

# Simulation-based marginal likelihood for cluster strong lensing cosmology

M. Killedar,<sup>1,2,3★</sup> S. Borgani,<sup>3,4,5</sup> D. Fabjan,<sup>4,6</sup> K. Dolag,<sup>2,7</sup> G. Granato,<sup>3,4</sup>  
M. Meneghetti,<sup>8,9,10</sup> S. Planelles<sup>3,4,5,11</sup> and C. Ragone-Figueroa<sup>4,12</sup>

<sup>1</sup>Centre for Translational Data Science, The University of Sydney, NSW 2006, Australia

<sup>2</sup>Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany

<sup>3</sup>Dipartimento di Fisica dell'Università di Trieste, Sezione di Astronomia, Via Tiepolo 11, I-34131 Trieste, Italy

<sup>4</sup>INAF – Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34131 Trieste, Italy

<sup>5</sup>INFN – National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy

<sup>6</sup>Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

<sup>7</sup>Max-Planck-Institut für Astrophysik, D-85748 Garching, Germany

<sup>8</sup>INAF – Osservatorio Astronomico di Bologna, Via Ranzani 1, I-40127 Bologna, Italy

<sup>9</sup>INFN – Sezione di Bologna, Viale Berti Pichat 6/2, I-40127 Bologna, Italy

<sup>10</sup>Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

<sup>11</sup>Departament d'Astronomia i Astrofísica, Universitat de València, E-46100 Burjassot (Valencia), Spain

<sup>12</sup>Instituto de Astronomía Teórica y Experimental (IATE), Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET), Observatorio Astronómico, Universidad Nacional de Córdoba, Laprida 854, X5000BGR Córdoba, Argentina

Accepted 2017 August 31. Received 2017 August 31; in original form 2015 July 17

## ABSTRACT

Comparisons between observed and predicted strong lensing properties of galaxy clusters have been routinely used to claim either tension or consistency with  $\Lambda$  cold dark matter cosmology. However, standard approaches to such cosmological tests are unable to quantify the preference for one cosmology over another. We advocate approximating the relevant Bayes factor using a marginal likelihood that is based on the following summary statistic: the posterior probability distribution function for the parameters of the scaling relation between Einstein radii and cluster mass,  $\alpha$  and  $\beta$ . We demonstrate, for the first time, a method of estimating the marginal likelihood using the X-ray selected  $z > 0.5$  Massive Cluster Survey clusters as a case in point and employing both  $N$ -body and hydrodynamic simulations of clusters. We investigate the uncertainty in this estimate and consequential ability to compare competing cosmologies, which arises from incomplete descriptions of baryonic processes, discrepancies in cluster selection criteria, redshift distribution and dynamical state. The relation between triaxial cluster masses at various overdensities provides a promising alternative to the strong lensing test.

**Key words:** gravitational lensing: strong – methods: numerical – methods: statistical – galaxies: clusters: general – cosmology: theory.

## 1 INTRODUCTION

The matter density parameter,  $\Omega_M$ , the vacuum density parameter,  $\Omega_\Lambda$ , the normalization of the matter power spectrum,  $\sigma_8$ , and the slope,  $n$ , of the power spectrum for the primordial density fluctuations ( $P(k) \propto k^n$ ) have a strong influence on the redshift at which clusters form and the amount of time they are given to evolve until we observe them (e.g. Kravtsov & Borgani 2012, for a review). For clusters of a fixed mass at the time of observation, lower values of  $\Omega_M$ ,  $\Omega_\Lambda$  or  $\sigma_8$  or higher values of the spectral index,  $n$ , would require the host haloes to

have formed earlier, and subsequently lead to higher concentrations (Cole & Lacey 1996; Navarro, Frenk & White 1997; Tormen, Bouchet & White 1997; van den Bosch 2002; Wechsler et al. 2002). Galaxy clusters gravitationally lens and distort the images of background galaxies; their lensing efficiency is a powerful probe of cosmology with the ability to constrain the aforementioned structure formation parameters (Bartelmann et al. 1998, 2003; Takahashi & Chiba 2001; Wambsganss, Bode & Ostriker 2004; Boldrin et al. 2016). This is partly because the cosmological model determines the formation history of clusters, but also because the critical surface mass density for lensing, a function of the angular diameter distances between observer, lens and source(s), is also dependent on these cosmological parameters. However, cosmological distances play a secondary role compared to the mass

\*E-mail: [madhura.killedar.astro@gmail.com](mailto:madhura.killedar.astro@gmail.com)

distribution of clusters (Wu & Hammer 1993; Hattori, Watanabe & Yamashita 1997; Oguri, Taruya & Suto 2001). The earliest comparisons between simulated clusters and the observed frequency of arc-like lensed galaxy images in a cluster sample revealed an order of magnitude difference between the observations and  $\Lambda$  cold dark matter ( $\Lambda$ CDM) predictions (Bartelmann et al. 1998; Li et al. 2005). This discrepancy, dubbed the ‘arc-statistics problem’, has the potential to be a point of tension for the standard  $\Lambda$ CDM model (see Meneghetti et al. 2013, for an overview). As such, many efforts have been made to provide explanations, beginning with discussions about the appropriate modelling of shape and redshift distribution of the background source galaxies (Bartelmann, Steinmetz & Weiss 1995; Hattori et al. 1997; Wambsganss, Bode & Ostriker 2004; Ho & White 2005; Li et al. 2005; Bayliss 2012), complex structure in the lensing mass (e.g. Bayliss et al. 2014) and the nature of dark matter (DM; e.g. Mahdi et al. 2014). Cluster selection criteria are another complicating factor (Serenio et al. 2015). Strong-lensing selection will sample the high-mass end of the cluster mass function (Comerford & Natarajan 2007) and preferentially target clusters aligned along the major axes (Miralda-Escude 1993; Hennawi et al. 2007; Oguri & Blandford 2009; Meneghetti et al. 2010). X-ray selection tends to create a sample including more merging clusters with complex morphology (Ritchie & Thomas 2002; Planelles & Quilis 2009), and yet high concentration (Rasia et al. 2013) and a higher fraction of cool-core systems (Jones et al., in preparation).

Comparisons between simulated and observed clusters have been conducted at a range of cluster lens redshifts. Dalal, Holder & Hennawi (2004) found that arc statistics associated with low-redshift cluster lenses are consistent with observations, although Horesh et al. (2011) maintain that observed number counts are higher than expected for clusters at  $z \lesssim 0.2$ . All studies so far have found that the discrepancy remains at high redshift ( $z \gtrsim 0.6$ ) for the most massive clusters or is unclear due to small number statistics for high-mass simulated clusters (Dalal et al. 2004; Horesh et al. 2011; Meneghetti et al. 2011). We improve upon these works by increasing the simulated sample size somewhat. However, given the sensitivity of arc-statistics to the assumed properties of background sources and the lack of detailed observations required for complete mass profile models, we compromise by using Einstein radii as a proxy for cluster mass concentration.

The Massive Cluster Survey (MACS) is one of few samples that have a simple, well-defined selection function, high completeness and ample data for strong lensing analysis. There are also updates to some of the clusters’ mass models care of the Cluster Lensing and Supernova survey with *Hubble* (CLASH; Postman et al. 2012; Zheng et al. 2012; Coe et al. 2013; Zitrin et al. 2015) and ongoing Frontier Fields programmes<sup>1</sup> (Johnson et al. 2014). The X-ray selected  $z > 0.5$  MACS cluster sample (Ebeling et al. 2007) has posed such a challenge due to large measured Einstein radii. However there are disagreements throughout the literature due the differing theoretical models and statistical methods. In this work, we propose a Bayesian approach to the cosmological test using strong lensing properties of this sample, and clusters modelled within hydrodynamic simulations.

Concerns about the lack of baryonic heating and cooling mechanisms in early cosmological simulations led to a number of studies examining the effect of these processes on cluster density profiles and lensing efficiency (Lewis et al. 2000; Puchwein et al. 2005; Rozo

et al. 2008; Duffy et al. 2010; Mead et al. 2010; Cui et al. 2012; Killedar et al. 2012; Rasia et al. 2013). Together, they paint a complex picture of numerous counteracting baryonic effects. ‘Run-away’ cooling flows in simulations have been found to steepen profiles and produce stronger lenses (Puchwein et al. 2005; Rozo et al. 2008). Yet, there is no significant link between intracluster medium (ICM) cooling signatures and lensing efficiency in strong lensing selected clusters (Blanchard et al. 2013). This apparent discrepancy is resolved with the additional component of feedback mechanisms that temper the overproduction of stellar mass while simultaneously reducing the strong lensing efficiency (Sijacki et al. 2007; Fabjan et al. 2010; McCarthy et al. 2010; Mead et al. 2010; Teysier et al. 2011; Killedar et al. 2012). In Killedar et al. (2012), we found that the quantitative difference between cluster lensing properties in  $N$ -body and hydrodynamic simulations depends mildly on the redshift of the cluster lenses, and the results suggested that for relaxed clusters, the inclusion of baryons would not affect lensing efficiencies characterized by Einstein radii or tangential arcs. However, cluster selection criteria and characterization of strong lensing efficiency affect the result of the comparison. Exploring in detail such effects is included in the analysis that we present here. We are not restricted to relaxed clusters in this sample, although we investigate the consequences of dynamical selection.

The paper is structured as follows: in Section 2 the basic theory and notation of gravitational lensing are introduced, along with the characterization of the Einstein radius; in Section 3 the MACS high- $z$  sample is described; the main details of the hydrodynamic simulations are briefly outlined in Section 4; in Section 5 we motivate a Bayesian approach and describe a method by which one estimates the marginal likelihood; in Section 6 we demonstrate and discuss selection effects and uncertainties with regards to the modelling of baryonic processes; we discuss how our findings differ from previous strong lensing studies of the MACS clusters in Section 7; in Section 8 we consider the potential of triaxial mass models to provide an alternative scaling relation; we finally summarize our findings in Section 9.

Throughout this work, the following values for cosmological parameters are adopted: present day vacuum density parameter,  $\Omega_{\Lambda,0} = 0.76$ ; matter density parameter,  $\Omega_{M,0} = 0.24$ ; baryon density parameter,  $\Omega_{b,0} = 0.04$ ; Hubble constant  $h = 0.72$ ; normalization of the matter power spectrum  $\sigma_8 = 0.8$  and primordial power spectrum  $P(k) \propto k^n$  with  $n = 0.96$ . Furthermore, characteristic overdensities at which cluster masses are provided are assumed to refer to the critical cosmic density,  $\rho_c = 3H^2/(8\pi G)$ .

## 2 STRONG LENSING EFFICIENCY

Throughout this work, we refer to gravitational lensing quantities following the notation of Schneider, Ehlers & Falco (1992), and assuming the thin lens approximation. Note that from here onwards, the redshift of the background source galaxies is denoted  $z_s$ . Since the strong lensing properties of the observational sample are usually determined for galaxies at the fixed source redshift of  $z_s = 2$ , we derive results for the same unique source redshift throughout this work.

### 2.1 Gravitational lensing

Images of a source at  $z_s$  are highly magnified when they appear on a locus known as the *critical curve*, where the Jacobian of the lens mapping formally diverges.

<sup>1</sup> <http://www.stsci.edu/hst/campaigns/frontier-fields/>

**Table 1.** Properties of MACS  $z > 0.5$  cluster sample. Column 2: redshift. Column 3: morphological code (see Ebeling et al. 2007). Columns 4 and 5: *Chandra* fluxes and luminosities in the 0.1–2.4 keV band, quoted from Ebeling et al. (2007). Column 6: masses within  $R_{500}$  assuming sphericity and hydrostatic equilibrium are cited from Mantz et al. (2010); 2009 January 21 *Chandra* calibration update fluxes mean that these are overestimated by 7–15 per cent. Column 7: effective Einstein radius at  $z_s = 2$  (see text for details and references).

MACS	$z$	Morph.	$f_X$ ( $10^{-12}$ erg s $^{-1}$ cm $^{-2}$ )	$L_X$ ( $10^{44}$ erg s $^{-1}$ )	$M_{500}$ ( $10^{14} M_\odot$ )	$\theta_{E, \text{eff}}$ (arcsec)
J0018.5+1626	0.5456	3	$2.14 \pm 0.03$	$19.6 \pm 0.3$	$16.5 \pm 2.5$	24
J0025.4–1222	0.5843	3	$0.81 \pm 0.02$	$8.8 \pm 0.2$	$7.6 \pm 0.9$	30
J0257.1–2325	0.5049	2	$1.80 \pm 0.03$	$13.7 \pm 0.3$	$8.5 \pm 1.3$	39
J0454.1–0300	0.5377	2	$1.88 \pm 0.04$	$16.8 \pm 0.6$	$11.5 \pm 1.5$	13
J0647.7+7015	0.5907	2	$1.49 \pm 0.03$	$15.9 \pm 0.4$	$10.9 \pm 1.6$	28
J0717.5+3745	0.5458	4	$2.74 \pm 0.03$	$24.6 \pm 0.3$	$15.0 \pm 2.85$	50.1
J0744.8+3927	0.6976	2	$1.44 \pm 0.03$	$22.9 \pm 0.6$	$11.9 \pm 2.8$	23.3
J0911.2+1746	0.5049	4	$1.00 \pm 0.02$	$7.8 \pm 0.3$	$9.0 \pm 1.2$	11
J1149.5+2223	0.5444	4	$1.95 \pm 0.04$	$17.6 \pm 0.4$	$14.6 \pm 3.1$	21.5
J1423.8+2404	0.5431	1	$1.80 \pm 0.06$	$16.5 \pm 0.7$	$6.6 \pm 0.9$	17.8
J2129.4–0741	0.5889	3	$1.45 \pm 0.03$	$15.7 \pm 0.4$	$10.6 \pm 1.4$	21.8
J2214.9–1359	0.5027	2	$1.85 \pm 0.03$	$14.1 \pm 0.3$	$13.2 \pm 2.3$	23

The angular separation of highly magnified and tangentially sheared background galaxies has a formal definition, which is strictly applicable only in the case of axially symmetric lenses. This separation is defined by the Einstein radius:

$$\theta_E = \sqrt{\frac{4GM}{c^2} \frac{D_{ds}}{D_d D_s}}, \quad (1)$$

where  $M$  denotes the enclosed mass and  $D_s$ ,  $D_d$  and  $D_{ds}$  are the angular diameter distances from the observer to the source, from the observer to the lens and from the lens to the source, respectively. However, galaxy clusters are not axially symmetric in general, and so critical curves are not circular. As such, the typical scale length may be characterized by the so-called ‘effective Einstein radius’,  $\theta_{E, \text{eff}}$ , according to

$$A = \pi \theta_{E, \text{eff}}^2, \quad (2)$$

where  $A$  is the area enclosed within the tangential critical curve. This is the definition used, for example, by Puchwein & Hilbert (2009), Zitrin et al. (2011) and Redlich et al. (2012), and is implemented throughout this work (however, see Section 7).

### 3 THE $z > 0.5$ MACS SAMPLE

MACS consists of the most X-ray luminous clusters, from which a 90 per cent complete sample of high-redshift clusters ( $z > 0.5$ ) was presented in Ebeling et al. (2007). Ultimately the clusters are chosen by following a flux limit and redshift criterion, and so should not suffer from the lensing-selection bias. For the *ROSAT* All-Sky Survey Bright Source Catalogue (RASS-BSC) the flux limit is  $f_X > 1 \times 10^{-12}$  erg s $^{-1}$  cm $^{-2}$  in the 0.1–2.4 keV band. However, follow-up observations with *Chandra* found that the lowest flux cluster had  $f_X = 0.8 \times 10^{-12}$  erg s $^{-1}$  cm $^{-2}$  in the same band; given the higher sensitivity of *Chandra*, we use this latter flux limit for our simulated cluster selection (see Section 6.1). Masses within  $R_{500}$  assuming sphericity and hydrostatic equilibrium were derived by Mantz et al. (2010). Changes to the flux measurements from the 2009 January 21 *Chandra* calibration update mean that these masses are likely overestimated by 7–15 per cent, therefore throughout this work, we reduce the masses by 10 per cent as a rough guide, and estimate a 10 per cent uncertainty. Lensing data from the CLASH programme have led to new mass

estimates for MACS J0717.5+3745, MACS J1149.5+2223, MACS J0744.8+3927 and MACS J0647.7+7015 (Umetsu et al. 2016).

The morphological codes listed in Ebeling et al. (2007) suggest that half the clusters are unrelaxed (see further discussion in Section 6.5).

The effective Einstein radii were originally presented in Zitrin et al. (2011). However a number of these have been revised with new imaging and spectroscopic data, as well as mass reconstruction techniques. MACS J0647.7+7015 has been re-analysed by Coe et al. (2013) and MACS J2129.4–0741, MACS J0744.8+3927 and MACS J1423.8+2404 by Zitrin et al. (2015) using *Hubble Space Telescope* (*HST*) data collected as part of the CLASH programme. New models for MACS J1149.5+2223 and MACS J0717.5+3745 have been presented by Johnson et al. (2014) using data collected from the Frontier Fields programme. The Einstein radii at  $z_s = 2$  are found to be 21.5 and 50.1 arcsec, respectively. In the case of the ‘baby bullet’ cluster MACS J0025.4–1222, the secondary critical curve was included in the analysis, despite being slightly separated from the primary critical curve. However in any case, simulated clusters undergoing a merger will show a large variation in strong lensing measurements depending on the projection; secondary critical curves will often be connected to the primary. In the absence of more complete information, we estimate a 1 per cent Gaussian uncertainty on the Einstein radii of MACS J1149.5+2223 and J0717.5+3745 (Johnson et al. 2014), and 10 per cent uncertainty on all the other MACS high- $z$  clusters (Zitrin, private communication). All relevant properties of the high- $z$  MACS clusters are listed in Table 1.

### 4 COSMOLOGICAL SIMULATIONS

The simulations analysed here are the same as described in Ragone-Figueroa et al. (2013) and Planelles et al. (2014). In the following, we provide a short overview, while we refer to the above papers for a comprehensive description.

#### 4.1 The set of simulated clusters

Simulations have been carried out using the TreePM–smoothed particle hydrodynamics (SPH) *GADGET-3* code, a newer version of

the original GADGET-2 code by Springel (2005) that adopted a more efficient domain decomposition to improve the work-load balance. A flat  $\Lambda$ CDM model whose cosmological parameters were chosen as follows: present day vacuum density parameter,  $\Omega_{\Lambda,0} = 0.76$ ; matter density parameter,  $\Omega_{M,0} = 0.24$ ; baryon density parameter,  $\Omega_{b,0} = 0.04$ ; Hubble constant  $h = 0.72$ ; normalization of the matter power spectrum  $\sigma_8 = 0.8$  and primordial power spectrum  $P(k) \propto k^n$  with  $n = 0.96$ .

Starting from a low-resolution cosmological box having size of  $1 h^{-1}$  Gpc, we selected 24 Lagrangian regions surrounding the most massive clusters identified at  $z = 0$ , all having virial mass of at least  $10^{15} h^{-1} M_{\odot}$ , plus further five Lagrangian regions surrounding clusters in the mass range  $(1-5) \times 10^{14} h^{-1} M_{\odot}$  (see Bonafede et al. 2011 for details). Initial conditions are then generated by increasing mass resolution, and correspondingly adding higher frequency modes to the density fluctuation field, within these regions. Resolution is progressively degraded outside these regions, so as to save computational time while still providing a correct description of the large-scale tidal field. The Lagrangian regions were large enough to ensure that only high-resolution particles are present within five virial radii of the central cluster.

Each Lagrangian region has been simulated in four different flavours: including only DM particles (*DM*); with non-radiative hydrodynamics (*NR*); including cooling star formation and supernova (SN) feedback (*CSF*) and further including active galactic nuclei (AGN) feedback (*AGN*).

The basic characteristics of these resimulation sets are described here below.

*DM.* Simulations including only DM particles that in the high-resolution region have a mass  $m_{\text{DM}} = 10^9 h^{-1} M_{\odot}$ . The Plummer-equivalent comoving softening length for gravitational force in the high-resolution region is fixed to  $\epsilon_{\text{pl}} = 5 h^{-1}$  kpc physical at  $z < 2$  while being fixed to  $\epsilon_{\text{pl}} = 15 h^{-1}$  kpc comoving at higher redshift.

*NR.* Non-radiative hydrodynamical simulations. Initial conditions for these hydrodynamical simulations are generated starting from those of the DM-only simulations, and splitting each particles in the high-resolution region into one DM and one gas particle, with their masses chosen so as to reproduce the assumed cosmic baryon fraction. The mass of each DM particle is then  $m_{\text{DM}} = 8.47 \times 10^8 h^{-1} M_{\odot}$  and the mass of each gas particle is  $m_{\text{gas}} = 1.53 \times 10^8 h^{-1} M_{\odot}$ . For the computation of the hydrodynamical forces we assume the minimum value attainable by the SPH smoothing length of the B-spline interpolating kernel to be half of the corresponding value of the gravitational softening length. No radiative cooling is included.

*CSF.* Hydrodynamical simulations including the effect of cooling, star formation, chemical enrichment and SN feedback. Star formation is described through the effective model by Springel & Hernquist (2003). The effect of SN feedback is included by using galactic wind having a velocity of  $500 \text{ km s}^{-1}$ . Chemical enrichment is described as in Tornatore et al. (2007) and includes the contributions from Type Ia and Type II SN, and of asymptotic giant branch (AGB) stars.

*AGN.* The same as CSF, but with the additional effect of AGN feedback. In the model for AGN feedback, released energy results from gas accretion on to supermassive black holes (SMBHs) that are initially seeded within resolved DM haloes and later grow by gas accretion and merging with other BHs. The description of BH accretion and AGN feedback used in our simulations is largely inspired by that originally presented by Springel, Di Matteo & Hernquist (2005a), with a number of modifications, whose details

and motivation are explained in Ragone-Figueroa et al. (2013); see also Planelles et al. (2014).

## 4.2 Properties of the simulated clusters

The clusters within the simulated regions are identified as follows. First, a standard friends-of-friends (FoF) algorithm is run over the DM particles in the high-resolution regions, using a linking length of 0.16 in units of the mean interparticle separation. Within each FoF group, we identify the position of the particle with the minimum gravitational potential, which is then taken as the centre from where clusters are then identified according to a spherical overdensity (SO) method. The mass,  $M_{500}$ , of each cluster is defined as the mass enclosed within the radius,  $R_{500}$ , at which the average density is 500 times the critical overdensity.

Throughout Section 6, we provide the results of analyses of simulated clusters chosen either by mass or X-ray luminosity within the 0.1–2.4 keV energy band. Since we aim to perform a self-consistent cluster selection with respect to the high- $z$  MACS cluster selection criteria, we estimate the X-ray luminosity for simulated clusters within the *AGN* simulations, which are those producing a relation between X-ray luminosity and mass consistent with observational results (Planelles et al. 2014). The X-ray luminosity is computed by summing the contributions to the emissivity,  $\epsilon_i$ , carried by all the gas particles within  $R_{500}$ :

$$L_X = \sum_i \epsilon_i = \sum_i n_{e,i} n_{H,i} \Lambda(T_i, Z_i) dV_i, \quad (3)$$

where  $n_{e,i}$  and  $n_{H,i}$  are the number densities of electrons and of hydrogen atoms, respectively, associated with the  $i$ th gas element of given density  $\rho_i$ , temperature  $T_i$ , mass  $m_i$ , metallicity  $Z_i$  and volume  $dV_i = m_i / \rho_i$ . Furthermore,  $\Lambda(T, Z)$  is the temperature- and metallicity-dependent cooling function computed within the [0.1–2.4] keV energy band.

## 4.3 Measuring Einstein radii

The lensing mass includes all matter within two virial radii<sup>2</sup> of the cluster centre. Using the Fourier techniques outlined in Killedar et al. (2012), we determine the positions of critical points for a projected lens. First, the projection is centred on the peak in the two-dimensional surface map, which is likely to reside in the largest critical curve within this field. Tangential critical points are identified within a square field of view (comoving  $1.5 h^{-1}$  Mpc across) on a fine 2048-pixel grid (giving an angular resolution of 0.1 arcsec at  $z = 0.5$ ). As large substructures can also be present, we remove critical points associated with any distinct secondary critical curves, before measuring the Einstein radius. The effective Einstein radius is defined by equation (2) where  $A$  is the angular area enclosed by the polygon bounded by the remaining critical points.

Snapshots of the cosmological simulations are taken at fixed redshifts ( $z = 0.5$  and  $0.6$ ) from which we may select galaxy cluster-scale objects as a representative description of lenses as predicted by  $\Lambda$ CDM. However, the MACS cluster high- $z$  sample spans a range of redshifts. Thus quantities that describe strong lensing, which ultimately reflect the mass distribution in the inner regions of the

<sup>2</sup> The virial radius is defined as the smallest radius of a sphere centred on the cluster, for which the mean density falls below the virial overdensity. The virial overdensity is measured relative to the critical density and calculated using the fitting formula of Bryan & Norman (1998).

lens, should be scaled in a way that makes the quantities sensitive to that mass, but robust to offsets in redshift between the lenses being compared. Rearranging equation (1) we find

$$\theta_E \sqrt{\frac{D_d}{D_{ds}}} \propto \sqrt{M}, \quad (4)$$

which provides a rough scaling for a strong lensing quantity that scales with enclosed mass.

## 5 ESTIMATING THE MARGINAL LIKELIHOOD

Strong lensing efficiencies, as characterized by the Einstein radii, scale well with the mass of clusters at large overdensities (see Killedar et al. 2012). If the  $z > 0.5$  MACS sample are, in fact, stronger lenses than predicted by the  $\Lambda$ CDM model, they will have larger Einstein radii for a given total mass at low overdensities (or a proxy thereof).

A Bayesian approach is advocated (see e.g. Sivia 1996; Trotta 2008; Jenkins & Peacock 2011), in which one determines the relative preference of two hypothetical cosmological models,  $C_1$  and  $C_2$ , in light of the data  $D$ :

$$\frac{P(C_1|D)}{P(C_2|D)} = R \frac{P(C_1)}{P(C_2)}, \quad (5)$$

where  $P(C_1)/P(C_2)$  denotes the prior preference for  $C_1$  over  $C_2$ , perhaps due to previously available data sets, while the evidence ratio or Bayes factor,<sup>3</sup>  $R$ , is defined as

$$R = \frac{P(D|C_1)}{P(D|C_2)}, \quad (6)$$

where the marginal likelihood  $P(D|C_j)$  denotes the probability that one would observe data  $D$  assuming a cosmology  $C_j$ . A large Bayes factor  $R \gg 1$  reflects a shift in preference for  $C_1$  and vice versa. Performing comparisons for many cosmological models would require numerous simulations, each run under various cosmologies; this is outside the scope of the current work. In this work, the aim is to estimate the marginal likelihood: the probability of observing the Einstein radii of the high- $z$  MACS sample under a single chosen hypothesis:  $\Lambda$ CDM with aforementioned parameters. This is non-trivial because a likelihood function related to the original observables ( $\theta_E$  and  $M_{500}$ ) is intractable; the finite number of objects from the simulations mean that the full  $\theta_E$ - $M_{500}$  space cannot be sampled.

To circumvent this problem, we assume a generative probabilistic model with the form of a power-law relation between the strong lensing and mass proxies, and perform a fitting to the following function in logarithmic space:<sup>4</sup>

$$\log \left[ \frac{M_{500}}{9 \times 10^{14} M_\odot} \right] = \alpha \log \left[ \frac{\theta_E}{20 \text{ arcsec}} \sqrt{\frac{D_d}{D_{ds}}} \right] + \beta, \quad (7)$$

with parameters ( $\alpha$  and  $\beta$ ) and aim to find the probability of observing *the scaling relationship*. However, rather than calculating

<sup>3</sup> Cosmological simulations are run one set of parameters at a time, so in the current framework we are dealing with parameter estimation as opposed to full model selection, i.e.  $C_1$  and  $C_2$  differ only by the value of their parameters.

<sup>4</sup> The pivot mass  $9 \times 10^{14} M_\odot$  is chosen to approximate the logarithmic average of the observed and simulated clusters. Similarly the pivot Einstein radius is chosen to be 20 arcsec.

precise values for  $\alpha$  and  $\beta$ , one would determine a probability distribution,  $P(\alpha, \beta)$ , that reflects the degree of belief in their respective values. Thus  $P(\alpha, \beta|D)$  acts as a summary statistic for the data set. The relevant linear regression method, following Hogg, Bovy & Lang (2010), and the priors on  $\alpha$  and  $\beta$  are outlined in Appendix A.

Next, we outline how to estimate the marginal likelihood. In the following,  $\mathcal{I}$  represents background information such as knowledge of the cluster selection criteria, the method of characterizing the Einstein radius and the assumption that there exists a power-law relation between strong lensing and mass.

(i) Fit equation (7) to the data to obtain the posterior probability distribution,  $P(\alpha, \beta|D, \mathcal{I})$ , for  $\alpha$  and  $\beta$ .

(ii) Computer simulations are run within the framework of a chosen cosmological hypothesis,  $C_j$ . In our case,  $C_j$  represents the assumption that  $\Lambda$ CDM (with aforementioned values for cosmological parameters) is the true description of cosmology.

(iii) Simulated galaxy clusters are selected according to specified criteria, ideally reflecting the criteria used to select the real clusters. Their masses,  $M_{500}$ , are noted.

(iv) Different on-sky projections of these three-dimensional objects produce different apparent measurements of structural properties. Therefore, we construct a large number  $N$  of mock samples by randomly choosing an orientation angle and calculating  $\theta_E$  for each cluster. The fair sampling means that for each mock data set, denoted  $D_i$  for  $i$  from 1 to  $N$ , that  $P(D_i|C_j, \mathcal{I}) = 1/N$ .

(v) Equation (7) is fit to each mock sample to determine a posterior probability distribution,  $P(\alpha_i, \beta_i|D_i, \mathcal{I})$ , over  $\alpha_i$  and  $\beta_i$ .

(vi) Consider the proposal  $H_i := D_i = D$ , in other words, that the mock sample resembles the real data. The resemblance is judged using the summary statistic. We assign a weight,  $w_i$ , to each mock sample according to its similarity to the real data:

$$\begin{aligned} w_i &\equiv P(H_i|D, D_i, C_j, \mathcal{I}) \\ &= \int P(H_i|\alpha, \beta, D_i, \mathcal{I}) P(\alpha, \beta|D, \mathcal{I}) d\alpha d\beta \\ &\approx \int P(\alpha_i = \alpha, \beta_i = \beta|\alpha, \beta, D_i, \mathcal{I}) P(\alpha, \beta|D, \mathcal{I}) d\alpha d\beta, \end{aligned} \quad (8)$$

noting that  $D$  is redundant once  $\alpha$  and  $\beta$  are known. The weight is equal to the integral over the product of the two (mock and real) posterior probability distributions.<sup>5</sup>

(vii) Finally, we estimate the marginal likelihood:

$$\begin{aligned} \mathcal{Z}_j &= \hat{P}(D|C_j) \\ &= \sum_{i=1}^N P(D = D_i, D_i|C_j, \mathcal{I}) \\ &= \sum_{i=1}^N P(D = D_i|D_i, C_j, \mathcal{I}) P(D_i|C_j, \mathcal{I}) \\ &\approx \frac{1}{N} \sum_{i=1}^N P(H_i|D, D_i, C_j, \mathcal{I}) \\ &= \frac{1}{N} \sum_{i=1}^N w_i, \end{aligned} \quad (9)$$

noting that  $C_j$  is redundant once  $D_i$  is known. Thus the marginal likelihood is estimated to be the mean weight over all mock samples.

<sup>5</sup> This is equivalent to the cross-correlation, evaluated at zero-shift.

However, given the following equivalence:

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \int P(\alpha_i = \alpha, \beta_i = \beta | D_i, \mathcal{I}) P(\alpha, \beta | D, \mathcal{I}) d\alpha d\beta \\ &= \int \left[ \frac{1}{N} \sum_{i=1}^N P(\alpha_i = \alpha, \beta_i = \beta | D_i, \mathcal{I}) \right] P(\alpha, \beta | D, \mathcal{I}) d\alpha d\beta, \end{aligned} \quad (10)$$

we can add the many posteriors of all mock samples and renormalize:  $P(\alpha, \beta | C_j, \mathcal{I}) = \frac{1}{N} \sum_{i=1}^N P(\alpha_i, \beta_i | D_i, \mathcal{I})$ , then identify  $\mathcal{Z}$  as the zero-shift cross-correlation of the two aforementioned posteriors at the origin. Indeed, this is our practical approach. Note that since the probability distribution functions (PDFs) in equation (10) that must be added or multiplied are initially Monte Carlo sampled, we are forced to estimate the functions on a regular 2D  $\{\alpha, \beta\}$  grid via some choice of kernel. Whether nearest grid point, Gaussian convolution or kernel density estimation, with bandwidth of  $0.1 < \Delta\alpha < 0.5$  and  $0.02 < \Delta\beta < 0.1$ , all produce almost identical results. Any variation in  $\mathcal{Z}$ , at most 5 per cent, is negligible compared to the uncertainties explored in the next section.

Our solution requires simulations that produce several mock data sets, each of which can be compared to the real data via a summary statistic and kernel distance metric, as in approximate Bayesian computation (ABC; see Cameron & Pettitt 2012; Weyant et al. 2013; Robin et al. 2014; Akeret et al. 2015; Ishida et al. 2015; Lin & Kilbinger 2015, for examples of applications of ABC within the astrophysical literature). It differs from standard ABC in two ways. First, what we propose is not a likelihood-free approach; indeed our aim is to calculate a marginal likelihood. Secondly, rather than rejecting – and wasting – mock samples that are dissimilar to the real data, they are down-weighted; the weights incorporate both the kernel and the distance metric in traditional ABC. The effect should be similar to probabilistic acceptance of mock samples as outlined in Wilkinson (2013) and soft ABC as mentioned in Park, Jitkrittum & Sejdinovic (2015). Model selection here would boil down to comparing mean weights rather than acceptance fraction.

While the choice of summary statistics is open to discussion (our relatively simplistic choice is justified by the small sample and large scatter), the actual metric used to define the weighting/distance is not arbitrary. The caveat is that this method is only possible in cases where one can infer the summary statistics as a probability distribution for any (mock or real) data set.

## 6 $\Lambda$ CDM STRONG LENSING RESULTS

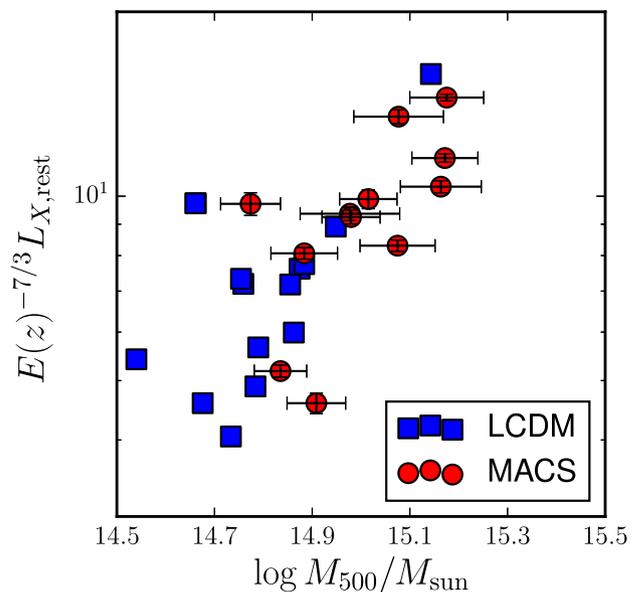
In this section we calculate  $\mathcal{Z}$ , where the data  $D$  are the high- $z$  MACS clusters and the cosmological model  $C$  is standard  $\Lambda$ CDM. For an initial demonstration in Section 6.1, we use our fiducial simulations and methodology, while in Sections 6.2–6.5 we examine the aspects of simulations and selection methods that affect this value.

### 6.1 AGN clusters at $z > 0.5$ selected by X-ray flux

The flux cut that was employed in the high- $z$  MACS sample can be translated into a luminosity cut and using the  $L_X$  (observer rest frame) determined for each cluster we can select simulated clusters for our sample (see Section 4.2):

$$f_{\text{cut}} = \frac{L_{X,\text{cut}}}{4\pi D_L^2}, \quad (11)$$

where  $D_L$  denotes the luminosity distance to the cluster. For the standard  $\Lambda$ CDM cosmology adopted in this work, and at  $z \approx$



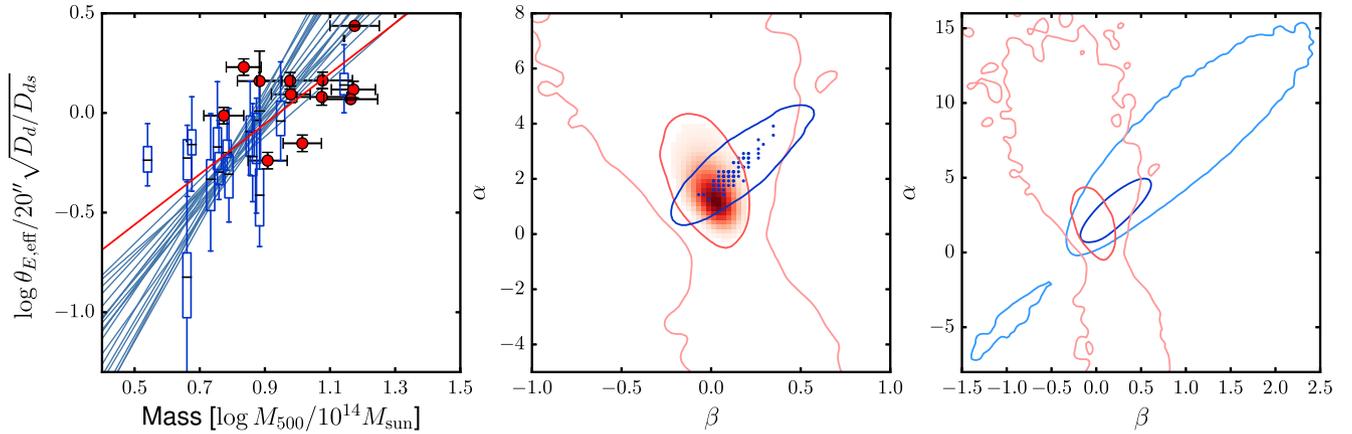
**Figure 1.** Cluster rest-frame X-ray luminosity, computed within the [0.1–2.4] keV energy band, as a function of cluster mass  $M_{500}$ . Blue squares denote  $z = 0.5$  clusters selected from the AGN simulations, while red circles denotes the high- $z$  MACS sample.

0.5, the threshold flux translates to a threshold luminosity of  $L_{X,\text{cut}} = 7.6 \times 10^{44} \text{ erg s}^{-1}$  (observer rest frame [0.1–2.4] keV). In the AGN simulation set, 15 clusters exceed the luminosity/flux threshold ( $L_X > L_{X,\text{cut}}$ ).<sup>6</sup>

There exists a tight correlation between core-excised X-ray luminosities of the high- $z$  MACS clusters and X-ray-based mass estimates (Mantz et al. 2010). We do not use core-excised luminosities to select our simulated clusters, however, nor do we use bolometric luminosities in the cluster rest frame, but rather the [0.1–2.4] keV band luminosities in the observer’s frame. This is done to best replicate the actual selection criteria. In Fig. 1 we show the  $L_X$ – $M$  relation for the simulated and observed clusters, where the X-ray luminosities are measured in the soft X-ray waveband. The  $L_X$ – $M$  self-similar relation is technically correct for bolometric luminosities, *but* Reiprich & Böhringer (2002) also suggest a relation with  $L_X$  in our band (Perrenod 1980). Self-similarity is generally better followed by relaxed clusters in hydrostatic equilibrium, while a high-luminosity sample would be biased towards unrelaxed clusters; for such a sample, one might expect higher luminosities for a fixed mass. Even with the inclusion of AGN feedback, the simulated clusters remain slightly overluminous for a given mass, relative to the observed clusters. There is a distinct lack of high-mass candidates in the simulated sample, witnessing that for the adopted cosmology even a box size as large as  $1 h^{-1}$  Gpc does not contain a large enough population of massive clusters at the redshift of interest,  $z = 0.5$ . Note that in the following analysis, we do not simply consider the observed distribution of Einstein radii, which would be severely biased by the lack of high-mass clusters, but rather the lensing–mass relation.

In the left-hand panel of Fig. 2 we show the relation between the Einstein radii and the cluster mass  $M_{500}$ . The  $z > 0.5$  clusters of the MACS sample are represented by red circles. For simulated clusters, the situation is more complicated. Since different lines of

<sup>6</sup> If more simulated clusters are available for follow-up studies, one could also account for cosmic variance.



**Figure 2.** Einstein radii statistics for  $z = 0.5$  clusters from the AGN simulations. Left: strong lensing efficiency, characterized by scaled Einstein radii,  $\theta_{E,\text{eff}}$ , plotted as a function of  $M_{500}$ . The range of Einstein radii for simulated clusters is shown by the blue box plots. The red circles represent the MACS  $z > 0.5$  clusters, whose masses have been corrected as described in Section 3. The red line marks the maximum a posteriori fit to observational data, while the thin blue lines mark the fit to 20 randomly chosen mock samples from simulations. Middle:  $1\sigma$  and  $2\sigma$  constraints on parameters of the strong lensing–mass relation given the MACS  $z > 0.5$  cluster data (red contours and shading). Overplotted in blue dots are the best fits to 80 mock observations of  $z = 0.5$  clusters from the AGN simulations. A typical  $1\sigma$  error is shown as a blue ellipse. Right: same as the middle panel, but the blue curves mark the  $1\sigma$  and  $2\sigma$  contours of  $P(\alpha, \beta | C, \mathcal{D})$ , having combined all mock observations. The value of  $\mathcal{Z}$  is the integral over the product of the PDFs marked by the red and blue contours.

sight provide a large variation in projected mass distribution, each cluster cannot be associated with an individual Einstein radius, nor a simple Gaussian or lognormal distribution (see Killedar et al. 2012). We therefore measure the Einstein radius for 80 different lines of sight and, for ease of visualization, describe the distribution of Einstein radii for each simulated cluster by a box plot.<sup>7</sup>

As described at the beginning of this section, we fit the observational data to the strong lensing–mass relation and after marginalizing out the nuisance parameter,  $V$ , present  $P(\alpha, \beta | D, \mathcal{I})$ , the posterior distribution for  $\alpha$  and  $\beta$ , denoted by red contours in the middle panel of Fig. 2. Many mock samples are individually fit to the lensing–mass relations; the maximum of the posterior is shown as a blue point, one for each mock, and a typical  $1\sigma$  error shown as a blue ellipse. By adding the posteriors for each mock sample and renormalizing, we estimate  $P(\alpha, \beta | C, \mathcal{D})$ , shown by the blue contours in the right-hand panel of Fig. 2. By multiplying by the two distributions, we find  $\mathcal{Z} = 0.25$ .

Note that one cannot comment on whether the marginal likelihood is *large* or *small*. One cannot use this value to claim ‘consistency’ or ‘tension’ with  $\Lambda$ CDM. However, if the same process is repeated for simulations under a different cosmological model, then the Bayes factor  $R$  can be calculated and, after accounting for priors, it may (or may not) reveal a preference for one of the cosmologies, in light of this data.

Note that we have allowed for negative  $\alpha$ , i.e. negative slopes for the strong lensing–mass relation, which seems counterintuitive but should not be ruled out on principle given the anticorrelation between concentration and cluster mass. If we do insist on positive slopes, by placing a boundary on the prior on  $\alpha$ , we find that  $\mathcal{Z}$  increases by about 5 per cent.

The use of the AGN simulation set and the cluster selection as described characterize our fiducial approach. Certain factors can be expected to impact the marginal likelihood, and consequently the Bayes factor when comparing cosmologies. For the remainder of

this section, we consider how the marginal likelihood for our specific cosmology may depend on other details, such as cluster redshift, selection criteria and the numerical implementation of baryonic processes.

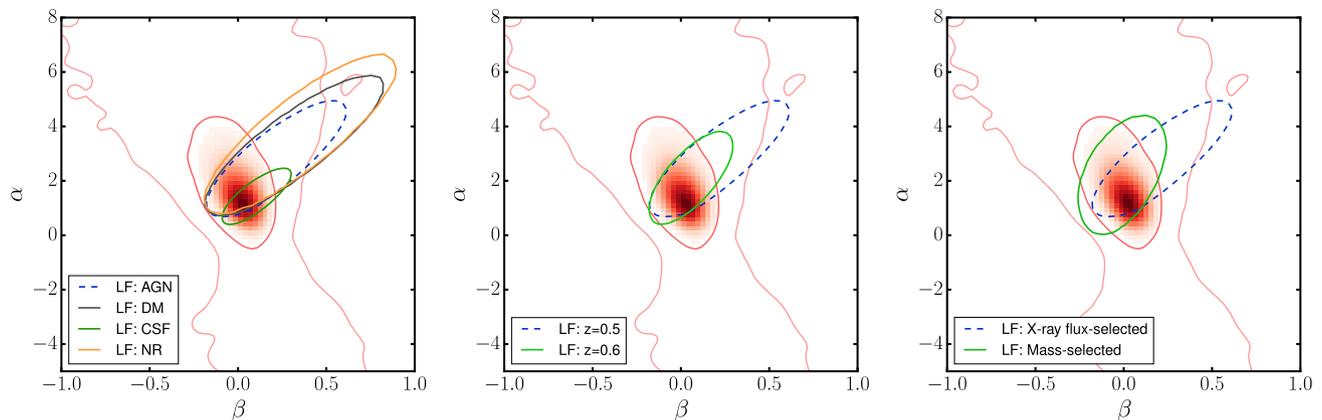
## 6.2 Effect of baryon processes

By using hydrodynamic simulations, we are now in the position of being able to select clusters in a manner more consistent with the selection of the observational sample. However, given the subgrid nature of the astrophysical processes and subsequent uncertainties in their implementation, we consider it prudent to determine the sensitivity of our scientific conclusions to the inclusion of baryonic processes and resulting gas distribution (see Ragone-Figueroa et al. 2013; Wurster & Thacker 2013 for detailed discussions on different implementations of AGN feedback). In this section we perform the same comparison as before but for the clusters in the DM, NR and CSF simulations. The clusters selected are the counterparts to the clusters selected in Section 6.1. That is, clusters are selected in the AGN simulation according to their X-ray luminosity as would be observed in the 0.1–2.4 keV band; then, the *same* clusters are selected in the other simulations.<sup>8</sup> Therefore, we are able to determine the effect of baryonic physics on the marginal likelihood estimated using the same sample of clusters.

Radiative processes can impact on cluster mass,  $M_{500}$ , and the mass profiles and consequently, strong lensing at  $z = 0.5$  (Killedar et al. 2012; Cui, Borgani & Murante 2014). CSF-like simulations, which suffer from overcooling and steepened profiles, result in up to 10 per cent increase in  $M_{500}$  relative to DM simulations at  $z = 0.5$ , with no significant trend with mass, but increase the Einstein radii for low-mass haloes in particular, resulting in a preference for a lower value of  $\alpha$ , as seen in the left-hand panel of Fig. 3. Haloes are more spherical as a consequence of adiabatic contraction; the

<sup>7</sup> In the box plots, we mark the median with a short black horizontal line, a blue box marking the 25th and 75th percentiles and stems to meet the furthest data points within 1.5 times the interquartile range.

<sup>8</sup> Note that these clusters are not selected based on their X-ray luminosity in the DM, NR and CSF simulations; the degree to which  $L_X$  is overestimated in NR and CSF simulations is greater than for AGN simulations, and X-ray luminosities are obviously not defined for the DM clusters.



**Figure 3.** Observed fit to the scaling relation as in the right-hand panel of Fig. 2 (red) and  $1\sigma$  contours of  $P(\alpha, \beta|C, \mathcal{I})$  derived from: (left-hand panel) the different simulation sets: *AGN* (blue; fiducial), *DM* (grey), *CSF* (green), *NR* (orange); (middle panel) simulated clusters at  $z = 0.5$  (blue; fiducial) and  $z = 0.6$  (green); (right-hand panel) simulated clusters selected by X-ray flux (blue; fiducial) and mass (green).

increased degree of sphericity causes, in turn, a decrease in the variation between mock samples, and therefore a much narrower  $P(\alpha, \beta|C, \mathcal{I})$  distribution. AGN feedback tempers the effect on both cluster mass and strong lensing. Ultimately, we find  $\mathcal{Z} = 0.74$  for the *CSF* simulations, about triple that derived from the *AGN* simulations, while the *DM* and *NR* simulations result in  $\mathcal{Z} = 0.18$  and  $0.14$ , respectively, about half that of *AGN*.<sup>9</sup> These simulations are extremes in terms of the astrophysical processes that are ignored. However, as they bracket our true ignorance of the thermal and kinetic effects of baryonic processes, then this implies an uncertainty in the  $\mathcal{Z}$  – and therefore  $R$  – of a factor of 3.

### 6.3 Changing the redshift of simulated clusters

11 out of the 12 high- $z$  MACS clusters lie within  $0.5 < z < 0.6$  (J0744.8+3927 lies at  $z \approx 0.7$ ). Thus far we have compared the strong lensing properties of the MACS clusters with simulated clusters by extracting the latter at the lower end of the redshift range:  $z = 0.5$ . We remind the reader that the Einstein radius is scaled in order to account for the different redshifts of the clusters being compared. However, this does not account for any structural differences due to clusters at  $z = 0.5$  being captured at a later stage of evolution than those at  $z = 0.6$ . Since cluster mass concentration at fixed total mass is expected to increase with redshift, then the choice of simulating clusters at only  $z = 0.5$  could potentially underestimate the  $\Lambda$ CDM prediction for strong lensing for clusters where this redshift is only the lower limit. Therefore, we repeat the comparison for simulated clusters at  $z \approx 0.6$  instead. At  $z = 0.6$ , the threshold flux translates to a threshold luminosity of  $L_{X, \text{cut}} = 11.9 \times 10^{44} \text{ erg s}^{-1}$  (observer rest frame [0.1–2.4] keV); 11 clusters in the *AGN* simulations satisfy this selection criteria.

The simulation-based  $P(\alpha, \beta|C, \mathcal{I})$  and the observed  $\alpha$  and  $\beta$  are shown in the middle panel of Fig. 3, akin to the right-hand panel of Fig. 2. The relationship between the Einstein radii and mass is similar to that of the  $z = 0.5$  simulated sample, notwithstanding the absence of any cluster with unusually small Einstein radius. The result is a more strongly peaked function  $P(\alpha, \beta|C, \mathcal{I})$  (blue

contours). In this case, we measure  $\mathcal{Z} = 0.49$ , which is almost double that derived from the  $z = 0.5$  simulations.

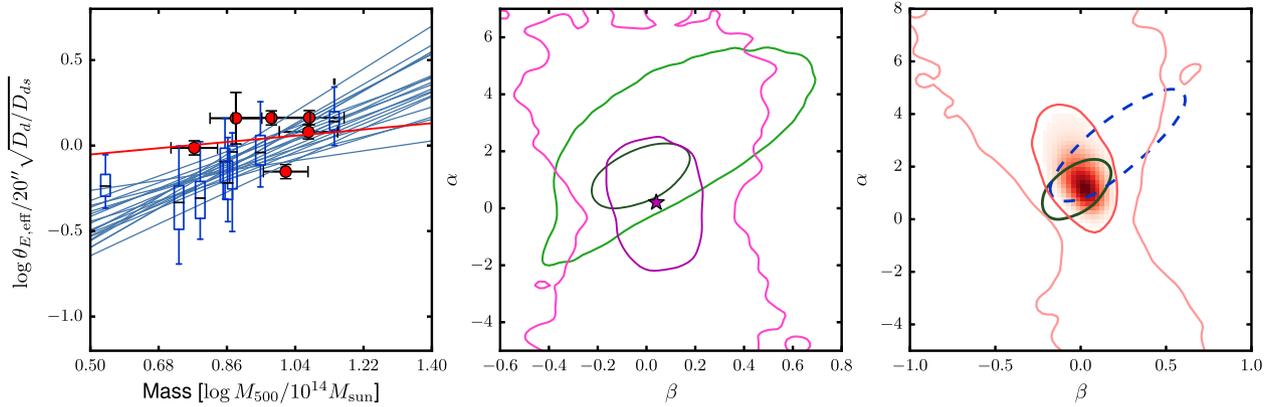
### 6.4 Cluster selection by mass

It is common practice to select simulated cluster samples using a mass threshold, or some other proxy for X-ray flux, when simulations do not contain gas dynamics (e.g. Horesh et al. 2010; Meneghetti et al. 2011). However, since the clusters in the MACS survey (among others) were selected by flux rather than luminosity, there is no corresponding mass threshold, strictly speaking. In fact, even if the selection is by luminosity, there will be a preference for high-concentration clusters, for a fixed mass, which are relatively X-ray brighter (Rasia et al. 2013).

On the other hand, mimicking X-ray selection for simulated clusters requires a robust treatment of the hot X-ray emitting ICM. While our simulated clusters follow the luminosity–temperature relation reasonably well (Planelles et al. 2014), we note that there is still some small degree of overluminosity in the scaling relation against mass, possibly due to violation of hydrostatic equilibrium in observational mass estimates. The single-redshift choice means that flux, luminosity and mass thresholds are equivalent. Thus we are able to investigate if the selection of clusters by X-ray luminosities introduces low-mass clusters into our simulated sample. To address this concern, we select simulated clusters by imposing a  $M_{500}$  mass threshold corresponding to the lowest mass  $z > 0.5$  MACS cluster.

Assuming that there are not many low-luminosity high-mass clusters introduced into the sample, we could expect, given the left-hand panel of Fig. 2, that in fact a few MACS clusters would have Einstein radii that are smaller than those typically measured in simulated clusters. Accounting for the 2009 *Chandra* calibration update, we estimate the lowest mass high- $z$  MACS cluster to have  $M_{500} = 4.2 \times 10^{14} h^{-1} M_{\odot}$ . Accordingly, we select the eight clusters from the *AGN* simulations with  $M_{500}$  above this value. Indeed, we find that compared to the simulated sample from Section 6.1, the seven lowest mass clusters drop out of the simulated sample, while no low-luminosity clusters are added. The function  $P(\alpha, \beta|C, \mathcal{I})$  for both assumed selection methods and observational fit to the scaling relation are shown in the right-hand panel of Fig. 3. By selecting simulated clusters by mass,  $\mathcal{Z}$  increases to 0.42.

<sup>9</sup> As an aside, we could assume that the cosmological model is correct and test cluster physics instead using such simulations but in that case perhaps the posterior for  $\alpha$  and  $V$  would be more revealing.



**Figure 4.** Left: as in the left-hand panel of Fig. 2 but for dynamically relaxed clusters; middle: constraints on  $\alpha$  and  $\beta$  for the relaxed subsample of the MACS clusters (magenta; best fit marked with a star), and  $P(\alpha, \beta | C, \mathcal{I})$  using the relaxed simulated clusters (green); right: constraints on  $\alpha$  and  $\beta$  for the complete MACS sample (red), and  $P(\alpha, \beta | C, \mathcal{I})$  using the complete sample of simulated clusters (blue dashed) and the relaxed subsample (green solid).

### 6.5 Cluster selection by dynamical state

Five of the 12  $z > 0.5$  MACS clusters are classified as dynamically relaxed, according to a morphological code described in Ebeling et al. (2007). This low fraction is not surprising for an X-ray flux-selected sample since cluster mergers are known to lead to large boosts in X-ray luminosity (e.g. Ritchie & Thomas 2002; Planelles & Quilis 2009). On the other hand, observational programmes such as CLASH collect clusters according to their X-ray contours and alignment of the brightest cluster galaxy (BCG) with the X-ray peak, in an effort to choose relaxed clusters. Here, we determine how this choice of relaxed clusters affects the marginal likelihood.

The background information  $\mathcal{I}$  now includes the dynamical state criterion, and so the mock data are limited to the relaxed subsample out of the fiducial AGN clusters. The dynamical state of the simulated clusters is defined following the method of Killedar et al. (2012). The method consists of computing the offset between the position of the particle with the minimum gravitational potential and centre of mass (COM), where the COM is calculated within a range of radii  $\zeta_i R_{\text{vir}}$ , with  $\zeta_i$  going from 0.05 to 2 in 30 logarithmic steps. A cluster is defined as relaxed if the offset is less than 10 per cent of  $\zeta_i R_{\text{vir}}$  for all radii. In the AGN simulations nine relaxed clusters exceed the luminosity/flux threshold ( $L_X > L_{X, \text{cut}}$ ). In both the MACS and simulated sample, half the clusters are deemed relaxed. The Einstein radii and masses of the relaxed subsample of clusters are shown in the left-hand panel of Fig. 4; with the observed fit described by the magenta contours and  $P(\alpha, \beta | C, \mathcal{I})$  shown in green in the middle panel. The inferred marginal likelihood is  $\mathcal{Z} = 0.39$ .

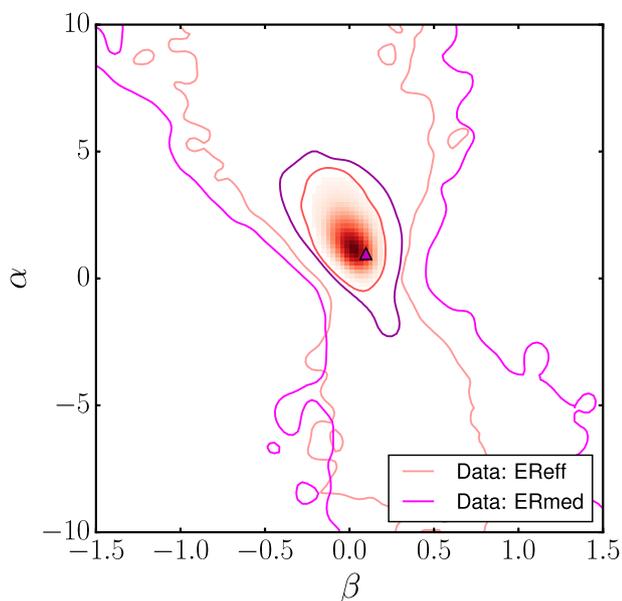
It would be unwise to use theoretical models based on the assumption of a relaxed sample *if* the observational sample did not include this criterion. In order to demonstrate this, consider now the effect of applying the incorrect selection criteria when modelling lenses. Excluding the most disturbed clusters from both the simulated and observational sample might tighten the fit on  $\alpha$  and  $\beta$ , but this is compensated for by a loosening fit due to the smaller sample size. Ultimately, as seen in the right-hand panel of Fig. 4, the function  $P(\alpha, \beta | C, \mathcal{I})$  derived from the relaxed simulated clusters (green contours) is more sharply peaked than that which is derived without this additional selection (blue contours). If this relaxed simulated sample is used to analyse the full observational sample (constraints shown in red), one would incorrectly derive a value of  $\mathcal{Z} = 0.7$ ; we remind the reader that including merging clusters in the simulated sample resulted in a smaller marginal likelihood by a factor of 4.

Relaxed subsamples as shown here provide much less data to work with, and thus a less powerful cosmological test. Furthermore, the dynamical state of simulated clusters has been evaluated in three dimensions, which is not exactly consistent with choices based on projected observables. Criteria presented by Merten et al. (2015), Meneghetti et al. (2014) and the automated methods of Mantz et al. (2015) have recently made it possible to mimic the morphological selection. If this can be applied to a large number of both simulated and observed clusters with mass estimates and strong lensing measurements, we could determine a more robust marginal likelihood.

## 7 COMPARISON TO PREVIOUS WORK

There have been four other works in the literature that analyse the strong lensing statistics of the high- $z$  MACS clusters and compare them to predictions from simulations (Horesh et al. 2011; Meneghetti et al. 2011; Zitrin et al. 2011; Waizmann et al. 2014). Several factors could lead to disagreements with our findings presented in Section 6. First, the theoretical predictions for the first three works were based on the adoption of 1-year *Wilkinson Microwave Anisotropy Probe* (WMAP-1; Spergel et al. 2003) cosmological parameters including, most notably, a high normalization for the matter power spectrum:  $\sigma_8 = 0.9$ . These parameters predict an earlier epoch for structure formation relative to the preferred model based on WMAP-7 results (Komatsu et al. 2011), and thus clusters are predicted to be more concentrated and stronger lenses than predicted by our simulations. Additionally, simulations in the above previous analyses were either collisionless  $N$ -body, or included only non-radiative physics; however as we have shown in Section 6.2, the effects of baryons are minor compared to the substantial scatter associated with cluster triaxiality. Finally, the inferred strong lensing properties of the high- $z$  MACS sample for all three studies described below were based on mass models constructed prior to the availability of the high-quality *HST* data within the CLASH and Frontier Fields programmes. All these factors aside, our main focus here is on the statistical methods used.

Horesh et al. (2011) measured the frequency of arc production in cluster lenses from the Millennium Simulation (Springel et al. 2005b) at three different redshift bins simulated clusters ( $z \sim 0.2, 0.4$  and  $0.6$ ). The high- $z$  MACS sample were compared with the highest redshift bin. They found that simulations



**Figure 5.** The posterior probability on  $\alpha$  and  $\beta$  for the high- $z$  MACS data for Einstein radii defined as  $\theta_{E,\text{eff}}$  (red) and  $\theta_{E,\text{med}}$  (magenta; best fit marked with a triangle).

underpredicted the number of arcs per cluster, but cautioned that there were too few simulated clusters available (four and one for their low and high mass threshold, respectively) to form a robust conclusion.

Meneghetti et al. (2011) based the  $\Lambda$ CDM predictions on clusters from six snapshots in the range  $0.5 < z < 0.7$  from the MareNostrum Universe non-radiative gas simulations (Gottlöber & Yepes 2007). The selection criteria included X-ray flux selection, but required a correction term for luminosities that are estimated from simulations that have a relatively simple description of the gas. Their comparisons revealed that the predicted lensing cross-section for giant arc-like images of sources at  $z_s = 2$  was half that of the observed value,<sup>10</sup> while the Einstein radii (characterized through the alternative ‘median’ radius) differed by 25 per cent. They then claimed to close the gap between observations and simulations with the inclusion of realistically modelled merging clusters. However, Redlich et al. (2012) have shown that the median Einstein radius is more sensitive to cluster mergers and will be boosted for a longer period during the merger. In fact, three of the MACS clusters were singled out for their unusually strong lensing qualities: MACS J0717.5+3745, MACS J0025.4–1222 and MACS J2129.4–0741; yet these are precisely the clusters for which the characterization of the Einstein radii makes the most difference:  $\theta_{E,\text{med}}/\theta_{E,\text{eff}} = 1.3, 1.9$  and  $2.2$ , respectively. Fig. 5 demonstrates that there is a much poorer scaling relationship between the Einstein radius if characterized in this alternative manner and the cluster mass for the  $z > 0.5$  MACS clusters due to merger-driven boosts.

Zitrin et al. (2011) determined a theoretical distribution for effective Einstein radii using smooth triaxial lens models with parameters constrained by  $N$ -body simulations. The lens structure was designed to match results from earlier collisionless simulations and they selected clusters above a mass threshold based on the flux limit using

<sup>10</sup> The values of  $\theta_{E,\text{med}}$  as quoted in Meneghetti et al. (2011) are based on the earlier mass models, prior to the availability of *HST* data within the CLASH and Frontier Fields programmes.

the (relaxed cluster)  $L_X$ – $M$  relation from Reiprich & Böhringer (1999). Einstein radii were found to be about 1.4 times larger than predicted by  $\Lambda$ CDM, measured by comparing the medians of the total distributions.

A recent study by Waizmann et al. (2014) takes a different approach to the strong lensing comparison by using order statistics. Semi-analytic models of cluster lensing and mass function are combined with general extreme value distributions to determine ‘exclusion constraints’ on the  $n$ -largest observed Einstein radii; the high- $z$  MACS sample is found to be consistent with  $\Lambda$ CDM.

The studies of Horesh et al. (2011) and Meneghetti et al. (2011) employ directly the results from  $N$ -body and non-radiative gas simulations, while Zitrin et al. (2011) and Waizmann et al. (2014) use semi-analytic models. In this work, we have modelled cluster lenses from hydrodynamic simulations, allowing a realistic description of baryonic effects and of unrelaxed clusters, as well as the directly comparable X-ray flux selection. However, unlike the semi-analytic models, the number of objects is still limited.

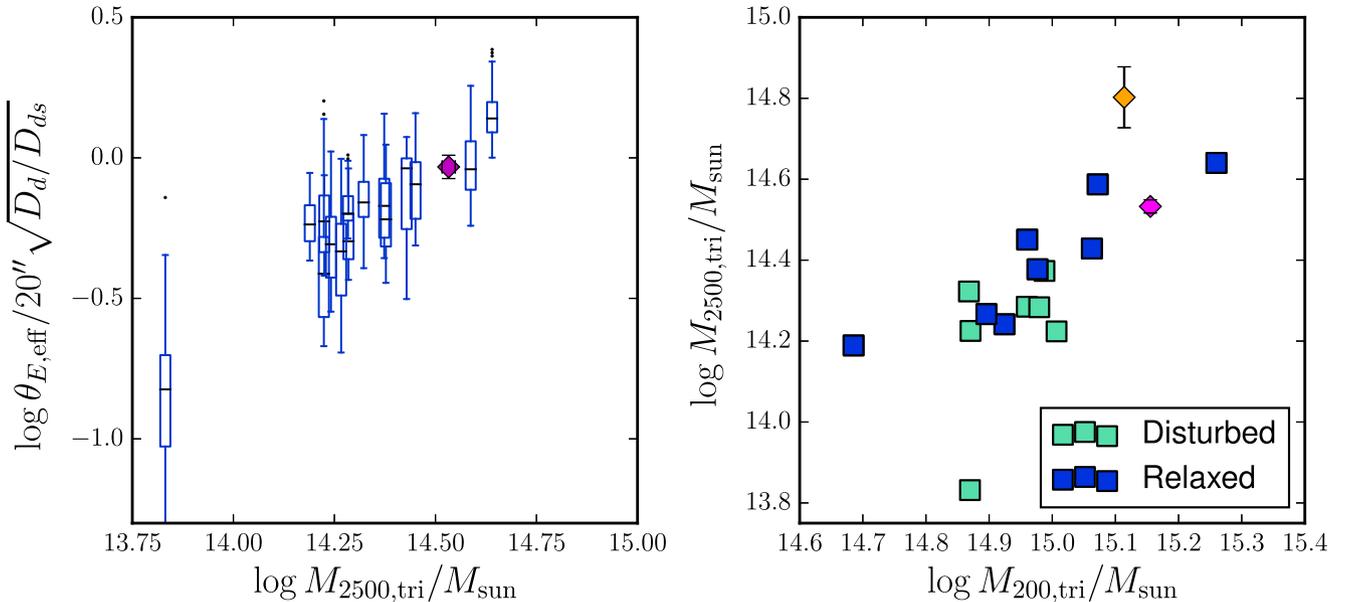
The key difference between this study and those conducted before lies in the statistical approach. Waizmann et al. (2014) employs order statistics, while the first three studies employ the Kolmogorov–Smirnov (KS) test in order to compare the distribution of a strong lensing property (either lensing cross-sections or Einstein radii) for the high- $z$  MACS sample to the similarly constructed distribution for a mock sample consisting of simulated clusters; otherwise the median of their distributions is compared as a consistency check. However,

- (i) comparing medians of distributions involves the loss of much of the information in the data;
- (ii) uncertainties in the measurement of the Einstein radii are ignored;
- (iii) different lines of sight through the same simulated cluster can produce different mock observations due to cluster triaxiality and the presence of substructures, leading to wildly varying results for the KS test;
- (iv) all aforementioned statistical approaches – and the majority of the literature analysing the strong lensing efficiency or concentration of galaxy clusters – do not formally provide a marginal likelihood that would allow one to judge the preference for the  $\Lambda$ CDM cosmological model *over other cosmologies*.

In this work instead, we take a Bayesian approach and provide a guide to performing the first step of the model-selection problem: determining the marginal likelihood, the probability of observing the lensing–mass relationship assuming a single power law form to the scaling relation. The rest of the comparison requires numerous simulations for other cosmologies and is therefore outside the scope of this work. However, we have laid the groundwork for a strong lensing test of cosmology.

## 8 TRIAXIAL MODELS

$\beta$ -profile fits to X-ray emission from the ICM provide a gas mass estimate that is converted to a total mass estimate assuming a gas/baryon fraction in clusters; the mass recovered is limited to within  $R_{500}$ . Weak lensing data can allow one to measure the total mass of a cluster out to larger radii and perhaps include shape measurements in combination with X-ray and/or Sunyaev–Zel’dovich (SZ) data (Marshall, Hobson & Slosar 2003; Mahdavi et al. 2007; Morandi, Pedersen & Limousin 2010; Sereno & Umetsu 2011; Limousin et al. 2013; Sereno et al. 2013). Triaxial model fits provide more realistic descriptions of cluster mass profiles, but they



**Figure 6.** Einstein radius as a function of enclosed mass within ellipsoid for an average overdensity of  $\Delta = 2500$ . Blue box plots represent simulated clusters, the magenta diamond represents MACS J1423.8+2404 and the orange diamond represents Abell 1689.

will generally also result in a higher mass estimate than if spherical symmetry was assumed (Corless & King 2007). Given the increasing availability of lensing and SZ data, and subsequent possibility of shape reconstruction, we consider it useful to explore the relationship between strong lensing efficiencies of clusters and their ellipsoidal mass. We have performed triaxial shape measurements on the simulated clusters for a fixed overdensity of 200 and 2500 to derive axis ratios (see Appendix B for details) and thus calculate the resulting mass within the ellipsoid ( $M_{200, \text{tri}}$  and  $M_{2500, \text{tri}}$ , respectively).

Despite the large line of sight related variation, the Einstein radius of a cluster tends to scale well with mass at high overdensities as seen clearly in the left-hand panel of Fig. 6. Therefore, we may use the mass at high overdensities in place of Einstein radius for a scaling relation as previously discussed. Consider, then the potential to use the relationship between triaxial masses at low and high overdensities, as shown in the right-hand panel of Fig. 6, and the associated fit parameters as summary statistics. The differences between mock samples are no longer due to triaxiality, but rather due to cosmic variance. A full simulation-based likelihood function will require a much more extensive cluster sample, so unfortunately it is out of the scope of this work.

As for observational data, we consider the following two galaxy clusters. MACS J1423.8+2404 is arguably the most dynamically relaxed cluster of the high- $z$  MACS sample and has a relatively low-substructure fraction (Limousin et al. 2010). This has made it an ideal candidate for reconstruction of the triaxial mass profile using multiple data sets. Morandi et al. (2010) have reconstructed the three-dimensional structure of this cluster using a combination of X-ray, SZ and weak lensing data, determining a best-fitting set of triaxial parameters and the enclosed mass for an overdensity of  $\Delta = 2500$ . Limousin et al. (2013) improved the algorithm used to combine data and provide a best-fitting set of generalized Navarro–Frenk–White (NFW) model parameters and enclosed mass for an overdensity of  $\Delta = 200$ . A second cluster, Abell 1689, is not part of the high- $z$  MACS sample but has been the subject of extensive study regarding its triaxial shape and mass reconstruction using a variety

of probes (Serenio & Umetsu 2011; Limousin et al. 2013; Sereno et al. 2013; Umetsu et al. 2015). We show both these clusters in Fig. 6 as well, although there are too few to draw any strong conclusions. At present, multiwavelength reconstruction methods are likely to be restricted to relaxed clusters, so we would advocate providing likelihood functions for relaxed clusters (as marked by dark blue squares) as well.

## 9 CONCLUSIONS

The MACS  $z > 0.5$  sample has been the subject of a number of studies in the literature (i.e. Horeish et al. 2011; Meneghetti et al. 2011; Zitrin et al. 2011; Sereno & Zitrin 2012) regarding the consistency of their strong lensing properties with  $\Lambda$ CDM. Since these studies were undertaken, several of the clusters have had revisions to their mass models as a result of the multiband data and precise *HST* imaging from the CLASH and Frontier Fields programmes; in some cases the new Einstein radii are significantly lower than previously estimated.

Our primary goal here is to step away from claims of ‘tension’ or ‘consistency’ with a single cosmological model. We have introduced a new approach to calculating a marginal likelihood for measuring the relationship between strong lensing efficiency and cluster lens mass, and demonstrated with the high- $z$  MACS clusters under the  $\Lambda$ CDM model and a choice of cosmological parameters consistent with *WMAP-7* results. Model predictions are based on a finite sample of galaxy clusters simulated with hydrodynamic processes. Since our interest lies in typical strong lensing properties of cluster lenses at fixed mass, we consider a scaling relation between the Einstein radius of the lens and its mass. We assume a power-law relation (see equation 7), and interpret the parameters of the scaling relations,  $\alpha$  and  $\beta$ , as summary statistics. We employ cosmological simulations to determine the probability of observing these parameters under  $\Lambda$ CDM. Folding this function with the observed fit, we derive the marginal likelihood. This forms the first step in a Bayesian model selection process.

The approach described above is an exciting new strategy for estimating the marginal likelihood for a given cosmology using strong lensing galaxy clusters. However, we recognize that the calculation involves running computer simulations that can take months. We find that the marginal likelihood would vary by a factor of 3 or so only if the most unrealistic simulations are employed, with significant overcooling or otherwise lacking radiative processes entirely. We invite development of cosmological simulations that include a large range of hydrodynamical processes with different philosophies for the implementation of subgrid physics, feedback models and implementation of hydrodynamical schemes. We expect simulations run under a range of cosmological models to be analysed in a manner equivalent to that demonstrated in this work to allow eventual model selection. Our findings suggest that if a model-comparison study was carried out using a simulation based on an alternative cosmological hypothesis and resulted in a Bayes factor of 20 or more (see equation 6), then *DM* simulations would be sufficient. However, in the event that the *R* is found to be smaller, then the computationally expensive hydrosimulations would be necessary.

Cluster selection is an important factor, however, particularly the choice by dynamical state. Samples that include clusters in unrelaxed dynamical states are problematic: they introduce a large amount of scatter in the lensing–mass scaling relation, which consequently weakens the constraints on  $\alpha$  and  $\beta$  and reduces the power of cosmological test. On the other hand, relaxed subsamples are smaller in number, which have a similar effect. Furthermore, since some measures of strong lensing (e.g.  $\theta_{E,med}$ ) can be highly sensitive to merging events (and line of sight on the sky), smaller samples will suffer more from occasional boosts in lensing efficiency, and weaken the strength of statistical tests. We warn against using theoretical models that assume a relaxed morphology to analyse an observational sample that includes clusters undergoing mergers. Instead, we recommend the use of effective Einstein radius to characterize strong lensing in a large homogeneous sample of relaxed clusters selected with automated methods, such as those of Mantz et al. (2015).

The relationship between masses at low and high overdensities is an interesting alternative to the strong lensing–mass scaling relation. Here, we advocate folding cluster triaxiality into the analysis in order to reduce variation between mock samples. Measuring these masses will become feasible over the coming years given the availability of multiwavelength data. Alternatively, weak lensing estimates alone would provide a more precise measure of cluster mass, albeit usually at lower overdensities or at fixed aperture radius (e.g. Applegate et al. 2014).

For follow-up studies, we aim to the following.

(i) Include more simulated clusters in a parent sample that would then allow one to account for ‘cosmic variance’ between mock samples, not just the orientation-related scatter. This would reduce any bias in the estimator of the marginal likelihood.

(ii) Use semi-analytic models or emulators in place of individual simulated clusters (e.g. Giocoli et al. 2012; Kwan et al. 2013; Jabot et al. 2014; Bonamigo et al. 2015). This would allow even easier modelling of cosmic variance, however, fitting functions here are still based on simulations.

(iii) Relax the assumption of the power law relationship between the Einstein radius and cluster lens mass. However a new summary statistic would need to be identified, possibly with more than two dimensions. Otherwise, alternative definitions of discrepancy distance could be explored (e.g. Park et al. 2015).

(iv) Apply generalized linear models (GLMs; e.g. de Souza et al. 2015a,b; Elliott et al. 2015) in place of the analysis as described in Section 5, in the case in which full error distributions on measurements become available.

While numerical simulations have opened up the possibility to explore the effects of cosmology and physics on non-linear structure formation, they have brought with them the necessity to break away from ‘textbook’ statistical methods that assume the existence of analytical models. We hope, therefore, that we have begun a discussion about alternatives that may prove as revolutionary to cosmological testing as the simulations themselves.

## ACKNOWLEDGEMENTS

The authors thank Johannes Buchner, Ewan Cameron, Geraint Lewis, Jiayi Liu, Hareth Saad Mahdi, Julio Navarro, Dennis Prangle, Matthias Redlich, Christian Robert, Alexandro Saro, Micha Schirmer, Mauro Sereno, Anja von der Linden and Adi Zitrin for helpful discussions, as well as Giuseppe Murante for his involvement in running the simulations. We also thank the anonymous referee for comments that aided in clarifying the paper. MK acknowledges support by the DFG project DO 1310/4-1, and a fellowship from the European Commission’s Framework Programme 7, through the Marie Curie Initial Training Network CosmoComp (PITN-GA-2009-238356). KD acknowledges the support by the DFG Cluster of Excellence ‘Origin and Structure of the Universe’. SP acknowledges support by a grant funded by the ‘Consorzio per la Fisica di Trieste’, by the PRIN-INAF09 project ‘Towards an Italian Network for Computational Cosmology’, by the PRIN-MIUR 201278X4FL grant ‘Evolution of Cosmic Baryons’, by the INDARK INFN grant and by the Spanish Ministerio de Economía y Competitividad (MINECO; grant AYA2013-48226-C-2-P) and Generalitat Valenciana (grant PROMETEO II/2014/069). Simulations have been carried out at the CINECA Supercomputing Centre in Bologna (Italy), with CPU time assigned through IS CRA proposals and through an agreement with University of Trieste. This work has been supported by the PRIN-MIUR 2012 grant ‘Evolution of Cosmic Baryons’, by the PRIN-INAF 2012 grant ‘Multi-scale Simulations of Cosmic Structure’ and by the INFN ‘INDARK’ grant. We acknowledge support from ‘Consorzio per la Fisica – Trieste’. This work has made use of the following open-source software: PYTHON,<sup>11</sup> NUMPY,<sup>12</sup> SCIPY,<sup>13</sup> MATPLOTLIB (Hunter 2007) and COSMOLOGY.<sup>14</sup> The bibliographic research was possible thanks to the tools offered by the NASA Astrophysical Data Systems.

## REFERENCES

- Akeret J., Refregier A., Amara A., Seehars S., Hasner C., 2015, *J. Cosmol. Astropart. Phys.*, 8, 043  
 Applegate D. E. et al., 2014, *MNRAS*, 439, 48  
 Bartelmann M., Steinmetz M., Weiss A., 1995, *A&A*, 297, 1  
 Bartelmann M., Huss A., Colberg J. M., Jenkins A., Pearce F. R., 1998, *A&A*, 330, 1  
 Bartelmann M., Meneghetti M., Perrotta F., Baccigalupi C., Moscardini L., 2003, *A&A*, 409, 449  
 Bayliss M. B., 2012, *ApJ*, 744, 156

<sup>11</sup> <http://www.python.org>

<sup>12</sup> <http://www.numpy.org>

<sup>13</sup> <http://www.scipy.org>

<sup>14</sup> <http://roban.github.com/CosmoPy/>

- Bayliss M. B., Johnson T., Gladders M. D., Sharon K., Oguri M., 2014, *ApJ*, 783, 41
- Blanchard P. K., Bayliss M. B., McDonald M., Dahle H., Gladders M. D., Sharon K., Mushotzky R., 2013, *ApJ*, 772, 24
- Boldrin M., Giocoli C., Meneghetti M., Moscardini L., Tormen G., Biviano A., 2016, *MNRAS*, 457, 2738
- Bonafede A., Dolag K., Staszczyn F., Murante G., Borgani S., 2011, *MNRAS*, 418, 2234
- Bonamigo M., Despali G., Limousin M., Angulo R., Giocoli C., Soucail G., 2015, *MNRAS*, 449, 3171
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Cameron E., Pettitt A. N., 2012, *MNRAS*, 425, 44
- Coe D. et al., 2013, *ApJ*, 762, 32
- Cole S., Lacey C., 1996, *MNRAS*, 281, 716
- Comerford J. M., Natarajan P., 2007, *MNRAS*, 379, 190
- Corless V. L., King L. J., 2007, *MNRAS*, 380, 149
- Cui W., Borgani S., Dolag K., Murante G., Tornatore L., 2012, *MNRAS*, 423, 2279
- Cui W., Borgani S., Murante G., 2014, *MNRAS*, 441, 1769
- Dalal N., Holder G., Hennawi J. F., 2004, *ApJ*, 609, 50
- de Souza R. S. et al., 2015a, *Astron. Comput.*, 12, 21
- de Souza R. S., Hilbe J. M., Buelens B., Riggs J. D., Cameron E., Ishida E. O., Chies-Santos A. L., Killedar M., 2015b, *MNRAS*, 453, 1928
- Despali G., Tormen G., Sheth R. K., 2013, *MNRAS*, 431, 1143
- Dubinski J., Carlberg R. G., 1991, *ApJ*, 378, 496
- Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, *MNRAS*, 405, 2161
- Ebeling H., Barrett E., Donovan D., Ma C.-J., Edge A. C., van Speybroeck L., 2007, *ApJ*, 661, L33
- Elliott J., de Souza R. S., Krone-Martins A., Cameron E., Ishida E. E. O., Hilbe J., 2015, *Astron. Comput.*, 10, 61
- Fabjan D., Borgani S., Tornatore L., Saro A., Murante G., Dolag K., 2010, *MNRAS*, 401, 1670
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Giocoli C., Meneghetti M., Bartelmann M., Moscardini L., Boldrin M., 2012, *MNRAS*, 421, 3343
- Goodman J., Weare J., 2010, *Appl. Math. Comput. Sci.*, 5, 65
- Gottlöber S., Yepes G., 2007, *ApJ*, 664, 117
- Hattori M., Watanabe K., Yamashita K., 1997, *A&A*, 319, 764
- Hennawi J. F., Dalal N., Bode P., Ostriker J. P., 2007, *ApJ*, 654, 714
- Ho S., White M., 2005, *Astropart. Phys.*, 24, 257
- Hogg D. W., Bovy J., Lang D., 2010, preprint ([arXiv:1008.4686](https://arxiv.org/abs/1008.4686))
- Horesh A., Maoz D., Ebeling H., Seidel G., Bartelmann M., 2010, *MNRAS*, 406, 1318
- Horesh A., Maoz D., Hilbert S., Bartelmann M., 2011, *MNRAS*, 418, 54
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Ishida E. E. O., Vitenti S. D. P., Penna-Lima M., Cisewski J., de Souza R. S., Trindade A. M. M., Cameron E., Busti V. C., 2015, *Astron. Comput.*, 13, 1
- Jabot F., Lagarrigues G., Courbaud B., Dumoulin N., 2014, preprint ([arXiv:1412.7560](https://arxiv.org/abs/1412.7560))
- Jenkins C. R., Peacock J. A., 2011, *MNRAS*, 413, 2895
- Johnson T. L., Sharon K., Bayliss M. B., Gladders M. D., Coe D., Ebeling H., 2014, *ApJ*, 797, 48
- Killedar M., Borgani S., Meneghetti M., Dolag K., Fabjan D., Tornatore L., 2012, *MNRAS*, 427, 533
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Kravtsov A. V., Borgani S., 2012, *ARA&A*, 50, 353
- Kwan J., Bhattacharya S., Heitmann K., Habib S., 2013, *ApJ*, 768, 123
- Lewis G. F., Babul A., Katz N., Quinn T., Hernquist L., Weinberg D. H., 2000, *ApJ*, 536, 623
- Li G.-L., Mao S., Jing Y. P., Bartelmann M., Kang X., Meneghetti M., 2005, *ApJ*, 635, 795
- Limousin M. et al., 2010, *MNRAS*, 405, 777
- Limousin M., Morandi A., Sereno M., Meneghetti M., Ettori S., Bartelmann M., Verdugo T., 2013, *Space Sci. Rev.*, 177, 155
- Lin C.-A., Kilbinger M., 2015, *A&A*, 583, A70
- McCarthy I. G. et al., 2010, *MNRAS*, 406, 822
- Mahdavi A., Hoekstra H., Babul A., Sievers J., Myers S. T., Henry J. P., 2007, *ApJ*, 664, 162
- Mahdi H. S., van Beek M., Elahi P. J., Lewis G. F., Power C., Killedar M., 2014, *MNRAS*, 441, 1954
- Mantz A., Allen S. W., Ebeling H., Rapetti D., Drlica-Wagner A., 2010, *MNRAS*, 406, 1773
- Mantz A. B., Allen S. W., Morris R. G., Schmidt R. W., von der Linden A., Urban O., 2015, *MNRAS*, 449, 199
- Marshall P. J., Hobson M. P., Slosar A., 2003, *MNRAS*, 346, 489
- Mead J. M. G., King L. J., Sijacki D., Leonard A., Puchwein E., McCarthy I. G., 2010, *MNRAS*, 406, 434
- Meneghetti M., Fedeli C., Pace F., Gottlöber S., Yepes G., 2010, *A&A*, 519, A90
- Meneghetti M., Fedeli C., Zitrin A., Bartelmann M., Broadhurst T., Gottlöber S., Moscardini L., Yepes G., 2011, *A&A*, 530, A17
- Meneghetti M., Bartelmann M., Dahle H., Limousin M., 2013, *Space Sci. Rev.*, 177, 31
- Meneghetti M. et al., 2014, *ApJ*, 797, 34
- Merten J. et al., 2015, *ApJ*, 806, 4
- Miralda-Escude J., 1993, *ApJ*, 403, 497
- Morandi A., Pedersen K., Limousin M., 2010, *ApJ*, 713, 491
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Oguri M., Blandford R. D., 2009, *MNRAS*, 392, 930
- Oguri M., Taruya A., Suto Y., 2001, *ApJ*, 559, 572
- Park M., Jitkrittum W., Sejdinovic D., 2015, preprint ([arXiv:1502.02558](https://arxiv.org/abs/1502.02558))
- Perrenod S. C., 1980, *ApJ*, 236, 373
- Planelles S., Quilis V., 2009, *MNRAS*, 399, 410
- Planelles S., Borgani S., Fabjan D., Killedar M., Murante G., Granato G. L., Ragone-Figueroa C., Dolag K., 2014, *MNRAS*, 438, 195
- Postman M. et al., 2012, *ApJS*, 199, 25
- Puchwein E., Hilbert S., 2009, *MNRAS*, 398, 1298
- Puchwein E., Bartelmann M., Dolag K., Meneghetti M., 2005, *A&A*, 442, 405
- Ragone-Figueroa C., Granato G. L., Murante G., Borgani S., Cui W., 2013, *MNRAS*, 436, 1750
- Rasia E., Borgani S., Ettori S., Mazzotta P., Meneghetti M., 2013, *ApJ*, 776, 39
- Redlich M., Bartelmann M., Waizmann J.-C., Fedeli C., 2012, *A&A*, 547, A66
- Reiprich T. H., Böhringer H., 1999, *Astron. Nachr.*, 320, 296
- Reiprich T. H., Böhringer H., 2002, *ApJ*, 567, 716
- Ritchie B. W., Thomas P. A., 2002, *MNRAS*, 329, 675
- Robin A. C., Reylé C., Fliri J., Czekaj M., Robert C. P., Martins A. M. M., 2014, *A&A*, 569, A13
- Rozo E., Nagai D., Keeton C., Kravtsov A., 2008, *ApJ*, 687, 22
- Schneider P., Ehlers J., Falco E. E., 1992, *Gravitational Lenses*. Springer-Verlag, Berlin
- Sereno M., Umetsu K., 2011, *MNRAS*, 416, 3187
- Sereno M., Zitrin A., 2012, *MNRAS*, 419, 3280
- Sereno M., Ettori S., Umetsu K., Baldi A., 2013, *MNRAS*, 428, 2241
- Sereno M., Giocoli C., Ettori S., Moscardini L., 2015, *MNRAS*, 449, 2024
- Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, *MNRAS*, 380, 877
- Sivia D. S., 1996, *Data Analysis: A Bayesian Tutorial*. Oxford Univ. Press, Oxford
- Spergel D. N. et al., 2003, *ApJS*, 148, 175
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Springel V., Di Matteo T., Hernquist L., 2005a, *MNRAS*, 361, 776
- Springel V. et al., 2005b, *Nature*, 435, 629
- Takahashi R., Chiba T., 2001, *ApJ*, 563, 489
- Teyssier R., Moore B., Martizzi D., Dubois Y., Mayer L., 2011, *MNRAS*, 414, 195
- Tormen G., Bouchet F. R., White S. D. M., 1997, *MNRAS*, 286, 865
- Tornatore L., Borgani S., Dolag K., Matteucci F., 2007, *MNRAS*, 382, 1050
- Trotta R., 2008, *Contemporary Phys.*, 49, 71
- Umetsu K. et al., 2015, *ApJ*, 806, 207

- Umetsu K., Zitrin A., Gruen D., Merten J., Donahue M., Postman M., 2016, *ApJ*, 821, 116
- van den Bosch F. C., 2002, *MNRAS*, 331, 98
- Waizmann J.-C., Redlich M., Meneghetti M., Bartelmann M., 2014, *A&A*, 565, A28
- Wambsganss J., Bode P., Ostriker J. P., 2004, *ApJ*, 606, L93
- Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, *ApJ*, 568, 52
- Weyant A., Schafer C., Wood-Vasey W. M., 2013, *ApJ*, 764, 116
- Wilkinson R. D., 2013, *Stat. App. Gen. Mol. Bio.*, 12, 129
- Wu X., Hammer F., 1993, *MNRAS*, 262, 187
- Wurster J., Thacker R. J., 2013, *MNRAS*, 431, 2513
- Zemp M., Gnedin O. Y., Gnedin N. Y., Kravtsov A. V., 2011, *ApJS*, 197, 30
- Zheng W. et al., 2012, *Nature*, 489, 406
- Zitrin A., Broadhurst T., Barkana R., Rephaeli Y., Benítez N., 2011, *MNRAS*, 410, 1939
- Zitrin A. et al., 2015, *ApJ*, 801, 44

## APPENDIX A: FITTING TO THE LENSING–MASS SCALING RELATION

We employ a Bayesian fitting procedure to determine the summary statistics,  $\alpha$  and  $\beta$ , parameters of the scaling relation between strong lensing efficiency and total cluster mass (equation 7). We also acknowledge that there is likely to be intrinsic scatter in this relationship directly comparable to the scatter in the concentration–mass relation, partly due to cluster triaxiality and substructure and partly from the varying formation histories of the clusters. Thus we also include a nuisance parameter,  $V$ , which represents intrinsic Gaussian variance, orthogonal to the line.

For this appendix, we change notation in order to reduce the subscripts: the mass of the  $i$ th cluster lens as  $M_i$ , and the scaled Einstein radius – however characterized – as  $E_i$ . Each data point is denoted by the vector  $\mathbf{Z}_i = [\log M_i, \log E_i]$ . Their respective uncertainties (on the logarithms) are denoted  $\sigma_M^2$  and  $\sigma_E^2$ . Since we assume the uncertainties for Einstein radii and cluster mass are uncorrelated, the covariance matrix,  $\mathbf{S}_i$ , reduces to

$$\mathbf{S}_i \equiv \begin{pmatrix} \sigma_M^2 & 0 \\ 0 & \sigma_E^2 \end{pmatrix}. \quad (\text{A1})$$

In the case of a mock sample of simulated clusters,  $\mathbf{S}_i = 0$ . We expect uncertainties in mass due to the choice of halo finder and Einstein radii (see our method describe in Section 4.3) to be negligible compared to the overall scatter.

Consider now the following quantities:  $\varphi \equiv \arctan \alpha$ , which denotes the angle between the line and the  $x$ -axis, and  $b_\perp \equiv \beta \cos \varphi$ , which is the orthogonal distance of the line to the origin. The orthogonal distance of each data point to the line is

$$\Delta_i = \hat{\mathbf{v}}^\top \mathbf{Z}_i - \beta \cos \varphi, \quad (\text{A2})$$

where  $\hat{\mathbf{v}} = [-\sin \varphi, \cos \varphi]$  is a vector orthogonal to the line.

Therefore, the orthogonal variance is

$$\Sigma_i^2 = \hat{\mathbf{v}}^\top \mathbf{S}_i \hat{\mathbf{v}}. \quad (\text{A3})$$

Following Hogg et al. (2010, see their equation 35), we calculate the (logarithm of the) likelihood over the three-dimensional parameter space  $\Theta_1 \equiv \{\alpha, \beta, V\}$  given the data  $D$ , which includes  $\mathbf{Z}$  and  $\mathbf{S}$ :

$$\ln \mathcal{L}(\Theta_1; D) = K - \sum_{i=1}^N \frac{1}{2} \ln(\Sigma_i^2 + V) - \sum_{i=1}^N \frac{\Delta_i^2}{2\Sigma_i^2 + V}, \quad (\text{A4})$$

where  $K$  is an arbitrary constant, and the summation is over all clusters in the considered sample.

While we ultimately (aim to) provide the parameter constraints on  $\alpha$  and  $\beta$ , flat priors for these tend to unfairly favour large slopes. A more sensible choice is flat for the alternative parameters  $\varphi$  and  $b_\perp$ . We apply a modified Jeffreys prior on  $V$ :

$$\pi(V) \propto \frac{1}{V + V_1}. \quad (\text{A5})$$

This is linearly uniform on  $V$  for small values and logarithmically uniform on  $V$  for larger values with a turnover,  $V_1$ , chosen to reflect the typical uncertainties.

Thus, for each  $\Theta_1$ , we may define an alternative set of parameters  $\Theta_2 \equiv \{\varphi, b_\perp, V\}$ , for which the prior is given by

$$\begin{aligned} \pi(\Theta_2) &= \pi(\varphi, b_\perp) \pi(V) \\ &\propto \pi(V), \end{aligned} \quad (\text{A6})$$

where  $\pi(V)$  is given by equation (A5). The prior on  $\Theta_1$  is then dependent on the magnitude of the Jacobian of the mapping between the two sets of parameters:

$$\begin{aligned} \pi(\Theta_1) &= \pi(\Theta_2) \det \frac{\partial \Theta_2}{\partial \Theta_1} \\ &\equiv \pi(\Theta_2) \frac{1}{(1 + \alpha^2)^{3/2}}. \end{aligned} \quad (\text{A7})$$

Boundaries on the priors are sufficiently large:<sup>15</sup>  $-8 \leq \beta \leq 8$ ;  $-40 \leq \alpha \leq 40$ ;  $0 \leq V \leq V_{\max}$ .  $V_{\max}$  is chosen to reflect the overall scatter in the data. The posterior is calculated following Bayes' theorem:

$$P(\Theta_1|D) \propto \mathcal{L}(\Theta_1; D) \pi(\Theta_1) \quad (\text{A8})$$

and is normalized. Since we are interested in the constraints on  $\alpha$  and  $\beta$ , we then marginalize over the nuisance parameter,  $V$ .

In practice, the posterior distribution was sampled by employing EMCEE (Foreman-Mackey et al. 2013), the PYTHON implementation of the affine-invariant ensemble sampler for Markov chain Monte Carlo (MCMC) proposed by Goodman & Weare (2010). The proposal distribution took the form of a small Gaussian ‘ball’ centred on the expected peak in the distribution as determined in earlier test runs. These tests also found that the autocorrelation times tend to be between 10 and 60 steps for each parameter. For both the data and each of the 80 mock samples, we ran 120 walkers for 600 steps each. In all cases, the 120 initial steps were considered to encompass the ‘burn-in’ phase and were discarded.

## APPENDIX B: TRIAXIALITY AND ENCLOSED MASS

Given a numerically simulated DM halo for which the underlying mass distribution is described by the distribution of discrete particles, its shape can be determined with the following iterative procedure, based on Dubinski & Carlberg (1991); see also a detailed discussion in Zemp et al. (2011).

- (i) Compute and diagonalize the inertia shape tensor:

$$S_{ij} \equiv \frac{\sum_k m_k w_k x_{k,i} x_{k,j}}{\sum_k m_k}, \quad (\text{B1})$$

where  $m_k$  is the mass of the  $k$ th particle,  $w_k$  is a weighting associated with that particle and  $x_{k,i}$  is the  $i$  component of the particle position vector. The summation is over all particles chosen for analysis.

<sup>15</sup> The physically motivated choice of restricting  $\alpha \geq 0$  is also explored, however, this has very minor effects on the final results despite removing the (small) secondary peak in the marginal posterior on  $\alpha$  and  $\beta$ .

- (ii) Identify the principal axes as the eigenvectors of  $S_{ij}$ .
- (iii) Calculate the eigenvalues of  $S_{ij}$  ( $\lambda_1 > \lambda_2 > \lambda_3$ ).
- (iv) Determine the axis ratios using

$$\frac{\lambda_3}{\lambda_1} = s^2 \quad \text{and} \quad \frac{\lambda_2}{\lambda_1} = q^2. \quad (\text{B2})$$

(v) Identify a new set of particles for analysis by calculating  $r_e$  for each particle based on the newly determined  $q$  and  $s$ . If  $r_e$  is less than a pre-determined  $r_e^*$ , the particle is included in the integration volume.

(vi) Repeat steps (i)–(iv) until both  $q$  and  $s$  pass convergence criteria or until a maximum number of iterations.

The choice of  $r_e^*$  determines the scale of the ellipsoidal volume at each iteration. We perform the iterative procedure outlined above with  $r_e^*$  defined such that the enclosed mass has a fixed average density of  $\Delta$  times the critical density. Note that numerous works in the literature scale the ellipsoid at each iteration by anchoring the major axis to a fixed radius instead, but while our approach is more numerically demanding, this criterion better reflects the required information about cluster shape (see also Despali, Tormen & Sheth 2013; Bonamigo et al. 2015). We repeat this process two times for each cluster, once for  $\Delta = 200$  and then  $\Delta = 2500$ .

This paper has been typeset from a  $\text{T}_\text{E}\text{X}/\text{L}^\text{A}\text{T}_\text{E}\text{X}$  file prepared by the author.