta)$ is feasible, but requires a large number of simulations for every observed lens. Instead, we use an approach that is based on the likelihood function of only the observables,

$$p(x|\theta) = \int dz \, p(x, z|\theta), \quad (17)$$

where the latent variables are marginalised out. If we are able to evaluate this likelihood function efficiently enables efficient frequentist and Bayesian inference. Unfortunately, even in our somewhat simplified simulator, each run of the simulation easily involves more than a thousand latent variables, and the integral over this enormous space clearly cannot be computed explicitly. The likelihood function $p(x|\theta)$ is thus intractable, providing a major challenge for both frequentist and Bayesian inference.

This issue is often tackled by reducing the high-dimensional data x to lower-dimensional summary statistics $v(x)$, for instance based on power spectra (Hezaveh et al. 2016a; Díaz Rivero et al. 2018; Díaz Rivero et al. 2018; Cyr-Racine et al. 2019; Brennan et al. 2019; Chatterjee & Koopmans 2018; Cyr-Racine et al. 2016). The likelihood $p(v|\theta)$ in the space of summary statistics can either be explicitly estimated through density estimation techniques such as histograms, kernel density estimation, or Gaussian processes, or replaced

by a rejection probability in an Approximate Bayesian Computation (ABC) technique. While the compression to summary statistics makes the analysis tractable, it typically loses information and hence reduces the statistical power of the analysis.

We follow an alternative approach of simulation-based inference introduced in Refs. Brehmer et al. (2018a,b,c). It consists of four steps:

1. During each run of the simulator, additional information that characterizes the subhalo population and lensing process is stored together with the simulated observed image.
2. This information is used to train a neural network to approximate the likelihood ratio function.
3. The neural network output is calibrated, ensuring that errors during training do not lead to wrong inference results.
4. The calibrated network output is then used in either frequentist or Bayesian inference techniques.

This approach enables amortized inference: after an upfront simulation and training phase, the neural network that estimates the likelihood ratio function can be evaluated very efficiently for any parameter point and observed image. Unlike MCMC based on the joint likelihood, this scales well to the expected large number of lenses expected in upcoming surveys.

In the remainder of this section, we will explain these four steps in detail.

3.1. Extracting additional information from the simulator

- 3.2. *Machine learning*
- 3.3. *Calibration*
- 3.4. *Inference*
- 4. RESULTS
- 5. EXTENSIONS
- 6. CONCLUSIONS

We thank Simon Birrer, Christopher Fassnacht, Daniel Gilman, and Neal Weiner for useful conversa-

tions. SM is partially supported by the NSF CAREER grant PHY-1554858 and NSF grant PHY-1620727.

Software: *Astropy* (Astropy Collaboration et al. 2013, 2018), *IPython* (Perez & Granger 2007), *LensPop* (Collett 2015), *matplotlib* (Hunter 2007), *NumPy* (van der Walt et al. 2011), *PyTorch* (Paszke et al. 2017), *SciPy* (Jones et al. 2001–).

REFERENCES

- Aaboud, M., Aad, G., Abbott, B., et al. 2019, *Journal of High Energy Physics*, 2019, 142, doi: [10.1007/JHEP05\(2019\)142](https://doi.org/10.1007/JHEP05(2019)142)
- Agrawal, P., Cyr-Racine, F.-Y., Randall, L., & Scholtz, J. 2017, *JCAP*, 2017, 021, doi: [10.1088/1475-7516/2017/08/021](https://doi.org/10.1088/1475-7516/2017/08/021)
- Agrawal, P., & Randall, L. 2017, *JCAP*, 2017, 019, doi: [10.1088/1475-7516/2017/12/019](https://doi.org/10.1088/1475-7516/2017/12/019)
- Akerib, D. S., Alsum, S., Araújo, H. M., et al. 2017, *PhRvL*, 118, 021303, doi: [10.1103/PhysRevLett.118.021303](https://doi.org/10.1103/PhysRevLett.118.021303)
- Albert, A., Anderson, B., Bechtol, K., et al. 2017, *ApJ*, 834, 110, doi: [10.3847/1538-4357/834/2/110](https://doi.org/10.3847/1538-4357/834/2/110)
- Aprile, E., Aalbers, J., Agostini, F., et al. 2018, *PhRvL*, 121, 111302, doi: [10.1103/PhysRevLett.121.111302](https://doi.org/10.1103/PhysRevLett.121.111302)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Bechtol, K., Drlica-Wagner, A., Balbinot, E., et al. 2015, *ApJ*, 807, 50, doi: [10.1088/0004-637X/807/1/50](https://doi.org/10.1088/0004-637X/807/1/50)
- Birrer, S., Amara, A., & Refregier, A. 2017, *JCAP*, 2017, 037, doi: [10.1088/1475-7516/2017/05/037](https://doi.org/10.1088/1475-7516/2017/05/037)
- Bode, P., Ostriker, J. P., & Turok, N. 2001, *ApJ*, 556, 93, doi: [10.1086/321541](https://doi.org/10.1086/321541)
- Bonaca, A., & Hogg, D. W. 2018, *ApJ*, 867, 101, doi: [10.3847/1538-4357/aae4da](https://doi.org/10.3847/1538-4357/aae4da)
- Bonaca, A., Hogg, D. W., Price-Whelan, A. M., & Conroy, C. 2019, *ApJ*, 880, 38, doi: [10.3847/1538-4357/ab2873](https://doi.org/10.3847/1538-4357/ab2873)
- Bond, J. R., & Szalay, A. S. 1983, *ApJ*, 274, 443, doi: [10.1086/161460](https://doi.org/10.1086/161460)
- Boyanovsky, D., de Vega, H. J., & Sanchez, N. G. 2008, *PhRvD*, 78, 063546, doi: [10.1103/PhysRevD.78.063546](https://doi.org/10.1103/PhysRevD.78.063546)
- Boyanovsky, D., & Wu, J. 2011, *PhRvD*, 83, 043524, doi: [10.1103/PhysRevD.83.043524](https://doi.org/10.1103/PhysRevD.83.043524)
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. 2018a, *PhRvL*, 121, 111801, doi: [10.1103/PhysRevLett.121.111801](https://doi.org/10.1103/PhysRevLett.121.111801)
- . 2018b, *PhRvD*, 98, 052004, doi: [10.1103/PhysRevD.98.052004](https://doi.org/10.1103/PhysRevD.98.052004)
- Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. 2018c, *arXiv e-prints*, arXiv:1805.12244, <https://arxiv.org/abs/1805.12244>
- Brennan, S., Benson, A. J., Cyr-Racine, F.-Y., et al. 2019, *MNRAS*, 2044, doi: [10.1093/mnras/stz1607](https://doi.org/10.1093/mnras/stz1607)
- Brewer, B. J., Huijser, D., & Lewis, G. F. 2016, *MNRAS*, 455, 1819, doi: [10.1093/mnras/stv2370](https://doi.org/10.1093/mnras/stv2370)
- Brooks, A. M. 2018, *arXiv e-prints*, arXiv:1812.00044, <https://arxiv.org/abs/1812.00044>
- Buckley, M. R., & DiFranzo, A. 2018, *PhRvL*, 120, 051102, doi: [10.1103/PhysRevLett.120.051102](https://doi.org/10.1103/PhysRevLett.120.051102)
- Buschmann, M., Kopp, J., Safdi, B. R., & Wu, C.-L. 2018, *PhRvL*, 120, 211101, doi: [10.1103/PhysRevLett.120.211101](https://doi.org/10.1103/PhysRevLett.120.211101)
- Carlberg, R. G. 2012, *ApJ*, 748, 20, doi: [10.1088/0004-637X/748/1/20](https://doi.org/10.1088/0004-637X/748/1/20)
- Carlberg, R. G., & Grillmair, C. J. 2013, *ApJ*, 768, 171, doi: [10.1088/0004-637X/768/2/171](https://doi.org/10.1088/0004-637X/768/2/171)
- Chang, L. J., Lisanti, M., & Mishra-Sharma, S. 2018, *PhRvD*, 98, 123004, doi: [10.1103/PhysRevD.98.123004](https://doi.org/10.1103/PhysRevD.98.123004)
- Chatterjee, S., & Koopmans, L. V. E. 2018, *MNRAS*, 474, 1762, doi: [10.1093/mnras/stx2674](https://doi.org/10.1093/mnras/stx2674)
- Ciotti, L., & Bertin, G. 1999, *A&A*, 352, 447
- Collett, T. E. 2015, *ApJ*, 811, 20, doi: [10.1088/0004-637X/811/1/20](https://doi.org/10.1088/0004-637X/811/1/20)
- Cui, X., Abdurkirim, A., Chen, W., et al. 2017, *PhRvL*, 119, 181302, doi: [10.1103/PhysRevLett.119.181302](https://doi.org/10.1103/PhysRevLett.119.181302)
- Cyr-Racine, F.-Y., Keeton, C. R., & Moustakas, L. A. 2019, *PhRvD*, 100, 023013, doi: [10.1103/PhysRevD.100.023013](https://doi.org/10.1103/PhysRevD.100.023013)

- Cyr-Racine, F.-Y., Moustakas, L. A., Keeton, C. R., Sigurdson, K., & Gilman, D. A. 2016, *PhRvD*, 94, 043505, doi: [10.1103/PhysRevD.94.043505](https://doi.org/10.1103/PhysRevD.94.043505)
- Dalal, N., & Kochanek, C. S. 2002, *ApJ*, 572, 25, doi: [10.1086/340303](https://doi.org/10.1086/340303)
- Dalcanton, J. J., & Hogan, C. J. 2001, *ApJ*, 561, 35, doi: [10.1086/323207](https://doi.org/10.1086/323207)
- Daylan, T., Cyr-Racine, F.-Y., Diaz Rivero, A., Dvorkin, C., & Finkbeiner, D. P. 2018, *ApJ*, 854, 141, doi: [10.3847/1538-4357/aaa1e](https://doi.org/10.3847/1538-4357/aaa1e)
- Despali, G., & Vegetti, S. 2017, *MNRAS*, 469, 1997, doi: [10.1093/mnras/stx966](https://doi.org/10.1093/mnras/stx966)
- Diaz Rivero, A., Cyr-Racine, F.-Y., & Dvorkin, C. 2018, *PhRvD*, 97, 023001, doi: [10.1103/PhysRevD.97.023001](https://doi.org/10.1103/PhysRevD.97.023001)
- Díaz Rivero, A., Dvorkin, C., Cyr-Racine, F.-Y., Zavala, J., & Vogelsberger, M. 2018, *PhRvD*, 98, 103517, doi: [10.1103/PhysRevD.98.103517](https://doi.org/10.1103/PhysRevD.98.103517)
- Drlica-Wagner, A., Mao, Y.-Y., Adhikari, S., et al. 2019, arXiv e-prints, arXiv:1902.01055. <https://arxiv.org/abs/1902.01055>
- Efstathiou, G. 1992, *MNRAS*, 256, 43P, doi: [10.1093/mnras/256.1.43P](https://doi.org/10.1093/mnras/256.1.43P)
- Errani, R., Peñarrubia, J., Laporte, C. F. P., & Gómez, F. A. 2017, *MNRAS*, 465, L59, doi: [10.1093/mnrasl/slz211](https://doi.org/10.1093/mnrasl/slz211)
- Fan, J., Katz, A., Randall, L., & Reece, M. 2013, *Physics of the Dark Universe*, 2, 139, doi: [10.1016/j.dark.2013.07.001](https://doi.org/10.1016/j.dark.2013.07.001)
- Fitts, A., Boylan-Kolchin, M., Elbert, O. D., et al. 2017, *MNRAS*, 471, 3547, doi: [10.1093/mnras/stx1757](https://doi.org/10.1093/mnras/stx1757)
- Fitts, A., Boylan-Kolchin, M., Bozek, B., et al. 2018, arXiv e-prints, arXiv:1811.11791. <https://arxiv.org/abs/1811.11791>
- Garrison-Kimmel, S., Wetzel, A., Bullock, J. S., et al. 2017, *MNRAS*, 471, 1709, doi: [10.1093/mnras/stx1710](https://doi.org/10.1093/mnras/stx1710)
- Han, J., Cole, S., Frenk, C. S., & Jing, Y. 2016, *MNRAS*, 457, 1208, doi: [10.1093/mnras/stv2900](https://doi.org/10.1093/mnras/stv2900)
- Hezaveh, Y., Dalal, N., Holder, G., et al. 2016a, *JCAP*, 2016, 048, doi: [10.1088/1475-7516/2016/11/048](https://doi.org/10.1088/1475-7516/2016/11/048)
- Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016b, *ApJ*, 823, 37, doi: [10.3847/0004-637X/823/1/37](https://doi.org/10.3847/0004-637X/823/1/37)
- Hiroshima, N., Ando, S., & Ishiyama, T. 2018, *PhRvD*, 97, 123002, doi: [10.1103/PhysRevD.97.123002](https://doi.org/10.1103/PhysRevD.97.123002)
- Hsueh, J.-W., Enzi, W., Vegetti, S., et al. 2019, arXiv e-prints, arXiv:1905.04182. <https://arxiv.org/abs/1905.04182>
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Johnston, K. V., Zhao, H., Spergel, D. N., & Hernquist, L. 1999, *ApJL*, 512, L109, doi: [10.1086/311876](https://doi.org/10.1086/311876)
- Jones, E., Oliphant, T., Peterson, P., et al. 2001–, *SciPy: Open source scientific tools for Python*. <http://www.scipy.org/>
- Kahlhoefer, F., Kaplinghat, M., Slatyer, T. R., & Wu, C.-L. 2019, arXiv e-prints, arXiv:1904.10539. <https://arxiv.org/abs/1904.10539>
- Kaplinghat, M., Keeley, R. E., Linden, T., & Yu, H.-B. 2014, *PhRvL*, 113, 021302, doi: [10.1103/PhysRevLett.113.021302](https://doi.org/10.1103/PhysRevLett.113.021302)
- Kaplinghat, M., Tulin, S., & Yu, H.-B. 2016, *PhRvL*, 116, 041302, doi: [10.1103/PhysRevLett.116.041302](https://doi.org/10.1103/PhysRevLett.116.041302)
- Keeton, C. R. 2001, arXiv e-prints, astro. <https://arxiv.org/abs/astro-ph/0102341>
- Koposov, S., Belokurov, V., Evans, N. W., et al. 2008, *ApJ*, 686, 279, doi: [10.1086/589911](https://doi.org/10.1086/589911)
- Koposov, S. E., Belokurov, V., Torrealba, G., & Evans, N. W. 2015, *ApJ*, 805, 130, doi: [10.1088/0004-637X/805/2/130](https://doi.org/10.1088/0004-637X/805/2/130)
- Lisanti, M., Mishra-Sharma, S., Rodd, N. L., & Safdi, B. R. 2018, *PhRvL*, 120, 101101, doi: [10.1103/PhysRevLett.120.101101](https://doi.org/10.1103/PhysRevLett.120.101101)
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201. <https://arxiv.org/abs/0912.0201>
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563, doi: [10.1086/177173](https://doi.org/10.1086/177173)
- . 1997, *ApJ*, 490, 493, doi: [10.1086/304888](https://doi.org/10.1086/304888)
- Neyman, J., & Pearson, E. S. 1933, *Philosophical Transactions of the Royal Society of London Series A*, 231, 289, doi: [10.1098/rsta.1933.0009](https://doi.org/10.1098/rsta.1933.0009)
- Paszke, A., Gross, S., Chintala, S., et al. 2017, in *NIPS-W*
- Perez, F., & Granger, B. E. 2007, *Computing in Science and Engineering*, 9, 21, doi: [10.1109/MCSE.2007.53](https://doi.org/10.1109/MCSE.2007.53)
- Peter, A. H. G., Rocha, M., Bullock, J. S., & Kaplinghat, M. 2013, *MNRAS*, 430, 105, doi: [10.1093/mnras/sts535](https://doi.org/10.1093/mnras/sts535)
- Read, J. I., Iorio, G., Agertz, O., & Fraternali, F. 2017, *MNRAS*, 467, 2019, doi: [10.1093/mnras/stx147](https://doi.org/10.1093/mnras/stx147)
- Refregier, A., Amara, A., Kitching, T. D., et al. 2010, arXiv e-prints, arXiv:1001.0061. <https://arxiv.org/abs/1001.0061>
- Sánchez-Conde, M. A., & Prada, F. 2014, *MNRAS*, 442, 2271, doi: [10.1093/mnras/stu1014](https://doi.org/10.1093/mnras/stu1014)
- Schneider, P., Ehlers, J., & Falco, E. E. 1992, *Gravitational Lenses*, doi: [10.1007/978-3-662-03758-4](https://doi.org/10.1007/978-3-662-03758-4)
- Sirunyan, A. M., Tumasyan, A., Adam, W., et al. 2017, *Physics Letters B*, 769, 520, doi: [10.1016/j.physletb.2017.02.012](https://doi.org/10.1016/j.physletb.2017.02.012)
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, 13, 22, doi: [10.1109/MCSE.2011.37](https://doi.org/10.1109/MCSE.2011.37)

- Van Tilburg, K., Taki, A.-M., & Weiner, N. 2018, JCAP, 2018, 041, doi: [10.1088/1475-7516/2018/07/041](https://doi.org/10.1088/1475-7516/2018/07/041)
- Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969, doi: [10.1111/j.1365-2966.2010.16865.x](https://doi.org/10.1111/j.1365-2966.2010.16865.x)
- Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341, doi: [10.1038/nature10669](https://doi.org/10.1038/nature10669)
- Verma, A., Collett, T., Smith, G. P., Strong Lensing Science Collaboration, & the DESC Strong Lensing Science Working Group. 2019, arXiv e-prints, arXiv:1902.05141. <https://arxiv.org/abs/1902.05141>
- Vogelsberger, M., Zavala, J., Cyr-Racine, F.-Y., et al. 2016, MNRAS, 460, 1399, doi: [10.1093/mnras/stw1076](https://doi.org/10.1093/mnras/stw1076)
- Vogelsberger, M., Zavala, J., & Loeb, A. 2012, MNRAS, 423, 3740, doi: [10.1111/j.1365-2966.2012.21182.x](https://doi.org/10.1111/j.1365-2966.2012.21182.x)
- Vogelsberger, M., Zavala, J., Schutz, K., & Slatyer, T. R. 2019, MNRAS, 484, 5437, doi: [10.1093/mnras/stz340](https://doi.org/10.1093/mnras/stz340)
- Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435, doi: [10.1146/annurev-astro-081817-051756](https://doi.org/10.1146/annurev-astro-081817-051756)
- Zahid, H. J., Sohn, J., & Geller, M. J. 2018, ApJ, 859, 96, doi: [10.3847/1538-4357/aabe31](https://doi.org/10.3847/1538-4357/aabe31)
- Zavala, J., Vogelsberger, M., & Walker, M. G. 2013, MNRAS, 431, L20, doi: [10.1093/mnrasl/sls053](https://doi.org/10.1093/mnrasl/sls053)