

MACS J0416.1–2403: Impact of line-of-sight structures on strong gravitational lensing modelling of galaxy clusters

G. Chirivì¹, S. H. Suyu^{1,2,3}, C. Grillo^{4,5}, A. Halkola¹¹, I. Balestra^{6,7}, G. B. Caminha^{8,9}, A. Mercurio¹⁰, and P. Rosati⁸

¹ Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Str. 1, 85741 Garching, Germany
e-mail: chirivig@mpa-garching.mpg.de

² Institute of Astronomy and Astrophysics, Academia Sinica, 23-141, Taipei 10617, Taiwan

³ Physik-Department, Technische Universität München, James-Franck-Straße 1, 85748 Garching, Germany

⁴ Dipartimento di Fisica, Università degli Studi di Milano, via Celoria 16, 20133 Milano, Italy

⁵ Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark

⁶ University Observatory Munich, Scheinerstrasse 1, 81679 Munich, Germany

⁷ INAF – Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, 34143 Trieste, Italy

⁸ Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, via Saragat 1, 44122 Ferrara, Italy

⁹ Osservatorio di Bologna INAF – Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy

¹⁰ INAF – Osservatorio Astronomico di Capodimonte, via Moiariello 16, 80131 Napoli, Italy

¹¹ Pyörrekuja 5 A, 04300 Tuusula, Finland

Received 23 June 2017 / Accepted 11 December 2017

ABSTRACT

Exploiting the powerful tool of strong gravitational lensing by galaxy clusters to study the highest-redshift Universe and cluster mass distributions relies on precise lens mass modelling. In this work, we aim to present the first attempt at modelling line-of-sight (LOS) mass distribution in addition to that of the cluster, extending previous modelling techniques that assume mass distributions to be on a single lens plane. We have focussed on the *Hubble* Frontier Field cluster MACS J0416.1–2403, and our multi-plane model reproduces the observed image positions with a rms offset of $\sim 0''.53$. Starting from this best-fitting model, we simulated a mock cluster that resembles MACS J0416.1–2403 in order to explore the effects of LOS structures on cluster mass modelling. By systematically analysing the mock cluster under different model assumptions, we find that neglecting the lensing environment has a significant impact on the reconstruction of image positions (rms $\sim 0''.3$); accounting for LOS galaxies as if they were at the cluster redshift can partially reduce this offset. Moreover, foreground galaxies are more important to include into the model than the background ones. While the magnification factor of the lensed multiple images are recovered within $\sim 10\%$ for $\sim 95\%$ of them, those $\sim 5\%$ that lie near critical curves can be significantly affected by the exclusion of the lensing environment in the models. In addition, LOS galaxies cannot explain the apparent discrepancy in the properties of massive sub-halos between MACS J0416.1–2403 and N-body simulated clusters. Since our model of MACS J0416.1–2403 with LOS galaxies only reduced modestly the rms offset in the image positions, we conclude that additional complexities would be needed in future models of MACS J0416.1–2403.

Key words. gravitational lensing: strong – galaxies: clusters: general – galaxies: clusters: individual: MACS J0416.1–2403 – dark matter

1. Introduction

Massive galaxy clusters are the largest gravitationally bound structures in the Universe and they are located at the nodes of the cosmic web. According to the currently accepted cosmological model, which consists of a cold dark matter dominated Universe with a cosmological constant (Λ CDM), more massive structures form by accretion and assembly of smaller self-bound individual systems (e.g. [Springel et al. 2006](#)). As such, galaxy clusters are not only a perfect laboratory to study the formation and evolution of structures in the Universe (e.g. [Dressler 1984](#); [Kravtsov & Borgani 2012](#), and references therein), but also to study the mass-energy density components, such as dark matter and dark energy, and to constrain cosmological parameters (e.g. [Jullo et al. 2010](#); [Caminha et al. 2017](#); [Rozo et al. 2010](#); [Planck Collaboration XX 2014](#), among others).

They are also very efficient cosmic telescopes. Indeed, the magnification effect produced by gravitational lensing with galaxy clusters provides a powerful tool to detect and study

high-redshift galaxies, that would be undetectable with currently available instruments. Gravitational lensing is a relativistic effect for which the light travelling from a source towards the observer is bent by the presence of matter in-between. Consequently, the source is observed at a different position than it actually is, distorted in shape, and in some cases also multiply imaged (in the so-called strong lensing regime). It also appears magnified by a factor μ (magnification). Once the magnification μ is known, due to surface brightness conservation, the intrinsic brightness and shape of the source galaxy can be reconstructed. This allows study of high-redshift galaxies, providing crucial probes of structure formation and galaxy evolution.

In the past decades, gravitational lensing with clusters has highly improved our knowledge of the mass distribution in clusters, and has led to the discovery of some of the highest-redshift galaxies (e.g. [Coe et al. 2013](#); [Bouwens et al. 2014](#)). Being such versatile laboratories for many studies, galaxy clusters have been searched for and studied in depth in recent years. In particular, the *Hubble* Frontier Fields initiative (HFF; P.I.: J. Lotz)

has exploited the *Hubble* Space Telescope (HST) sensitivity and the magnification effect of six targeted strong lensing clusters to detect the high-redshift Universe. These six clusters were observed in seven optical and near-IR bands using the Advanced Camera for Survey (ACS) and the Wide Field Camera 3 (WFC3), achieving an unprecedented depth of ~ 29 mag (AB). The HFF initiative has provided a sample of high-redshift galaxies that will allow one to investigate their properties in a statistically significant way.

As previously mentioned, to reconstruct the intrinsic brightness of these far away sources, the magnification effect needs to be accounted for. This can be done only by modelling in detail the mass distribution of the lensing cluster. High precision reconstruction of cluster lenses is necessary to avoid systematic errors affecting the cluster mass reconstruction or the source brightness substantially. In recent years, the models of clusters of galaxies have been performed with increasing precision (e.g. Grillo et al. 2015; Caminha et al. 2017, among others). However, current analyses using single-plane approach seem to have reached a limit in reproducing the observables. This approach consists of modelling the lens as if it were an isolated system on a single redshift-plane, and has led to models with an rms between the observed and predicted positions of the multiple images of the strongly lensed background sources of $\sim 0''.5-1''$, which is greater than the observational uncertainty ($\sim 0''.05$). To account for the residual offset, it might be necessary to also consider the lensing environment in the model. In fact, we expect that the contribution of the line-of-sight (LOS) matter can affect the observables more than the observational uncertainties. Therefore, this effect needs to be taken into account to study its contribution to the recovered offset between observations and current cluster models (e.g. D'Aloisio et al. 2014; Caminha et al. 2016).

In this work we have simulated galaxy cluster lensing observations, accounting for the LOS effects using multi-plane lensing formalism, and study the effects of the LOS structures on the magnification and position of the images. We then analysed these simulated data to study the cluster mass distribution and to reconstruct the high-redshift galaxies' intrinsic brightness. We analysed toy models of clusters, and then a mock model that is as similar as possible to the HFF cluster MACS J0416.1–2403, to make our study more realistic.

The work is organised as follows: in Sect. 2 we describe our analysis' method, the multi-plane formalism, and the profiles we use to parametrise the lens cluster. In Sect. 3 we analyse toy models of simplistic clusters and in Sect. 4 we present our model of the mass distribution of a real cluster, namely the HFF cluster MACS J0416.1–2403. We used our best-fit model to generate lensing observables from a simulated mock system that mimics MACS J0416.1–2403 and we model the simulated cluster with different assumptions. We discuss our results in Sect. 5.

Throughout this work, we have assumed a flat Λ -CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega_\Lambda = 1 - \Omega_M = 0.70$. From the redshift of the lens ($z_c = 0.396$) in Sect. 4.1, one arc-second at the lens plane in MACS J0416.1–2403 corresponds to ~ 5.34 kpc. The magnitudes are all given in the AB system. The uncertainty we have considered for all the image positions corresponds to the pixel scale $0''.065$, unless stated otherwise.

2. Multiple lens plane modelling

In this work, the modelling was obtained using GLEE, a software developed by A. Halkola and S. H. Suyu (Suyu & Halkola 2010; Suyu et al. 2012). This software uses parametrised mass profiles to describe the dark matter halos and the galaxies and a Bayesian

analysis to infer the best-fit parameter values and their variances and degeneracies. It also includes the possibility of considering lenses at different redshifts (multiple lens plane modelling; Suyu et al. in prep.). In the following sections we introduce the multiple lens plane formalism (Sect. 2.1), we describe the lens profiles that we use for the lens galaxies and dark matter halos (Sect. 2.2) and the scaling relations we used for the cluster members and the LOS galaxies (Sect. 2.3). We discuss how we determine the model parameters using a Bayesian approach in Sect. 2.4.

2.1. Multi-plane formalism

In this section we briefly revisit generalised multi-plane gravitational lens theory (e.g. Blandford & Narayan 1986; Schneider et al. 1992), which considers the fact that a light ray can be bent multiple times by several deflectors during its path. This theory takes into account the effect of secondary lenses at different redshifts using the thin-lens approximation for every deflector on its redshift plane. The lens equation in this formalism is (following Gavazzi et al. 2008)

$$\theta_j(\theta_1) = \theta_1 - \sum_{i=1}^{j-1} \frac{D_{ij}}{D_j} \alpha(\theta_i), \quad (1)$$

where θ_1 is the image position on the first plane (observed image plane), θ_j is the image position on the j th plane and $\alpha(\theta_i)$ is the deflection angle on the i th plane. In this recursive equation the deflection angle on one plane depends on the deflection angle of all the previous planes. The source position β , which is on the N th plane, corresponds to

$$\beta = \theta_N(\theta_1) = \theta_1 - \sum_{i=1}^{N-1} \frac{D_{iN}}{D_N} \alpha(\theta_i). \quad (2)$$

Therefore the total scaled deflection angle α_{tot} is the sum of all the deflection angles on all planes, namely

$$\alpha_{\text{tot}} = \sum_{i=1}^{N-1} \frac{D_{iN}}{D_N} \alpha(\theta_i). \quad (3)$$

The magnification in the multi-plane case is calculated from the Jacobian matrix obtained by differentiating the total scaled deflection angle α_{tot} , namely

$$\mathbf{A} = \frac{\partial \beta}{\partial \theta} = \frac{\partial \theta_N}{\partial \theta_1} = \mathbf{I} - \frac{\partial \alpha_{\text{tot}}}{\partial \theta_1}, \quad (4)$$

the magnification is

$$\mu = \frac{1}{\det(\mathbf{A})}. \quad (5)$$

To compute the average surface mass density $\Sigma(<R)$, we obtain the convergence κ from GLEE and multiply it by the critical surface mass density Σ_{crit} using the definition of convergence,

$$\kappa = \frac{\Sigma}{\Sigma_{\text{crit}}}, \quad (6)$$

where Σ_{crit} is

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_d D_{ds}}. \quad (7)$$

We then computed the average¹ surface mass density, namely

$$\Sigma(<R) = \frac{\int_0^R \Sigma(R') 2\pi R' dR'}{\pi R^2}. \quad (8)$$

¹ Average circularised, since we integrated all the mass distributions over circles.

In the multi-plane case, these quantities need to be properly defined using the multi-plane formalism: the deflection angle which we obtained the multi-plane κ from is the α_{tot} expressed by Eq. (3). The quantity derived from differentiating this α_{tot} is what we call the “effective” convergence κ_{Eff} . We then multiplied this quantity by Σ_{crit} to obtain the effective average surface mass density $\Sigma_{\text{Eff}}(< R)$. Therefore, this quantity is not a physical surface density, but the gradient of the total scaled deflection angle α_{tot} , which receives contribution from all mass components on all the planes.

2.2. Lens mass distributions

In GLEE, we used parametric mass profiles to portray the cluster members component, the LOS perturber galaxies component, and the contribution of the remaining intra-cluster mass (mainly dark matter) respectively. We modelled the luminous mass component (i.e. members and LOS galaxies) with a truncated dual pseudo-isothermal elliptical mass distribution (dPIE; Elíasdóttir et al. 2007; Suyu & Halkola 2010) with vanishing core radius. Their dimensionless projected surface mass density – convergence – is

$$\kappa_{\text{dPie}}(x, y) \Big|_{z_s=\infty} = \frac{\theta_E}{2} \left(\frac{1}{R_{\text{em}}} - \frac{1}{\sqrt{R_{\text{em}}^2 + r_t^2}} \right), \quad (9)$$

where (x, y) are the co-ordinates on the lens plane, θ_E is the lens Einstein radius, r_t is the truncation radius. The mass distribution is then suitably rotated by its orientation angle θ and shifted by the centroid position of the co-ordinate system used. The convergence for a given source redshift z_s is

$$\kappa(x, y) \Big|_{z_s} = \kappa(x, y) \Big|_{z_s=\infty} \frac{D_{\text{ds}}}{D_s}. \quad (10)$$

The 2D elliptical mass radius is

$$R_{\text{em}} = \sqrt{\frac{x^2}{(1+e)^2} + \frac{y^2}{(1-e)^2}}, \quad (11)$$

and the ellipticity is

$$e = \frac{1-q}{1+q} \quad (12)$$

where q is the axis ratio. Therefore, the parameters that identify this profile are its central position (x_c, y_c) , its axis ratio q , its orientation θ , its strength θ_E , and its truncation radius r_t .

The 3D density corresponding to this κ is

$$\rho_{\text{dPie}}(r) \propto \left(r^2 (r^2 + r_t^2) \right)^{-1}, \quad (13)$$

where r is the 3D radius. As we can see, for r greater than the truncation radius, this density distribution is truncated, that is it scales as r^{-4} . For isothermal profiles there is a direct relation between the value of the velocity dispersion σ of the lens and that of its strength θ_E (for a source at $z_s = \infty$), namely

$$\frac{\sigma}{c} = \sqrt{\frac{\theta_E}{4\pi}}. \quad (14)$$

To model the remaining mass of the cluster, especially the contribution of the dark matter halos, we use a 2D pseudo-isothermal elliptical mass distribution (PIEMD; Kassiola & Kovner 1993) or a softened power-law elliptical mass

distributions (SPEMD; Barkana 1998). The convergence of the PIEMD profile, is

$$\kappa_{\text{piemd}}(x, y) \Big|_{z_s=\infty} = \frac{\theta_E}{2\sqrt{R_{\text{em}}^2 + r_c^2}}, \quad (15)$$

where, again, (x, y) are the co-ordinates in the lens plane, R_{em} is the elliptical mass radius, θ_E is the lens Einstein radius and r_c is the core radius. The mass distribution is then suitably rotated by its orientation angle θ and shifted by the centroid position of the co-ordinate system used. The parameters that identify this profile are its central position (x_c, y_c) , its axis ratio q , its orientation θ , its strength θ_E , and its core radius r_c . In the case of the SPEMD, there is one additional parameter, which is the slope γ of the profile, since its convergence, is

$$\kappa_{\text{spemd}}(x, y) \Big|_{z_s=\infty} = \theta_E \left(x^2 + \frac{y^2}{q^2} + r_c^2 \right)^{-\gamma}, \quad (16)$$

where q is the axis ratio, r_c is the core radius, γ is the power law index, which is 0.5 for an isothermal profile (Barkana 1998).

2.3. Scaling relation

Throughout this work, to reduce the number of model parameters and therefore increase the computational efficiency of our modelling, we assumed scaling relations for the Einstein radii and truncation radii of the cluster members, that is we scaled them with respect to a reference galaxy, which we choose as the cluster member with median luminosity. The scaling was done following Grillo et al. (2015),

$$\theta_{E,i} = \theta_{E,g} \left(\frac{L_i}{L_g} \right)^{0.7}, \quad r_{t,i} = r_{t,g} \left(\frac{L_i}{L_g} \right)^{0.5}, \quad (17)$$

where $\theta_{E,i}$ and $r_{t,i}$ are the Einstein radius and the truncation radius of the i th galaxy with luminosity L_i , and $\theta_{E,g}$, $r_{t,g}$, L_g are the properties of the reference galaxy. This is equivalent to having cluster members with a total mass-to-light ratio $\frac{M_T}{L} \propto L^{0.2}$, also known as the tilt of the fundamental plane (Faber et al. 1987; Bender et al. 1992), since

$$\frac{M_{T,i}}{L_i} \sim \frac{\sigma_i^2 r_{t,i}}{L_i} \sim \frac{L_i^{0.7} L_i^{0.5}}{L_i} \sim L_i^{0.2}. \quad (18)$$

To test how the values of the parameters change if we assume a different scaling relation, we also use

$$\theta_{E,i} = \theta_{E,g} \left(\frac{L_i}{L_g} \right)^{0.5}, \quad r_{t,i} = r_{t,g} \left(\frac{L_i}{L_g} \right)^{0.5}. \quad (19)$$

This is equivalent to having cluster members with constant total mass-to-light ratio, since, as shown in Grillo et al. (2015)

$$\frac{M_{T,i}}{L_i} \sim \frac{\sigma_i^2 r_{t,i}}{L_i} \sim \frac{L_i^{0.5} L_i^{0.5}}{L_i} \sim L_i^0. \quad (20)$$

2.4. Determination of model parameters

GLEE uses a Bayesian analysis to infer the best-fit parameters, that is those that maximise the likelihood on the image plane (given uniform priors on the parameters), namely the probability of the data (the observed image positions \mathbf{X}_{obs} , which corresponds to θ_1 in Sect. 2.1) given a set of parameters $\boldsymbol{\eta}$ of the model. This likelihood

$$\mathcal{L}(\mathbf{X}_{\text{obs}}|\boldsymbol{\eta}) \propto \exp \left[-\frac{1}{2} \sum_{j=1}^{N_{\text{sys}}} \sum_{i=1}^{N_{\text{imj}}} \frac{|\mathbf{X}_{i,j}^{\text{obs}} - \mathbf{X}_{i,j}^{\text{pred}}(\boldsymbol{\eta})|^2}{\sigma_{i,j}^2} \right] \quad (21)$$

describes the offset between the observed $X_{i,j}^{\text{obs}}$ and predicted $X_{i,j}^{\text{pred}}$ image positions for all the images $N_{\text{im},j}$ of the observed image system j of N_{sys} , where $\sigma_{i,j}$ is the observational uncertainty of the i th image of the j th source. We used a simulated annealing technique to find the global minimum and recover the best-fit parameter values, Markov chain Monte Carlo (MCMC methods based on Dunkley et al. 2005) and EMCEE (Foreman-Mackey et al. 2013) to sample their posterior probability distributions. The posterior probability is obtained using Bayes' Theorem

$$P(\boldsymbol{\eta}|\mathbf{X}_{\text{obs}}) \propto \mathcal{L}(\mathbf{X}_{\text{obs}}|\boldsymbol{\eta})P(\boldsymbol{\eta}), \quad (22)$$

where $P(\boldsymbol{\eta})$ is the prior probability on the model parameters, that we always assume to be uniform. For the centroid position of each halo we use a broad uniform prior range, that is $\sim 60''$ from the respective BCG centre, and for the axis ratio a range from 0.2 to 1. For the Einstein and truncation radii of all the galaxies and halos we set a minimum prior value to zero, for the orientation of the halos we assumed no boundaries on the priors. The model parameter values we show throughout this work, unless stated otherwise, are the median of the one-dimensional marginalised posterior PDF, with the quoted uncertainties showing the 16th and 84th percentiles (i.e. the bounds of a 68% credible interval). To obtain realistic uncertainties for the model parameters, we ran a second MCMC analysis where we increased the image position uncertainty to roughly the value of the *rms*, to account for the model imperfections (e.g. lack of treatment for elliptical cluster members, simplistic halo models, neglecting small dark-matter clumps). These new uncertainties are those shown in the tables and used to quantify the ability of the models to recover the input values. The root mean square values are quoted for the best-fit models, and do not depend on the adopted error on the observed image positions. We computed the root mean square error via

$$rms = \sqrt{\frac{1}{N} \sum_{i=1}^N |X_{\text{pred}} - X_{\text{obs}}|^2}, \quad (23)$$

with the model-predicted image positions X_{pred} and the observed image positions X_{obs} . When the number of predicted and observed image positions of a given image system is not the same, for each observed image in that system GLEE selects the closest predicted image to compute the χ^2 of the model. We also used the closest predicted image to compute the *rms*.

3. Toy models

We begin with simple toy models of lenses at multiple redshifts in order to gain intuition for multi-plane lens modelling. We generated mock lensing observables and analysed the impact of the introduction of multi-plane lenses by fitting the same image positions with both a multi-plane-lens and a single plane-lens model.

3.1. Two lenses at different redshifts

We started our analysis with a simple toy model composed of two lenses (modelled as SISs and denoted respectively as SIS1 and SIS2), at redshifts $z_{\text{SIS1}} = 0.5$, $z_{\text{SIS2}} = 0.7$ aligned along the LOS and with Einstein radii of $2''.6$ and $1''.6$ respectively, and a point source at redshift $z_s = 2$, as shown in Fig. 1. We studied different configurations, in terms of lens location and lensed image positions, and tested how well, for the same simulated observed

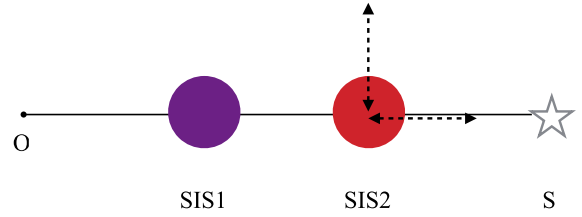


Fig. 1. Positional setup of the two lenses. The first lens (purple) is at $z_{\text{SIS1}} = 0.5$, the second lens (red) at $z_{\text{SIS2}} = 0.7$. Both the lenses are SISs. The dotted arrows indicate the direction in which we shift the second lens to experiment the effects on the image positions and magnification.

image positions, we were able to reconstruct the properties of the system using a single lens at $z_{\text{sd}} = 0.5$ whose centre is coincident with the input position of SIS1.

3.1.1. Variation of mass of SIS2

We find that the variation of the mass of SIS2 still allows us to fit the image and source positions and reconstruct the magnification with a single lens, if the two lenses are aligned. Expectedly, we obtain an Einstein radius of the single deflector $\theta_{\text{E,sd}}$ which increases as we increase the mass of the second lens SIS2. The image positions, the magnifications and the source position predicted by the single-lens models can reproduce the input values to the 4th-digit precision.

3.1.2. Variation of position from optical axis of SIS2

Variations in the position of SIS2 with respect to the optical axis (we choose a range of $0''.2$ – $1''.5$) still allow a perfect fit to the same image positions with a single lens (with a 4th-digit precision). The Einstein radius of the single lens $\theta_{\text{E,sd}}$ is unchanged with respect to changes of SIS2's distance from the optical axis (we kept the Einstein radius of the SIS2 fixed, while its distance to the optical axis is increased). This is probably a result of the particular properties of the SIS profile, where the deflection angle is independent of the location. Moreover, we chose a range where the maximum shift of position of SIS2 is lower than the Einstein radius of SIS1 ($\theta_{\text{E,SIS1}} = 2''.6$). However, we find that the magnification of the multi-plane system decreases as SIS2 moves further away from the optical axis (drops from ten to six for the image with positive parity and from eight to 0.3 for the image with negative parity), while the magnification of the single-plane system stays constant, since $\theta_{\text{E,sd}}$ is itself unchanged.

3.1.3. Variation of redshift of SIS2

Finally, if we shift SIS2 on the optical axis to higher redshifts, we observe that $\theta_{\text{E,sd}}$ decreases, while the magnification increases accordingly for both the multi-plane and single plane lens system, as shown in Fig. 2. Moreover, the predicted source position moves closer to the optical axis when the SIS2 moves to higher redshifts and when the Einstein radius of the single lens decreases.

3.2. Mock cluster lensing mass distribution 2

3.2.1. Input

We now create a more realistic mock system, composed by a cluster at $z_c = 0.4$ with a halo and ten elliptical galaxies

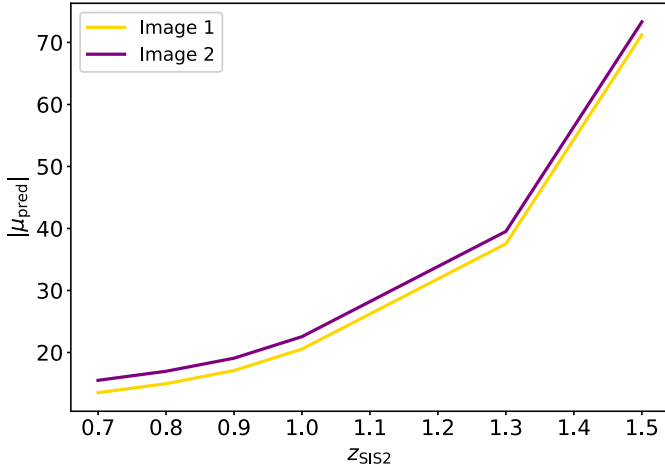


Fig. 2. Absolute value of the magnification for the two images of a source at redshift $z_s = 2$ as a function of redshift of SIS2. The lens SIS1 is kept fixed at redshift $z_{\text{SIS1}} = 0.5$ and SIS2 is free to move along the optical axis within redshift $z_{\text{SIS2}} = 0.8$ and $z_{\text{SIS2}} = 1.5$. The predicted magnification shown here corresponds to the multi-plane model, but we observe the same trend for the model with a single lens.

having different, realistic luminosities, axis ratios and orientations. We assume a total mass-to-light ratio corresponding to the tilt of the fundamental plane, and we scaled the mass and the truncation radii according to Eq. (17). We added two foreground perturbers at $z_{\text{fd}} = 0.2$, one close and one far away in projection from the cluster centre, and one close-in-projection background perturber at $z_{\text{bd}} = 0.6$. All the perturbers are massive (Einstein radii of $2''$) and have the same Einstein radii, random ellipticity (between 0.6 and 1) and orientation, as shown in Table 1. We adopted a SPEMD profile for the halo, and dPIEs for the cluster members and LOS galaxies. We used this configuration to simulate mock lensing data, and we obtained a set of 16 multiple image positions of the three background sources, shown in Fig. 3. We then modelled the parameters, that is all the halo parameters mentioned in Sect. 2.2, the cluster members' Einstein and truncation radii (with the scaling relation in Eq. (17) unless otherwise stated) and perturbers' Einstein radius (truncation radius was fixed to $15''$), with both the multi-plane set-up and the single-plane set-up (i.e. the cluster only).

3.2.2. Full multi-lens-plane modelling

In the full multi-lens-plane model, since it has the same properties as the input (i.e. perturbers, scaling relation for cluster members etc.), we recover, within the errors, the initial parameters we have used to simulate, as shown in the MP-full column of Table 1. The modelled image positions and the magnifications are perfectly fitted, meaning that they have a null total-rms offset.

3.2.3. Single cluster-plane modelling

The single plane case, as shown in column SP in Table 1, shows a shift of the halo centroid position of $\sim 1''.4$, due to the removal of the foreground perturber which was lensing it, an overestimation of the halo Einstein radius of $\sim 3''$. The halo's profile slope is less peaky in the centre, and it has a core radius that is $\sim 2''.5$ bigger. In terms of image positions and magnification, we find that the image positions' total rms offset is $\sim 0''.55$, and the magnification is generally greater, up to three times higher.

3.2.4. Mock cluster model 2: Assuming constant total mass-to-light ratio

In this experiment, we tested how the parameters change if we modelled assuming a different total mass-to-light relation. We chose a constant total mass-to-light ratio, which scales the Einstein and truncation radii of the cluster members as shown in Eq. (19). If we model the observables with the multi-plane system, we find, as shown in column MP-constML in Table 1, that the halo centroid position is shifted by $\sim 0.2''$. Even if the lensing effects of the foreground galaxy are still present, the halo is slightly more elliptical and the slope is slightly bigger. The cluster members have instead a bigger Einstein radius and a smaller truncation radius. All the remaining parameters are recovered within the errors. Modelling with the single-plane (SP-constML in Table 1) instead changes the halo centre by $\sim 1''.5$, the halo Einstein radius by $\sim 2''.5$, which is less than what we recover in the single-plane with the previous mass-to-light ratio assumption, and underestimates the cluster galaxies masses. This model also predicts a shallower halo profile.

In terms of image positions, we find that the total rms for the multi-plane system is $\sim 0''.12$ and the magnification is generally smaller, within a ratio of ~ 0.6 – 0.8 , showing that this approximation works slightly worse than the original total mass-to-light ratio used to simulate. The single-plane system, instead, has a total rms of $\sim 0''.5$, which appears to be less than what we recover using as total mass-to-light the tilt of the fundamental plane. This shows that, in the single-plane case, this approximation works slightly better than the standard SP. This is also confirmed by the magnification, which is still higher, but only up to twice as high. This unexpected behaviour might be explained by the fact that, as shown in Fig. 4, assuming a constant total mass-to-light ratio implies, given the same Einstein radius of the reference member, bigger Einstein radii for the other members (radius of circles in Fig. 4 proportional to Einstein radius). This is compensated by a decrease of the halo Einstein radius, that is, closer to its input value, to preserve the total mass within the total Einstein radius of the cluster. Probably a closer Einstein radius value for the halo is more efficient in reproducing the overall effect on image positions and magnification.

3.2.5. Mock cluster model 2: Assuming spherical galaxies

The next test is to model all the galaxies (members and LOS galaxies) as if they were spherical. The multi-plane system modelling, as shown in the column MP-s in Table 1, shows an offset of the halo centroid position of $\sim 0''.7$, and an underestimation of its Einstein radius of $\sim 3''$. The cluster galaxies and the background perturber have a bigger Einstein radii, while the foreground close-in-projection perturber has a smaller Einstein radii than the original parameter used to simulate. The image position rms is close to the observational uncertainties of $\sim 0''.065$, and the magnification is generally smaller, within a factor of 0.7 – 1 . This shows that, despite the low rms, assuming spherical galaxies creates quite a substantial offset in centroid position and Einstein radius. This suggests we should be cautious in interpreting the reconstructed cluster mass distribution, since having small rms, even comparable to positional uncertainty, does not guarantee unbiased recovery of the lens parameters. However, since our model is very simplistic (i.e. few cluster members and very massive perturbers along the LOS), we suspect these results might not be so prominent in more realistic cases where there are many more galaxies and the ellipticity averages out. In the single-plane case, we have that the parameters are consistent,

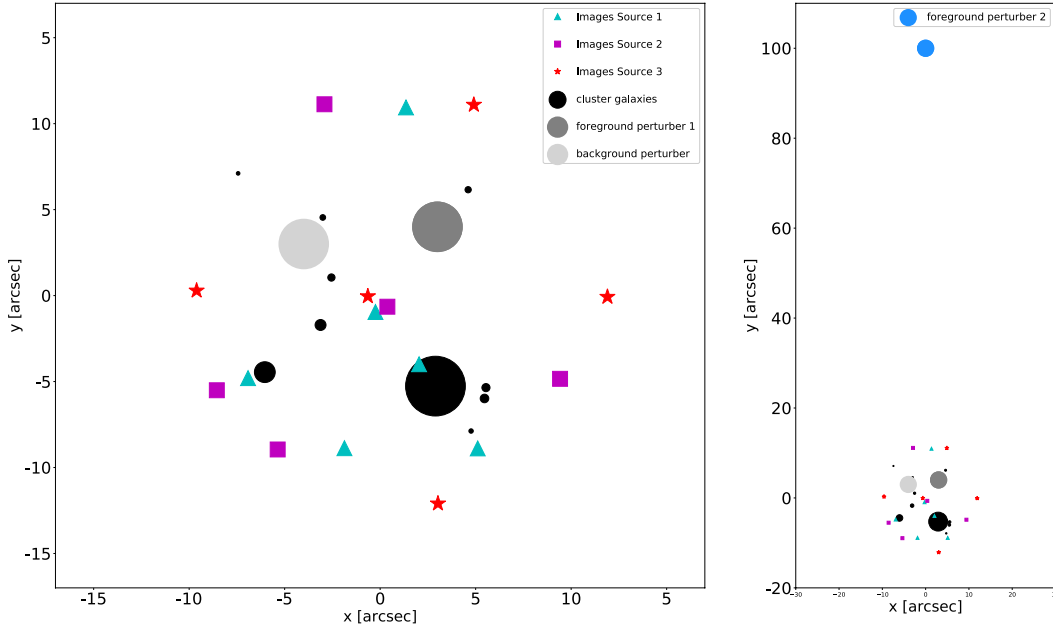


Fig. 3. Mock cluster lensing mass distribution 2. The black circles represent the lenses (cluster members), the grey circles the foreground galaxy (darker grey) and the background galaxy (lighter grey). The cyan triangles, magenta squares and red stars represent the images of the three sources, respectively at $z_{s1} = 1.5$, $z_{s2} = 2.0$, $z_{s3} = 2.5$. The circles radii are proportional to the galaxy's luminosity relative to the BCG (biggest circle). One far-in-projection foreground galaxy, shown in blue in the right panel, was added to this system (100'' away from the BCG).

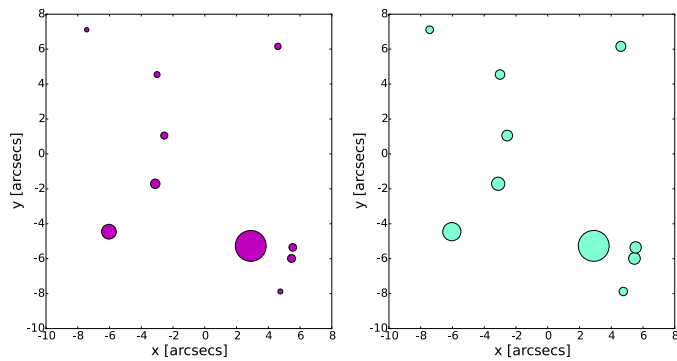


Fig. 4. Cluster galaxies' positional distribution of mock mass distribution 2, in the case of the total mass-to-light ratio equaling the tilt of the fundamental plane (*left*) and equaling to a constant (*right*) for the multi-plane case. The circles radii are proportional to the galaxy's Einstein radius relative to the BCG (biggest circle) mass.

within the errors, to the case with elliptical galaxies and the total rms is $\sim 0''.55$.

3.2.6. Mock cluster model 2: Assuming spherical cluster members

To further investigate the spherical assumption on galaxies, we produced a multi-plane model similar to that done previously, but assuming only the cluster members to be spherical. Therefore, the perturbors maintain the original ellipticity that was used in the simulation. As shown in column MP-sm in Table 1, this test still produces an offset in the centroid position, but by only $\sim 0''.4$; it underestimates the halo Einstein radius, but by $\sim 1''.2$, and overestimates the cluster member Einstein radius by $\sim 0''.4$, which is, instead, greater than the case where all the galaxies were spherical. This model also still overestimates the background perturber and the far-in-projection foreground perturber mass, but by a smaller amount, and does recover the close-in-projection foreground mass. Unlike the previous models, the halo slope in this model is underestimated, which makes the profile smoother.

The image position rms is close to the observational uncertainties of $\sim 0''.083$, and the magnification is generally smaller, within a factor of 0.7–0.9. As already mentioned in Sect. 3.2.5, we suspect these results will be mitigated in more realistic clusters cases.

3.2.7. Mock cluster model 2: Average surface mass density

To explore the differences in mass among the models, we computed the average surface mass density $\Sigma(< R)$ (Eq. (8)), as explained in Sect. 2.1. Figure 5 shows the relative error on the $\Sigma(< R)$, from top to bottom, for both the single plane and multi-plane cluster models as compared to the input. We note that in the bottom panel of Fig. 5, we plot the effective average surface mass density $\Sigma_{\text{Eff}}(< R)$, since it contains contributions from both the cluster and the perturbors, which are at different redshifts. We used the cluster redshift and the median source at redshift $z = 2$ for computing D_d , D_s and D_{ds} for the multi-plane $\Sigma_{\text{Eff}}(< R)$. The central panel of Fig. 5 shows, instead, the $\Sigma(< R)$ of the cluster for the multi-plane model, which we obtained from the convergence κ of the multi-plane best-fit models excluding the LOS perturbors' contribution. We see that the different models in the top and bottom panels of Fig. 5 tend to agree very well at $\sim 10''$, which is around the Einstein radius of the system. In the central panel of Fig. 5, however, the model with the spherical galaxies assumption differs by $\sim 10\%$ from the one with elliptical galaxies. This might be due to the fact that, for that multi-plane model, we computed κ for the cluster only, whereas lensing constrains tightly the total κ , including the contribution of the LOS perturbors. This is also confirmed by Fig. 6, where we see that the spherical assumption is always offset from the input model (which has elliptical galaxies and mass-to-light ratio equal to the tilt of fundamental plane). Moreover, it appears that the spherical model MP-s gives more mass to the perturbors, as shown already in Table 1. If we look at the case with spherical cluster members only (MP-sm), we see that the offset to the model is smaller than that of MP-s, where also the perturbors are assumed spherical. From Fig. 7 we see that this is valid also in the single-plane case. Probably leaving to the perturbors their original ellipticity gives

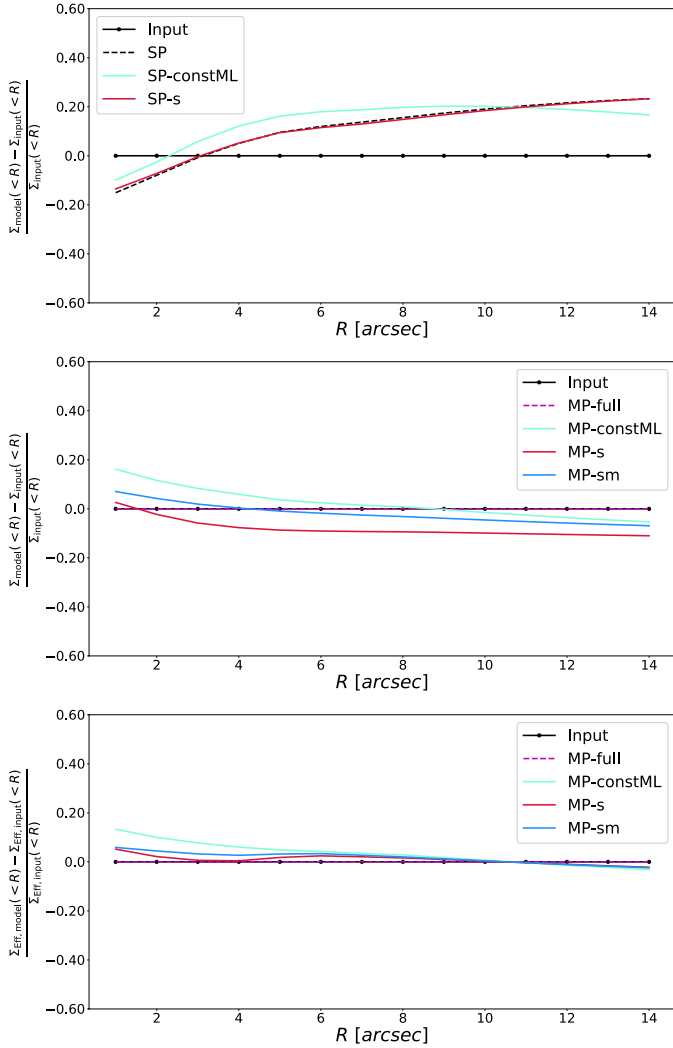


Fig. 5. Relative error on the average surface mass density $\Sigma(<R)$ as a function of radius from the BCG for the best-fit models of Mock mass distribution 2. *From top to bottom:* single plane cluster, cluster (without LOS perturbers) from multi-plane models, and total multi-plane configuration. In each plot, the lines represent the input (solid black), the multi-plane model (dashed magenta), the single-plane model (dashed black) the model where we assumed constant mass-to-light ratio (light blue), the model where we assumed spherical galaxies (red) and, only in the multi-plane cases, the model where we assumed only spherical cluster members (blue). We note that in the total multi-plane configuration the Σ_{Eff} is derived from the total deflection angle and computed for a source at $z_s = 2$, as explained in Sect. 2.1. We observe that at $\theta_{E,\text{tot}} \sim 10''$ all models converge to a certain value of $\Sigma(<R)$ and Σ_{Eff} in the *top* and *bottom* panels, showing that strong lensing provides accurate mass enclosed within the Einstein radius.

them less mass (see Table 1), thus the total cluster mass is recovered better. Finally, we observe that the multi-plane models are generally more peaked in the centre, as compared to single-plane ones, since their core radius is smaller than that of the single plane case (see Table 1).

3.2.8. Mock cluster model 2: Generic effects of LOS perturbers in the toy model

We find that the halo orientation and ellipticity are robust parameters. As we also notice in mock cluster mass distribution 1

(discussed in Appendix), they mainly stay, within the errors, close to the original values used to simulate. We also see that constraining the truncation radius of the galaxies $r_{t,g}$, especially that of the perturbers, is difficult, and we note that multi-plane models are generally better at constraining it than the single plane models (see Table 1). We think this might be caused by the limited amount of multiple image systems and by the fact that the truncation radius is of similar size as the total Einstein radius of the system ($\sim 15''$). However, in general this parameter does not appear correlated to the other parameters, and it is sampled as a flat probability distribution. We also note that the mass of the far away perturber is a quantity that is not generally correlated to the other parameters, except for a slight correlation to the cluster galaxies mass and the halo mass and slope. Moreover, its posterior probability distribution is flat, therefore this quantity is mostly unconstrained. This might be due to the fact that this distant galaxy is very far away from the multiple image region, which in our mock system extends to $\sim 15''$. In the multi-plane models we find a strong correlation between the core radius and the slope of the halo, which might be due to the fact that if we increase the core radius, the slope of the halo becomes less peaky, and that the background mass is strongly correlated with the mass of the foreground and with the halo mass. In all the single-plane cases, we find slight degeneracies between the halo ellipticity, mass and the cluster galaxies' mass. These degeneracies, as mentioned in Sect. A.3, would keep approximately the same total mass enclosed within the multiple images. Moreover, in mock cluster mass distribution 1 (discussed in the Appendix) we explore the effect of the single perturber, that is, only foreground and only background. We find that including the foreground is more important for a precise image reconstruction and for recovering the input parameter of the cluster, due to the lensing effect of foreground galaxies on the cluster itself.

4. MACS J0416.1–2403 mass distribution

4.1. Observation of MACS J0416.1–2403

In this section we present our models of the mass distribution of the HFF cluster MACS J0416.1–2403 (from here on, MACS0416), which is a massive luminous cluster at $z_c = 0.396$, with a very large spectroscopic data set along the LOS (Balestra et al. 2016). This cluster, shown in Fig. 8, was first discovered within the MAssive Cluster Survey (MACS) in the X-rays by Mann & Ebeling (2012). It shows an elongation in the NE-SW direction and was firstly identified as a merger. Since its discovery, MACS0416 has been extensively studied (e.g. Zitrin et al. 2013; Jauzac et al. 2014, 2015; Grillo et al. 2015; Diego et al. 2015; Caminha et al. 2017; Hoag et al. 2016; Kawamata et al. 2016; Natarajan et al. 2017; Bonamigo et al. 2017) as it represents an efficient lens for magnifying sources and producing multiple images (strong lensing regime), as shown in Fig. 9. In most recent studies (Grillo et al. 2015; Caminha et al. 2017) the image position was reconstructed with extremely high precision ($\Delta_{\text{rms}} \sim 0''.3$ in Grillo et al. (2015), with a set of 30 spectroscopically identified images, and $\Delta_{\text{rms}} \sim 0''.59$ in Caminha et al. (2017), with a set of 107 spectroscopically confirmed images), reaching the limit of what can be achieved by neglecting LOS contribution in the modelling. To study the effect of the LOS galaxies, we have created a mock mass distribution by building a model of MACS0416 that reproduces the observables as close as possible, and we use the model parameters to produce simulated observables. We then modelled the mock MACS0416

Table 1. Constraints on lens parameters for different models of mock cluster lensing mass distribution 2.

Parameters	Input	MP-full	SP	MP-constML	SP-constML	MP-s	SP-s	MP-sm
$\theta_{E, \text{fd1}} ["]$	2.00	$1.99^{+0.09}_{-0.08}$	–	$1.98^{+0.12}_{-0.12}$	–	$2.38^{+0.08}_{-0.08}$	–	$1.97^{+0.07}_{-0.07}$
$\theta_{E, \text{fd2}} ["]$	2.00	$2.02^{+0.19}_{-0.19}$	–	$3.10^{+1.40}_{-2.00}$	–	$3.10^{+1.40}_{-1.90}$	–	$3.00^{+1.40}_{-1.70}$
$x_{\text{halo}} ["]$	0.00	$0.00^{+0.08}_{-0.08}$	$0.45^{+0.13}_{-0.13}$	$0.20^{+0.16}_{-0.16}$	$0.63^{+0.12}_{-0.13}$	$0.63^{+0.08}_{-0.09}$	$0.53^{+0.11}_{-0.14}$	$0.43^{+0.07}_{-0.07}$
$y_{\text{halo}} ["]$	0.00	$0.00^{+0.09}_{-0.08}$	$1.30^{+0.20}_{-0.20}$	$0.01^{+0.09}_{-0.09}$	$1.30^{+0.19}_{-0.19}$	$-0.31^{+0.05}_{-0.04}$	$1.10^{+0.18}_{-0.17}$	$-0.10^{+0.05}_{-0.05}$
$\frac{b}{a}_{\text{halo}}$	0.80	$0.800^{+0.002}_{-0.002}$	$0.80^{+0.03}_{-0.04}$	$0.75^{+0.02}_{-0.02}$	$0.84^{+0.03}_{-0.04}$	$0.71^{+0.02}_{-0.03}$	$0.85^{+0.04}_{-0.05}$	$0.74^{+0.02}_{-0.02}$
$\theta_{\text{halo}} [\text{rad}]$	0.000	$0.003\pi^{+0.004}_{-0.004}$	$0.00^{+0.02}_{-0.02}$	$0.00^{+0.05}_{-0.05}$	$0.00^{+0.02}_{-0.02}$	$0.017\pi^{+0.004}_{-0.004}$	$0.00^{+0.02}_{-0.02}$	$0.007\pi^{+0.003}_{-0.003}$
$\theta_{E, \text{halo}} ["]$	10.00	$9.99^{+0.16}_{-0.16}$	$12.90^{+1.00}_{-0.79}$	$9.83^{+0.38}_{-0.40}$	$12.50^{+0.93}_{-0.72}$	$6.87^{+0.47}_{-0.39}$	$12.90^{+2.20}_{-1.00}$	$8.81^{+0.27}_{-0.28}$
$r_{c, \text{halo}} ["]$	2.00	$2.00^{+0.24}_{-0.23}$	$4.50^{+0.77}_{-0.74}$	$1.80^{+0.32}_{-0.28}$	$4.90^{+0.71}_{-0.72}$	$1.10^{+0.14}_{-0.12}$	$4.10^{+0.77}_{-1.30}$	$1.81^{+0.17}_{-0.17}$
γ_{halo}	0.40	$0.40^{+0.02}_{-0.02}$	$0.56^{+0.13}_{-0.11}$	$0.46^{+0.04}_{-0.04}$	$0.68^{+0.09}_{-0.13}$	$0.39^{+0.02}_{-0.02}$	$0.52^{+0.10}_{-0.15}$	$0.44^{+0.02}_{-0.02}$
$\theta_{E, \text{mg}} ["]$	0.17	$0.17^{+0.05}_{-0.05}$	$0.13^{+0.02}_{-0.02}$	$0.47^{+0.03}_{-0.03}$	$0.28^{+0.05}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$0.12^{+0.02}_{-0.03}$	$0.19^{+0.01}_{-0.01}$
$r_{t, \text{mg}} ["]$	1.93	$1.95^{+0.14}_{-0.13}$	$28.38^{+10.67}_{-4.41}$	$1.1^{+0.15}_{-0.13}$	$8.97^{+6.65}_{-4.21}$	$4.38^{+1.54}_{-1.10}$	$65.28^{+18.81}_{-52.55}$	$1.87^{+0.17}_{-0.21}$
$\theta_{E, \text{bd}} ["]$	2.00	$2.00^{+0.24}_{-0.25}$	–	$2.30^{+0.41}_{-0.42}$	–	$3.48^{+0.20}_{-0.23}$	–	$3.01^{+0.22}_{-0.21}$
$\text{rms}_{\text{best-fit}} ["]$		4×10^{-5}	0.549	0.126	0.511	0.069	0.547	0.083
$\chi^2_{\text{best-fit}}$		7×10^{-6}	1284	66	1085	20	1255	29

Notes. The first column refers to the values used to simulate, the other columns refer to different models, in order, the full multi-plane, the single cluster-plane, the full multi-plane and the single cluster-plane with constant mass-to-light ratio, the multi-plane and the single cluster-plane with spherical galaxies, and finally the full multi-plane with spherical cluster members only. The values are the medians of the posterior probability distributions of the lens parameters together with their 1σ uncertainties. The subscripts mg refers to the cluster member with the median luminosity among the members (F160W magnitude of 21.47). The orientation is measured counter clockwise from positive x -axis.

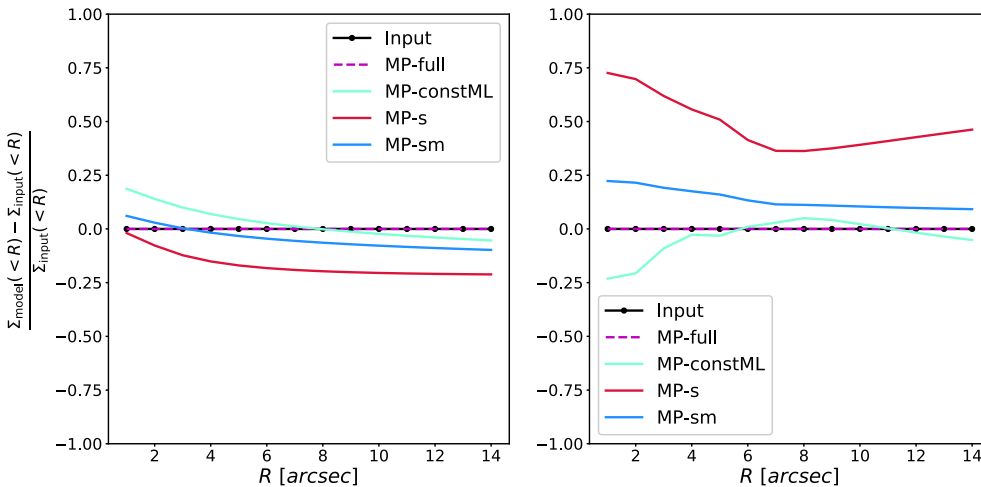


Fig. 6. Relative error on the average surface mass density $\Sigma(< R)$ as a function of radius from the BCG for the multi-plane best-fit models of Mock mass distribution 2. *Left:* cluster halo only. *Right:* cluster members only. In each plot, the four lines represent the full multi-plane model (dashed magenta), the model where we assumed constant mass-to-light ratio (light blue), the model where we assumed spherical galaxies (red), and the model where we assumed only spherical cluster members (blue) as compared to the input simulated cluster (solid black).

with different assumptions and compared them to the input mock model to assess the impact of LOS perturbers, as in previous sections.

4.2. MACS J0416.1–2403 best-fit model

We modelled the mass distribution of MACS0416 with a single-plane setup very similar to that presented by Caminha et al. (2017), but optimised independently with GLEE. This model consists of 193 cluster members, three dark matter halos and it is modelled on a set of 107 images (shown in Fig. 9) corresponding to 37 sources, all of them spectroscopically confirmed. We then included 11 chosen foreground and background galaxies into the model. These are listed in Table 2 and discussed in Sect. 4.2.1. We used truncated dual pseudo-isothermal elliptical

mass distributions (dPIEs) to represent the galaxies (members and perturbers), and pseudo-isothermal elliptical mass distribution (PIEMDs) for the halos, as these profiles were shown to reproduce better the observables (Grillo et al. 2015). We assumed all the galaxies to be spherical. The main reason for this choice is that measuring shapes of galaxies in a crowded cluster field is non-trivial and is thus beyond the scope of this work. The halos of galaxies located in the cores of clusters might have been significantly stripped and thus the shapes of these halos might deviate significantly from the shapes of the stellar component (e.g. Suyu et al. 2012; Harvey et al. 2016, among others). Furthermore, we suspect the effects of the ellipticities of the cluster members would average out with the substantial number of cluster members in MACS J0416.1–2403, as opposed to the case in mock mass distribution 2 (Sect. 3.2)

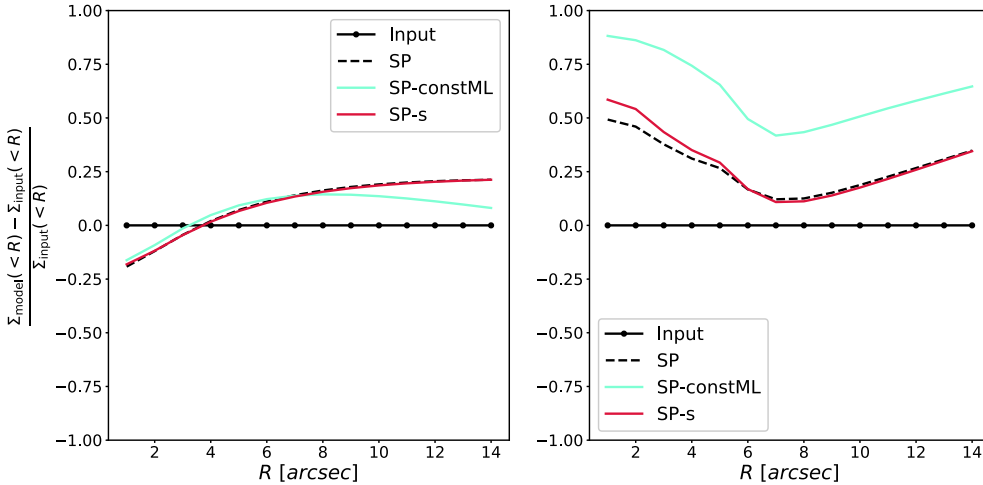


Fig. 7. Relative error on the average surface mass density $\Sigma(<R)$ as a function of radius from the BCG for the single plane best-fit models of mock mass distribution 2. *Left:* cluster halo only. *Right:* cluster members only. In each plot, the three lines represent the single plane model (dashed black), the model where we assumed constant mass-to-light ratio (light blue), and the model where we assumed spherical galaxies (red) as compared to the input simulated cluster (solid black).

where there are few cluster members. Two halos are located on the north-east (halo 1), south-west (halo 2) and centred, respectively, on the northern and southern BCGs. A third smaller halo is located in the eastern part of the northern halo.

To obtain the best-fit model that includes the perturbers, we model the halo parameters (centroid position (x_h, y_h) , axis ratio $\frac{b}{a}$, orientation θ , Einstein radius $\theta_{E,h}$, core radius $r_{c,h}$), the Einstein radii and truncation radii of the cluster members and perturbers. We scaled the Einstein radii and truncation radii of the cluster members using the tilt of fundamental plane for the total mass-to-light ratio (Eq. (18)), as described in Sect. 3.2.1, using Eq. (17). This mass-to-light ratio was shown to better reproduce the image position of this cluster by Grillo et al. (2015). As a reference galaxy we used a relatively isolated member (magnitude of 18.1 in the F160W band) which is marked in white in Fig. 8, together with the two BCGs, G1 and G2. We did not include in the scaling relation of the cluster members a luminous galaxy which is very close to the bright foreground galaxy (and also to a less massive foreground galaxy) in the south-west region of the cluster (see Fig. 10). The reason for this choice is that the light contamination from the foreground galaxy in that region does not allow an accurate estimation of the cluster member’s magnitude, which is also affected by the magnification due to the foreground galaxies. We quantified the same magnification effect due to the foreground galaxies on other close-by cluster members, but found that it was negligible. We also find that allowing that particular member to vary freely (both its Einstein radius, and truncation radius) decreases substantially the χ^2 of around $\sim 18\%$. We scaled the Einstein radius and truncation radius of the perturbers as explained in Sect. 4.2.1. We used an error on the images which corresponds to the observational uncertainty $\sim 0''.06$ (one pixel). However, we used a special treatment for images with high magnification forming arcs, since they have a more elliptical shape. We therefore introduced elliptical errors for those systems, with the minor axis of around one pixel, the major axis between $0''.2-0''.4$ (depending on the spatial extent of the arc) and orientated along the direction of the arc.

4.2.1. LOS secondary lenses

The secondary lenses are chosen among the brightest ($\text{mag} < 21.5$) and closest to the multiple image region objects

of the HST image, selected using near-IR (F160W) luminosities. We expect these brightest LOS galaxies to have the most significant effects on the modelling, compared to other fainter LOS galaxies. For the background perturber we have taken into account the magnification effect due to the cluster lensing. We kept the number of perturbers relatively low (11) for computational efficiency. Figure 8 shows the cluster MACS0416 with the perturbers, indicated by colour depending on whether they are foreground or background galaxies. The brightest foreground perturber in the southern region of the cluster was already included in previous models (Johnson et al. 2014; Richard et al. 2014; Grillo et al. 2015; Caminha et al. 2017), but at the cluster redshift $z=0.1124$ (first object in Table 2). Each of the 11 LOS perturber has two additional parameters (Einstein and truncation radii), and we scale the Einstein and truncation radii of these LOS galaxies with respect to the brighter ones. We scaled separately the foreground and background, using the definition of apparent magnitude and correcting the observed magnitude of the background for the magnification effect due to the presence of the foreground and the cluster, as follows:

$$m - m_{\text{ref}} = -2.5 \log_{10} \left(\frac{F}{F_{\text{ref}}} \right), \quad (24)$$

where the flux F

$$F = \frac{L}{4\pi D_1^2}, \quad (25)$$

with D_1 the luminosity distance to the object. Substituting in the definition of magnitude,

$$m - m_{\text{ref}} = -2.5 \log_{10} \left(\frac{\frac{L}{4\pi D_1^2}}{\frac{L_{\text{ref}}}{4\pi D_{1,\text{ref}}^2}} \right), \quad (26)$$

we obtain

$$m - m_{\text{ref}} = -2.5 \log_{10} \left(\frac{L}{L_{\text{ref}}} \left(\frac{D_{1,\text{ref}}}{D_1} \right)^2 \right). \quad (27)$$

If we now want to obtain the luminosity, we revert the equation to

$$\frac{L}{L_{\text{ref}}} = 10^{\frac{m_{\text{ref}} - m}{2.5}} \left(\frac{D_1}{D_{1,\text{ref}}} \right)^2, \quad (28)$$

and we scale it with the magnification μ ,

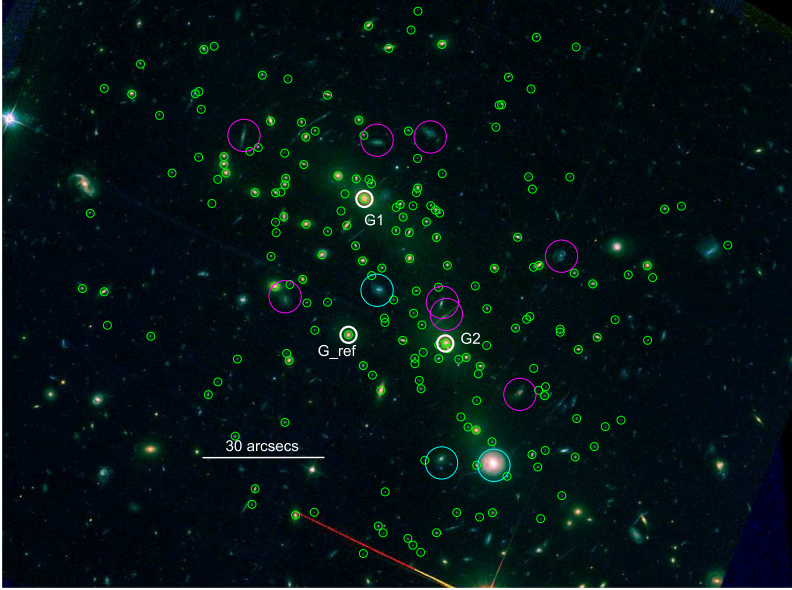


Fig. 8. Colour image of MACS J0416.1–2403 obtained through a combination of the HST/ACS and WFC3 filters. We indicate the selected 11 secondary lenses that we included in our model to account for the LOS contribution, using different colours for foreground (cyan) and background (magenta). In green we circle the cluster members. The two BCGs and the reference galaxy we use for the scaling relations are shown in white. North is up and east is left.

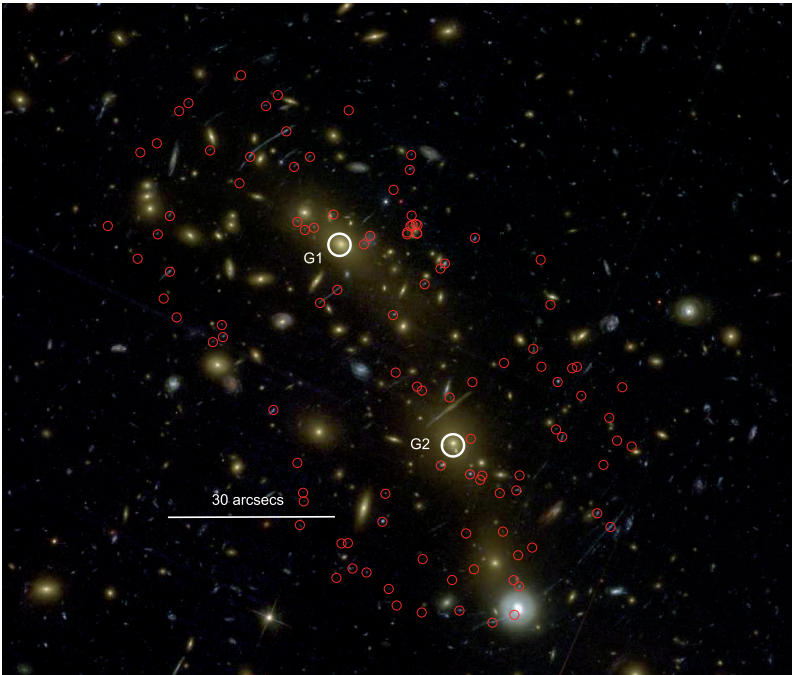


Fig. 9. Colour image of MACS J0416.1–2403 obtained through a combination of the HST/ACS and WFC3 filters. We show the selected 107 images we included in our model, corresponding to 37 sources that range from redshifts ~ 1 to ~ 6 . The two BCGs are shown in white. North is up and east is left.

$$L_r = \frac{\mu_{\text{ref}}}{\mu} \frac{L}{L_{\text{ref}}} = \frac{\mu_{\text{ref}}}{\mu} 10^{\frac{m_{\text{ref}} - m}{2.5}} \left(\frac{D_l}{D_{l,\text{ref}}} \right)^2, \quad (29)$$

where the magnification value for each perturber galaxy is obtained from the multi-plane best-fit model with only the foreground galaxies and the cluster. Therefore, the scaling is obtained by multiplying the reference quantity to the luminosity ratio L_r as shown in Eq. (17). We let the bright foreground galaxy at $z = 0.1124$ free to vary, and we scale the remaining two foreground galaxies with respect to the galaxy at $z_{f,\text{ref}} = 0.1126$, and the eight background galaxies with respect to that at $z_{b,\text{ref}} = 0.5004$. The reference perturbers were chosen since they have the greater apparent magnitude among, respectively, the foreground and background LOS galaxies, as shown in Table 2.

4.2.2. Results of best-fit model of MACS J0416.1–2403

Our best-fit model can reproduce the image positions with an rms $\sim 0''.53$ (see Fig. 11), that is approximately 8–9 pixels. Previous works with the same set of images (Caminha et al. 2017) was able to reproduce the observed image positions with a rms of $\sim 0''.59$; therefore our model with the addition of the LOS galaxies shows an improvement of $\sim 0''.06$. Our model has a $\chi^2 \sim 7.5 \times 10^3$. The best-fit mass distribution parameters are shown in Table 3. We remark that $1''$ at the redshift of the first foreground galaxy $z_{\text{fd}} \sim 0.1$ corresponds to ~ 1.8 kpc, while at the cluster redshift $z_c = 0.396$, $1''$ corresponds to ~ 5.34 kpc. Therefore, the truncation radius of the bright foreground galaxy is ~ 300 kpc, which is a typical value for massive galaxies. However, we find from the posterior probabilities that

Table 2. Selected secondary lenses added to the MACS J0416.1–2403 cluster, reported in redshift order, from the lowest to the highest.

RAJ2000	DEJ2000	ID	z	QF	R	$F814W$	$F160W$
64.0285	−24.0857	CLASHVLTJ041606.8–240508	0.1124	3	–	16.70	16.09
64.0323	−24.0854	CLASHVLTJ041607.7–240507	0.1126	3	20.66	20.76	20.25
64.0370	−24.0738	CLASHVLTJ041608.9–240425	0.1533	3	21.25	21.17	21.03
64.0323	−24.0747	CLASHVLTJ041607.8–240429	0.4678	3	21.05	21.49	19.92
64.0331	−24.0631	IDVLTJ041607.9–240347	0.4848	3	–	22.46	20.99
64.0322	−24.076	IDVLTJ041607.7–240433	0.5004	3	–	20.47	19.49
64.0473	−24.0633	CLASHVLTJ041611.3–240347	0.5277	3	21.22	21.44	20.80
64.0372	−24.0637	CLASHVLTJ041608.9–240349	0.5376	3	21.63	21.41	20.76
64.0441	−24.0745	CLASHVLTJ041610.6–240427	0.7093	3	21.32	21.03	19.96
64.0233	−24.0714	CLASHVLTJ041605.6–240417	0.7358	3	21.58	21.33	21.42
64.0265	−24.0809	CLASHVLTJ041606.4–240451	0.7362	3	21.20	20.70	19.79

Notes. The first two columns report the RA-DEC co-ordinates, the third column the ID, the fourth column the redshift, the fifth the quality flag of redshift (where a value of three represents “secure”), and the last three the R , $F814W$ and $F160W$ magnitudes. The line separates the galaxies that are on the foreground and on the background of the cluster.

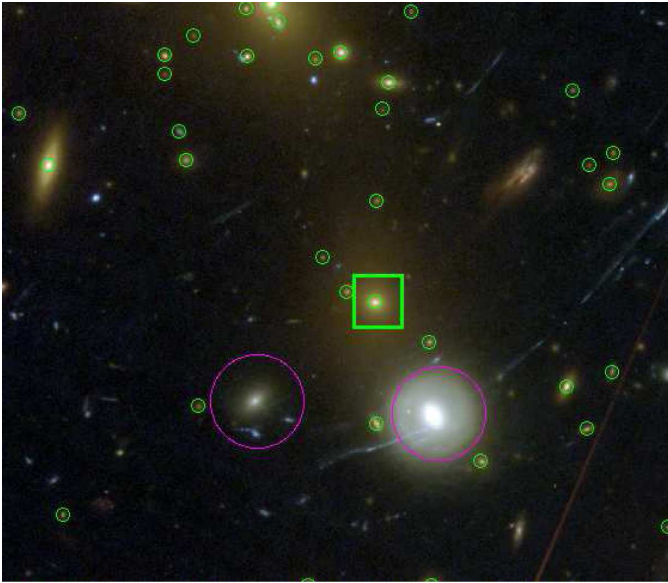


Fig. 10. Colour image of MACS J0416.1–2403 obtained through a combination of the HST/ACS and WFC3 filters. Green circles mark the cluster members. The massive member (green square) is not scaled using scaling relations, since it is very close to the two foreground galaxies (shown in magenta) and its observed magnitude might be affected by their light contamination.

the truncation radii of the LOS galaxies are mostly unconstrained due to the fact that they are isolated galaxies. The truncation radii of the cluster members are instead smaller, due to the tidal stripping of their dark matter halos. Figure 11 shows the observed image positions of MACS J0416.1–2403 and the predicted image positions of our best-fit model, which are in very good agreement, as can also be seen by the histogram of their positional offset. However, our model predicts 106 images of 107 observed, since two images in the north-east region of the cluster are predicted as a unique one (see Fig. 11)². The predicted magnification of the 106 images is shown in Figs. 12 and 13. One image has a very high predicted magnification

² These images belong to a source at redshift $z_s = 3.2387$, that has three arc-shape images with high magnification (>100). Two of them are very close to each other and they are predicted by our model to be one image.

($\sim 40000^{+2809}_{-2879}$), that is not shown in the plot for visualisation convenience³. In Fig. 14, we plot the average surface mass density for the best-fit model of MACS J0416.1–2403. We find that neglecting the LOS contribution does not affect the total Σ reconstruction significantly (see Fig. 14). We also find that most of the contribution on the outskirts (from $\sim 20''$ from the northern BCG), is due to the halo mass, while in the very centre ($\sim 5''$ from the northern BCG), the contribution of the galaxies is more prominent. We have used the unlensed positions (i.e. those they would have if the perturbers were not there) for the halos and members when computing the average surface mass density of the cluster, halos and members separately. We also computed the velocity dispersion of the cluster galaxies, that, for isothermal profiles, is related to the Einstein radius of the galaxy by

$$\theta_E = 4\pi \frac{\sigma_v^2}{c^2}, \quad (30)$$

for a source at infinite redshift. The velocity dispersion of a galaxy is also related to its circular velocity by

$$v_c = \sqrt{2}\sigma. \quad (31)$$

We then looked (Fig. 15) at the number of galaxies with a certain circular velocity, therefore mass, and we confirm the trend already pointed out by Grillo et al. (2015), that the amount of small mass halos seems to be in better agreement with simulations, whereas we find more high mass sub-halos than predicted by cosmological simulations (details on cosmological simulations can be found in Bonafede et al. 2011; Contini et al. 2012; Grillo et al. 2015). Regarding the radial distribution of substructures, we computed the position of the cluster members by removing the lensing effect of the foreground galaxies, and we compute the radial distance from the barycentre $R_b = (1'.21, -6'.85)$, which was obtained by the weighted average

$$R_b = \frac{\int \Sigma(R') R' dR'}{\int \Sigma(R') dR'}. \quad (32)$$

³ This image belongs to the system of arcs in the north-east part of the cluster mentioned previously ($z_s = 3.2387$). Again, this shows that this system, being close to the critical curves, is very highly magnified.

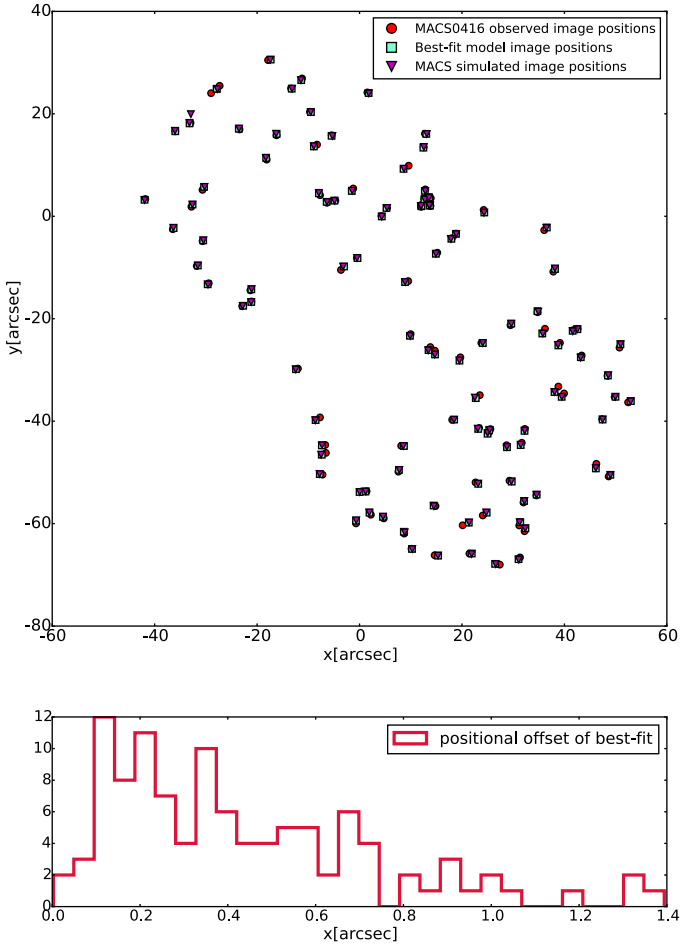


Fig. 11. *Upper panel:* predicted image positions (squares) of MACS J0416.1–2403 and the simulated image positions (triangles) of our MACS model (with introduced gaussian scatter), in comparison to the observed image positions (circles) of MACS J0416.1–2403. *Lower panel:* positional offsets between the observed and the model-predicted image positions. The rms for our best-fit model of MACS J0416.1–2403 is $\sim 0''.53$.

We find our model to slightly underpredict the number of substructures at small radii and at large radii ($\sim 300\text{--}400$ kpc), while to overpredict at radii $\sim 200\text{--}300$ kpc with respect to the model of Grillo et al. (2015). This implies a better agreement with cosmological simulations at smaller radii and at large radii, and a worse agreement for radii $\sim 200\text{--}300$ kpc.

4.3. Mock MACS input

Our best-fit model of MACS0416 and its environment can reproduce the image positions with a rms $\sim 0''.53$ (see Fig. 11), which is, as already mentioned, the current best-fit model obtained with this set of images. Our set of simulated constraints are the 106 predicted image positions from the best-fit model of MACS J0416.1–2403. Since one image of one system was not predicted by our model (see Sect. 4.2.2 for discussion), we decided to tweak the observed position of the third image manually in a way that its position was on the other side of the critical curve with respect to the second image and therefore such that it was predicted by our best-fit model. We then used the image position predicted by best-fit MACS J0416.1–2403 model as the 107th constraint for creating the mock system that we call mock MACS, making our set of constraints to a total of

Table 3. Constraints on lens parameters for our model of MACS J0416.1–2403 mass distribution (including three dark matter halos, 193 cluster members, three foreground galaxies and eight background galaxies).

Parameters	Best-fit parameter values
$\theta_{E, \text{fd1}}$ ["]	$1.60^{+0.21}_{-0.20}$
$r_{t, \text{fd1}}$ ["]	$147.00^{+2.40}_{-9.20}$
$\theta_{E, \text{fd2}}$ ["]	$0.05^{+0.07}_{-0.03}$
$r_{t, \text{fd2}}$ ["]	$98.40^{+1.20}_{-3.90}$
x_{h1} ["]	$-1.50^{+0.56}_{-0.46}$
y_{h1} ["]	$0.35^{+0.39}_{-0.53}$
$\frac{b}{a_{h1}}$	$0.33^{+0.02}_{-0.02}$
θ_{h1} [rad]	$2.50^{+0.02}_{-0.02}$
$\theta_{E, h1}$ ["]	$17.00^{+0.97}_{-0.90}$
$r_{c, h1}$ ["]	$8.33^{+0.50}_{-0.48}$
x_{h2} ["]	$21.00^{+0.41}_{-0.32}$
y_{h2} ["]	$-39.30^{+0.60}_{-0.73}$
$\frac{b}{a_{h2}}$	$0.38^{+0.01}_{-0.01}$
θ_{h2} [rad]	$2.19^{+0.01}_{-0.01}$
$\theta_{E, h2}$ ["]	$30.80^{+1.00}_{-1.30}$
$r_{c, h2}$ ["]	$12.60^{+0.52}_{-0.48}$
x_{h3} ["]	$-34.20^{+1.00}_{-1.60}$
y_{h3} ["]	$10.50^{+1.30}_{-1.00}$
$\frac{b}{a_{h3}}$	$0.85^{+0.11}_{-0.21}$
θ_{h3} [rad]	$2.60^{+0.31}_{-0.21}$
$\theta_{E, h3}$ ["]	$5.90^{+1.40}_{-1.00}$
$r_{c, h3}$ ["]	$9.50^{+2.60}_{-1.70}$
$\theta_{E, \text{mg}}$ ["]	$0.19^{+0.01}_{-0.01}$
$r_{t, \text{mg}}$ ["]	$0.76^{+0.12}_{-0.09}$
$\theta_{E, \text{pg}}$ ["]	$1.40^{+0.16}_{-0.17}$
$r_{t, \text{pg}}$ ["]	$4.82^{+0.13}_{-0.31}$
$\theta_{E, \text{bd}}$ ["]	$0.26^{+0.13}_{-0.13}$
$r_{t, \text{bd}}$ ["]	$23.00^{+11.00}_{-6.50}$

Notes. The first halo (“h1”) is the one located in the northern part of the cluster, the second halo (“h2”) in the southern one, the third (“h3”) is the smaller northern halo. The subscripts “fd” and “bd” refer, respectively, to the parameters of the foreground and background galaxies. The subscripts “mg” and “pg” refer to the parameters of the galaxy with median luminosity among the cluster members (F160W magnitude of 21.53) and of the peculiar member we allow to vary out of the scaling relations. The values are the medians of the posterior probability distributions of the lens parameters together with their 1σ uncertainties. The orientation is measured counter clockwise from positive x -axis. The centre of the co-ordinates is the position of the northern BCG (G1).

107 images. As mentioned before, the images’ positional uncertainties corresponds to the observational uncertainty $\sim 0''.06$, but we used a special treatment for images with high magnification forming arcs, introducing elliptical errors for those systems, as explained in Sect. 4.2. To obtain the simulated image positions, we then shifted the 107 predicted image positions (of the best-fit model) by a random number, drawn from a 2D-Gaussian distribution with $\sigma = 0''.06$, in both x and y direction to introduce a random scatter. In the case of highly magnified images forming arcs, we draw random numbers from a 2D elliptical Gaussian

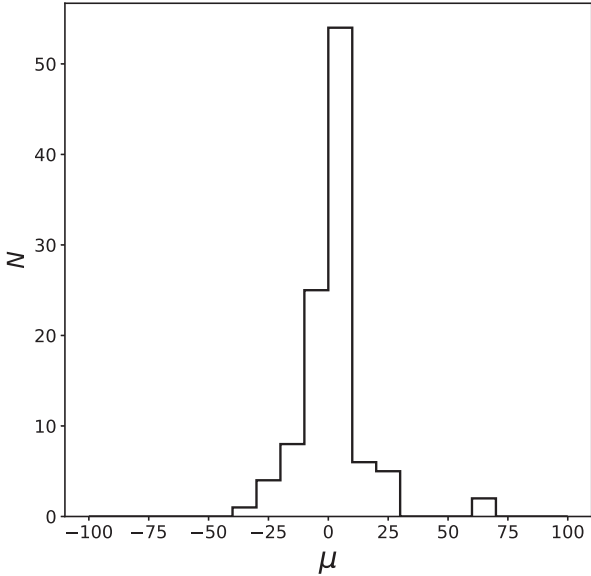


Fig. 12. Predicted magnification for the best-fit model of MACS J0416.1–2403. There is one image with predicted magnification of $\sim 4 \times 10^4$ (not included in our plot for visualisation convenience), which is one of the arc-shaped images in the north-east region ($-20''$, $20''$) of MACS J0416.1–2403, as shown in Fig. 11.

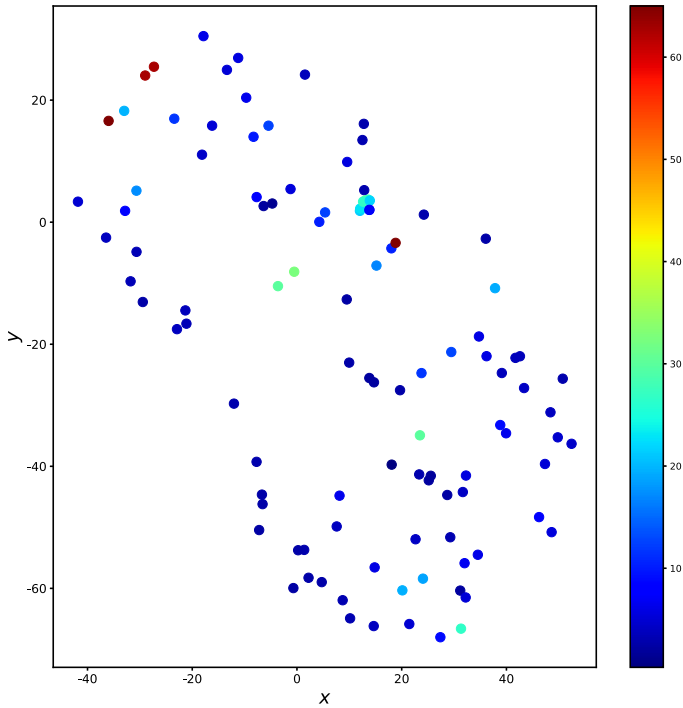


Fig. 13. Magnification for the image positions predicted by best-fit model of MACS J0416.1–2403. There is one image with predicted magnification of $\sim 4 \times 10^4$ and one with predicted magnification of ~ 600 whose magnification value was set to the border value of $\mu = 65$ for visualisation convenience. The two images are in the systems with high magnification and whose positional uncertainty was treated as elliptical due to their arc-like shape.

distribution, with σ_1, σ_2 equal to the minor and major axis of the error ellipses on those images. We then rotated the Gaussian to align with the direction of the arcs, since arcs are tangentially orientated. We used these 107 shifted image positions as our observables. The initial χ^2 of the input mock MACS cluster is 246.

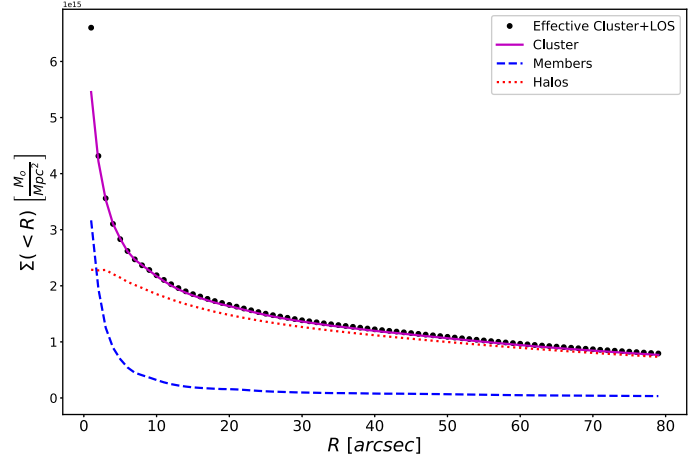


Fig. 14. Average surface mass density $\Sigma(< R)$ as a function of radius (from G1) for the best-fit model of MACS J0416.1–2403 for the total mass of the cluster (magenta line), for the cluster members (dashed blue) and for the dark matter halos (dotted red). The black points represents the effective average surface mass density $\Sigma_{\text{eff}}(< R)$ (for source at redshift $z_s = 3$) of the cluster and the LOS galaxies, as explained in Sect. 2.1. Of particular interest is the $\Sigma(< R)$ within $\sim 80''$, which approximately corresponds to the HST FoV at $z = 0.396$ (~ 420 kpc).

4.4. Mock MACS models

Once we had simulated a set of images (107 in total, from 37 sources), we modelled all the halo parameters (centroid position, ellipticity, orientation, Einstein radius, core radius) and the Einstein radii and truncation radii of the cluster members and of the perturbers (see Table 4), with different assumptions to assess the effect of these LOS perturbers. We modelled the mock MACS mass distribution using:

1. MP-full: the full multi-plane treatment, i.e. including all the LOS perturbers at the different redshifts;
2. SP1: assuming only the bright foreground perturber at the cluster redshift (similar to models in Johnson et al. 2014; Richard et al. 2014; Grillo et al. 2015; Caminha et al. 2017);
3. SP: assuming no LOS perturbers, and including only cluster members;
4. MP-fore: multi-plane including only the three foreground galaxies;
5. MP-back: multi-plane including only the eight background galaxies;
6. MP-constML: full multi-plane with different scaling relation, that is, using constant mass-to-light ratio as shown in Eq. (19);
7. SP-constML: single plane using constant mass-to-light ratio (Eq. (19)).

We describe the results of each of the models below.

4.4.1. MACS MP-full model results

In the case of the multi-plane modelling, where we modelled the parameters in the same set-up as the input, we recover, within the 1σ uncertainties, the input parameters, as shown in Table 4. The rms of this model is $0.08''$, which is very close to the observational uncertainty. We see that the offset between predicted and observed image positions is very small (Fig. 16) and the distribution of predicted and observed magnification ratio is centred around 1 with very small scatter in the tails, as shown in Fig. 17. One of the images with high magnification (and therefore elliptical errors) is predicted to have a different image parity than

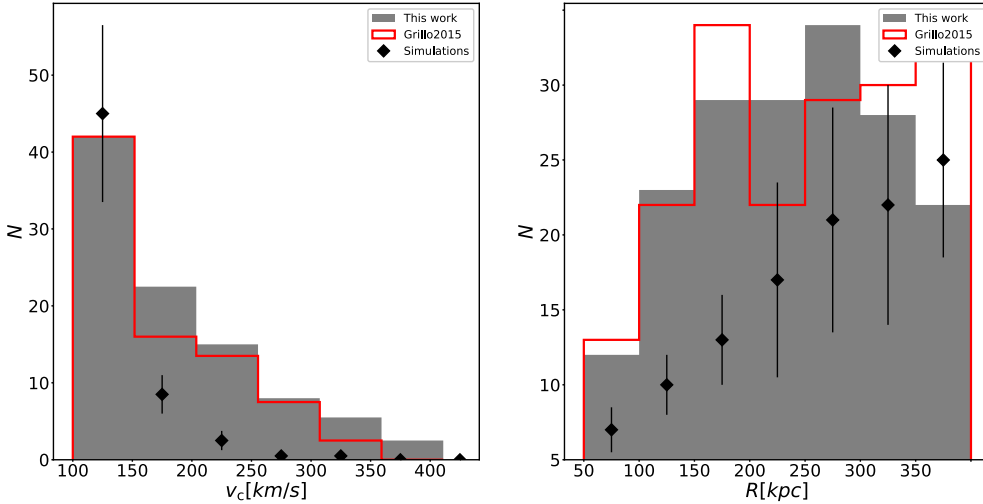


Fig. 15. Number of sub-halos as a function of the projected distance from the cluster lens centre (*right*) and as a function of circular velocity (*left*) within an aperture of ~ 400 kpc for our best-fit model of MACS J0416.1–2403 (grey filled) as compared to the model of Grillo et al. (2015) (red) and to the median values and the 1σ uncertainties obtained from cosmological simulations used therein (Bonafede et al. 2011; Contini et al. 2012; Grillo et al. 2015).

observed. This is due to the fact that it is close to a critical curve, and a slight change on the model can change its relative position with respect to the curve.

4.4.2. MACS SP1 model results

If we model the input mass distribution keeping only the brightest foreground at the cluster redshift (similarly to what was done in Grillo et al. 2015; Caminha et al. 2017), we find that the mass of this perturber is decreased. If we look at the scaled Einstein radii of the perturbers, as shown in Eq. (A.1), using $z_s = 3$ (roughly the mean redshift), we see that the difference is less significant, but still the scaled Einstein radius at the real redshift is higher than that at the cluster redshift. This mass is compensated by an increase in the mass of the southern halo and of the peculiar member (Fig. 10) that has a freely varying mass profile (without scaling relation imposed) due to its proximity to the bright foreground, which is anyway not substantial. However, most of the parameters are recovered. The total rms is $\sim 0.2''$, as shown in Table 4. This model is however pretty good at predicting the image positions and magnification (magnification ratio median ~ 0.97 as shown in Table 5), and does not show any parity flip in any of the predicted images.

4.4.3. MACS SP model results

In the single plane model, which includes only the cluster, the lack of the LOS galaxies' mass is compensated by an increase in the mass of the cluster members and halos. This model also shifts the centroid position of the southern halo (becoming closer to the bright massive foreground perturber) and of the smaller halo by almost $1''$. The total rms of this model is $\sim 0'.32$, which is very close to the model precision reached by the single-lens-plane models in recent years. Therefore, residual rms between the observed image positions and the image positions predicted by single-lens-plane models can actually be due to the lack of appropriate treatment of the lens environment (as previously suggested by e.g., Jullo et al. 2010). Moreover, this shows that including the LOS galaxies at the wrong redshift performs better than not including them at all (Table 4 and Sect. 4.4.2). However, the reconstruction of magnification, and therefore the intrinsic brightness of the source, is still possible with the single plane models with an error of $\sim 10\%$. Some more care is needed with images close to critical curves and highly magnified, some of

which are predicted to have an opposite image parity in this model as well.

4.4.4. MACS MP-fore model results

If we include only the foreground objects at their correct redshift, we find a slight increase of mass in both the members and the halos, to compensate for the lack of the background perturbers, but we can still recover the input within $1-2\sigma$, as shown in Table 4. The rms of this model is $0'.11$, which is also quite close to the observational uncertainty. As in the MP-full model, we find that the offset between predicted and observed image positions is very small and the distribution of predicted and observed magnification ratio is centred around one with very small scatter in the tails, as shown in Fig. 17 and Table 5. Moreover, the same image as in Sect. 4.4.1 is predicted with opposite image parity.

4.4.5. MACS MP-back model results

In the case of the multi-plane modelling with only the background galaxies, we find that the rms is $\sim 0'.3$, as in the single plane model. We see that the centroid of the two massive halos are shifted by $\sim 1''$, because the model did not take into account the lensing effect of the foreground galaxies. Moreover, to account for the lack of foreground, we find that the members and the southern halo are more massive. We note that this is probably due to the fact that the two most massive foreground galaxies are in the southern region of the cluster. Magnification is reconstructed with $\sim 10\%$ error, and this model predicts an image with flipped parity, which is very close to the southern BCG, and thus more sensitive to the increase in mass of members and of the southern halo.

4.4.6. MACS MP-constML model results

If we scale the cluster members and LOS galaxies assuming a different mass-to-light ratio, as for example that of Eq. (19), we find, as expected, that the offset in the predicted and observed image positions is greater than that of the multi-plane model, and it is $0'.2$. As shown in Figs. 16 and 17, the image positions and magnification are, overall, quite well reconstructed. However, in this case we have five images that show a flip in parity, one of which is very close to the southern BCG and is very highly magnified, and the others are part of the system in the north-west region of the cluster, which shows eight images of two clumps of

Table 4. Constraints on lens parameters for different models of the mock MACS mass distribution (three dark matter halos, 193 cluster members, three foreground galaxies and eight background galaxies).

Parameters	Input	MP-full	SP1	SP	MP-fore	MP-back	MP-constML	SP-constML
$\theta_{E,fd1}$ ["]	1.70	$1.73^{+0.04}_{-0.04}$	$1.19^{+0.05}_{-0.05}$	–	$1.67^{+0.04}_{-0.06}$	–	$1.82^{+0.07}_{-0.09}$	–
$r_{t,fd1}$ ["]	148.58	$147.90^{+0.79}_{-9.80}$	$112.90^{+0.97}_{-0.84}$	–	$111.70^{+0.18}_{-0.12}$	–	$147.00^{+2.00}_{-2.50}$	–
$\theta_{E,fd2}$ ["]	0.01	$0.03^{+0.03}_{-0.02}$	–	–	$0.02^{+0.03}_{-0.01}$	–	$0.06^{+0.06}_{-0.03}$	–
$r_{t,fd2}$ ["]	99.71	$99.34^{+0.46}_{-3.20}$	–	–	$94.64^{+0.11}_{-0.14}$	–	$59.00^{+11.00}_{-5.70}$	–
x_{h1} ["]	–1.51	$-1.55^{+0.09}_{-0.09}$	$-1.92^{+0.12}_{-0.17}$	$-1.2^{+0.22}_{-0.31}$	$-1.42^{+0.07}_{-0.08}$	$-1.70^{+0.35}_{-0.28}$	$-1.20^{+0.16}_{-0.15}$	$0.03^{+0.89}_{-0.91}$
y_{h1} ["]	0.51	$0.56^{+0.09}_{-0.08}$	$1.03^{+0.13}_{-0.10}$	$0.30^{+0.18}_{-0.18}$	$0.62^{+0.05}_{-0.05}$	$0.48^{+0.24}_{-0.30}$	$0.36^{+0.14}_{-0.14}$	$-0.36^{+0.77}_{-0.69}$
$\frac{b}{a}_{h1}$	0.34	$0.34^{+0.04}_{-0.01}$	$0.34^{+0.01}_{-0.01}$	$0.27^{+0.01}_{-0.01}$	$0.34^{+0.05}_{-0.01}$	$0.28^{+0.01}_{-0.02}$	$0.34^{+0.01}_{-0.01}$	$0.27^{+0.03}_{-0.06}$
θ_{h1} [rad]	2.500	$2.501^{+0.003}_{-0.004}$	$2.52^{+0.01}_{-0.01}$	$2.54^{+0.01}_{-0.01}$	$2.50^{+0.01}_{-0.05}$	$2.54^{+0.01}_{-0.01}$	$2.49^{+0.01}_{-0.01}$	$2.53^{+0.03}_{-0.03}$
$\theta_{E,h1}$ ["]	16.66	$16.68^{+0.14}_{-0.13}$	$15.20^{+0.20}_{-0.24}$	$11.4^{+0.35}_{-0.19}$	$16.47^{+0.15}_{-0.08}$	$13.0^{+0.39}_{-0.71}$	$15.90^{+0.30}_{-0.39}$	$11.00^{+1.20}_{-2.00}$
$r_{c,h1}$ ["]	7.92	$7.92^{+0.08}_{-0.09}$	$7.30^{+0.14}_{-0.16}$	$6.12^{+0.29}_{-0.21}$	$7.92^{+0.10}_{-0.09}$	$6.72^{+0.22}_{-0.42}$	$6.66^{+0.22}_{-0.18}$	$5.00^{+0.77}_{-1.20}$
x_{h2} ["]	20.73	$20.71^{+0.07}_{-0.07}$	$20.63^{+0.10}_{-0.09}$	$21.37^{+0.18}_{-0.15}$	$20.54^{+0.06}_{-0.06}$	$21.5^{+0.27}_{-0.28}$	$20.18^{+0.10}_{-0.10}$	$20.30^{+0.35}_{-0.49}$
y_{h2} ["]	–38.72	$-38.67^{+0.11}_{-0.09}$	$-38.16^{+0.14}_{-0.16}$	$-39.31^{+0.16}_{-0.23}$	$-38.34^{+0.07}_{-0.09}$	$-39.8^{+0.45}_{-0.50}$	$-38.06^{+0.21}_{-0.19}$	$-37.90^{+0.94}_{-0.62}$
$\frac{b}{a}_{h2}$	0.39	$0.39^{+0.01}_{-0.01}$	$0.38^{+0.01}_{-0.01}$	$0.34^{+0.01}_{-0.01}$	$0.39^{+0.04}_{-0.01}$	$0.34^{+0.01}_{-0.01}$	$0.38^{+0.01}_{-0.03}$	$0.33^{+0.11}_{-0.01}$
θ_{h2} [rad]	2.194	$2.196^{+0.002}_{-0.014}$	$2.196^{+0.004}_{-0.020}$	$2.17^{+0.01}_{-0.03}$	$2.20^{+0.01}_{-0.01}$	$2.17^{+0.01}_{-0.01}$	$2.20^{+0.01}_{-0.01}$	$2.18^{+0.01}_{-0.02}$
$\theta_{E,h2}$ ["]	31.26	$31.50^{+0.33}_{-0.24}$	$34.89^{+0.37}_{-0.21}$	$39.38^{+0.31}_{-0.27}$	$32.41^{+0.08}_{-0.14}$	$38.60^{+0.96}_{-0.51}$	$31.00^{+0.51}_{-0.48}$	$39.70^{+2.00}_{-1.10}$
$r_{c,h2}$ ["]	12.52	$12.70^{+0.14}_{-0.15}$	$13.93^{+0.18}_{-0.12}$	$16.4^{+0.23}_{-0.18}$	$12.87^{+0.08}_{-0.09}$	$16.20^{+0.44}_{-0.29}$	$12.10^{+0.21}_{-0.18}$	$15.60^{+0.84}_{-0.62}$
x_{h3} ["]	–33.73	$-33.74^{+0.15}_{-0.18}$	$-33.27^{+0.25}_{-0.32}$	$-33.88^{+0.33}_{-0.45}$	$-33.81^{+0.19}_{-0.09}$	$-35.60^{+1.20}_{-0.83}$	$-34.10^{+0.41}_{-0.42}$	$-32.80^{+2.60}_{-3.20}$
y_{h3} ["]	10.52	$10.60^{+0.12}_{-0.12}$	$10.20^{+0.18}_{-0.17}$	$12.00^{+0.23}_{-0.42}$	$10.67^{+0.08}_{-0.17}$	$11.80^{+0.46}_{-1.00}$	$10.70^{+0.24}_{-0.23}$	$11.60^{+1.30}_{-1.10}$
$\frac{b}{a}_{h3}$	0.90	$0.97^{+0.02}_{-0.06}$	$0.96^{+0.03}_{-0.04}$	$0.59^{+0.05}_{-0.06}$	$0.92^{+0.02}_{-0.02}$	$0.74^{+0.06}_{-0.09}$	$0.96^{+0.03}_{-0.05}$	$0.61^{+0.15}_{-0.17}$
θ_{h3} [rad]	2.56	$2.53^{+0.13}_{-0.21}$	$2.62^{+0.18}_{-0.18}$	$2.83^{+0.08}_{-0.06}$	$2.57^{+0.05}_{-0.07}$	$2.97^{+0.18}_{-0.12}$	$2.30^{+0.32}_{-0.55}$	$2.86^{+0.29}_{-0.15}$
$\theta_{E,h3}$ ["]	5.89	$5.60^{+0.29}_{-0.22}$	$4.95^{+0.16}_{-0.21}$	$7.34^{+0.22}_{-0.33}$	$5.39^{+0.07}_{-0.07}$	$6.14^{+0.51}_{-0.46}$	$6.14^{+0.41}_{-0.32}$	$7.30^{+2.20}_{-1.30}$
$r_{c,h3}$ ["]	8.95	$8.40^{+0.53}_{-0.39}$	$7.12^{+0.39}_{-0.30}$	$12.80^{+0.37}_{-0.49}$	$7.93^{+0.07}_{-0.10}$	$10.8^{+1.10}_{-0.72}$	$8.73^{+0.65}_{-0.58}$	$11.00^{+3.70}_{-1.90}$
$\theta_{E,mg}$ ["]	0.180	$0.172^{+0.001}_{-0.002}$	$0.179^{+0.002}_{-0.001}$	$0.200^{+0.002}_{-0.003}$	$0.182^{+0.002}_{-0.003}$	$0.190^{+0.011}_{-0.004}$	$0.311^{+0.001}_{-0.001}$	$0.39^{+0.02}_{-0.02}$
$r_{t,mg}$ ["]	0.72	$0.76^{+0.02}_{-0.03}$	$0.72^{+0.02}_{-0.02}$	$0.96^{+0.03}_{-0.04}$	$0.73^{+0.01}_{-0.01}$	$0.92^{+0.05}_{-0.04}$	$0.63^{+0.01}_{-0.01}$	$0.64^{+0.02}_{-0.06}$
$\theta_{E,pg}$ ["]	1.54	$1.58^{+0.04}_{-0.05}$	$1.85^{+0.21}_{-0.10}$	$1.55^{+0.11}_{-0.11}$	$1.60^{+0.04}_{-0.04}$	$1.56^{+0.14}_{-0.12}$	$1.53^{+0.07}_{-0.08}$	$1.50^{+0.29}_{-0.27}$
$r_{t,pg}$ ["]	4.98	$4.92^{+0.07}_{-0.15}$	$2.16^{+0.22}_{-0.39}$	$4.89^{+0.09}_{-0.22}$	$4.94^{+0.05}_{-0.09}$	$4.68^{+0.23}_{-0.63}$	$4.85^{+0.12}_{-0.28}$	$4.83^{+0.17}_{-0.95}$
$\theta_{E,bd}$ ["]	0.13	$0.13^{+0.02}_{-0.02}$	–	–	–	$0.13^{+0.08}_{-0.07}$	$0.25^{+0.03}_{-0.04}$	–
$r_{t,bd}$ ["]	33.50	$33.50^{+1.50}_{-4.70}$	–	–	–	$23.00^{+3.20}_{-3.20}$	$34.90^{+3.40}_{-5.80}$	–
$rms_{best-fit}$ ["]		0.08	0.19	0.32	0.11	0.34	0.20	0.45
$\chi^2_{best-fit}$		174	791	3021	246	3153	1124	3917

Notes. The first halo (“h1”) is the one located in the northern part of the cluster, the second halo (“h2”) in the southern part, the third (“h3”) is the smaller northern halo. The subscripts “fd” and “bd” refer, respectively, to the parameters of the foreground and background galaxies. The subscripts “mg” and “pg” refer to the parameters of the cluster galaxy with median luminosity among the cluster members (F160W magnitude of 21.53) and of the peculiar member we allow to vary out of the scaling relations. The values are the medians of the posterior probability distributions of the lens parameters together with their 1σ uncertainties. The orientation is measured counter clockwise from positive x -axis. The centre of the co-ordinates is the position of the northern BCG (G1).

the same source galaxy around two cluster members (system 14 in [Caminha et al. 2017](#), more detailed discussion in Sect. 4.4.9).

4.4.7. MACS SP-constML model results

If we model the input mass distribution with the single plane model and a constant mass-to-light ratio, as in Eq. (19), we find that the image positions are predicted with an overall rms of $\sim 0''.45$. This model is, expectedly, the worst at reproducing the image positions and magnification, and even has two predicted pairs of images on the same side of the critical curves. These images belong to the system of multiple images already mentioned in Sect. 4.4.6 (system 14 in [Caminha et al. 2017](#)) that correspond to the same source at $z = 3.2213$. We show the critical curves of that region in Fig. 20 (right panel). and we

discuss in Sect. 4.4.8. It also predicts two of the images with high magnification with a parity flip with respect to the input magnification.

4.4.8. Results

As expected, the multi-plane model is the one that reproduces the image positions best, with a rms of around $\sim 0''.08$, value similar to the observational uncertainty. However, accounting for the perturbers’ mass, even if at the wrong redshift (as in model SP1 in Fig. 16), allows the images positions to be reproduced better among the single plane models, with a rms of $\sim 0''.2$. The standard single-plane model (SP) has a rms of $\sim 0''.32$. We therefore suspect that part of the offset in single plane models of galaxy clusters might actually be due to the exclusion of the

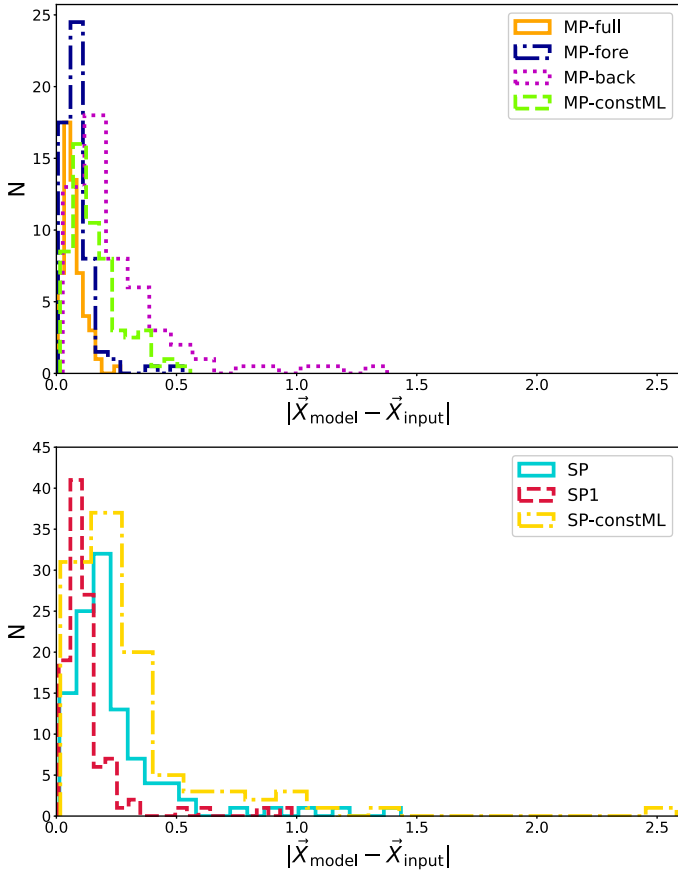


Fig. 16. Positional offset of the simulated observed vs predicted image positions of the different models of mock MACS. *Top panel:* multi-plane models, and *bottom panel:* single plane models. As expected, the full multi-plane model (black) is the one that reproduces the image position more closely.

Table 5. Magnification ratios between the reconstructed and input magnifications for different mock MACS models.

Model	$\mu_{\text{model}}/\mu_{\text{input}}$
MP-full	$1.00^{+0.05}_{-0.01}$
MP-fore	$1.00^{+0.03}_{-0.02}$
MP-back	$0.93^{+0.07}_{-0.11}$
MP-constML	$0.98^{+0.07}_{-0.05}$
SP	$0.92^{+0.06}_{-0.10}$
SP1	$0.97^{+0.04}_{-0.06}$
SP-constML	$0.90^{+0.07}_{-0.16}$

Notes. The listed values are the medians of the probability distribution functions of the magnification ratio, with 1σ uncertainties.

perturbers. We also explored the effect of taking into account only the foreground and background galaxies, and we find that the inclusion of the foreground is more important for a better fit of the observables, confirming what was found for mock mass distribution 1 (discussed in the Appendix) and in McCully et al. (2017), for a more simplistic mock cluster. Actually, for the mock MACS we find that including only the background gives a comparable rms to the single plane models, in other words, not including any LOS galaxy at all. Modelling with the wrong

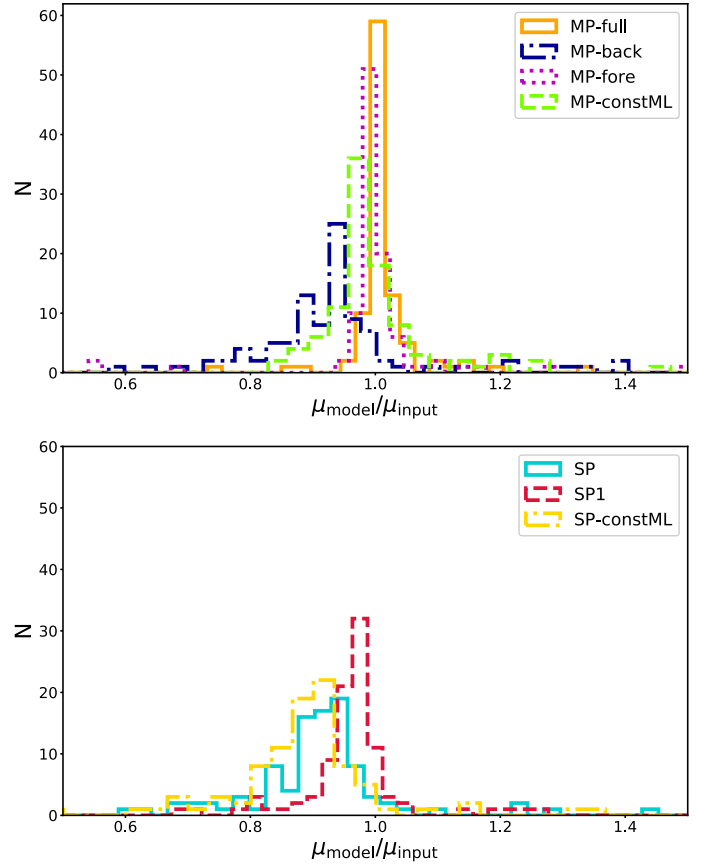


Fig. 17. Magnification ratio of the different mock MACS models and the input mock MACS. *Top panel:* multi-plane models, and *bottom panel:* single plane models. Interestingly, some images are predicted with high magnification ratio, and some are predicted with a different parity. These images are all the arc-shaped images that lie close to the critical curves.

scaling relation, as expected, increases the offsets of the predicted and observed image positions even more, for both the multi-plane and single plane model.

4.4.9. Source magnification reconstruction

As for the image positions, the magnification is reconstructed with different precision by the different models. Figure 17 shows the magnification ratio (predicted vs. input) distribution for the different models. We find that the MP-full model distribution is centred around one, showing that this model is the best at reproducing the input magnification. Also the MP-fore and SP1 models are able to reproduce the magnification quite well, with a median ratio of ~ 0.97 – 1 , as shown in Table 5. All the other models tend to predict a lower magnification than the input. In Table 5 we list the median values of the magnification ratio distributions and the respective 1σ uncertainties (16th and 84th percentiles). However, all the distributions have median values within 0.90 – 1.00 , therefore the median error on the magnification reconstruction is still within $\sim 10\%$. Interestingly, some image systems with high magnification and on (or close to) a critical curve, in almost all the models are predicted with a high magnification ratio and/or with a parity flip (we do not show these outliers in Fig. 17 for ease of visualisation). An example of a system with high magnification ratio and parity flip is shown in Fig. 20. In this figure we show the eight simulated observed

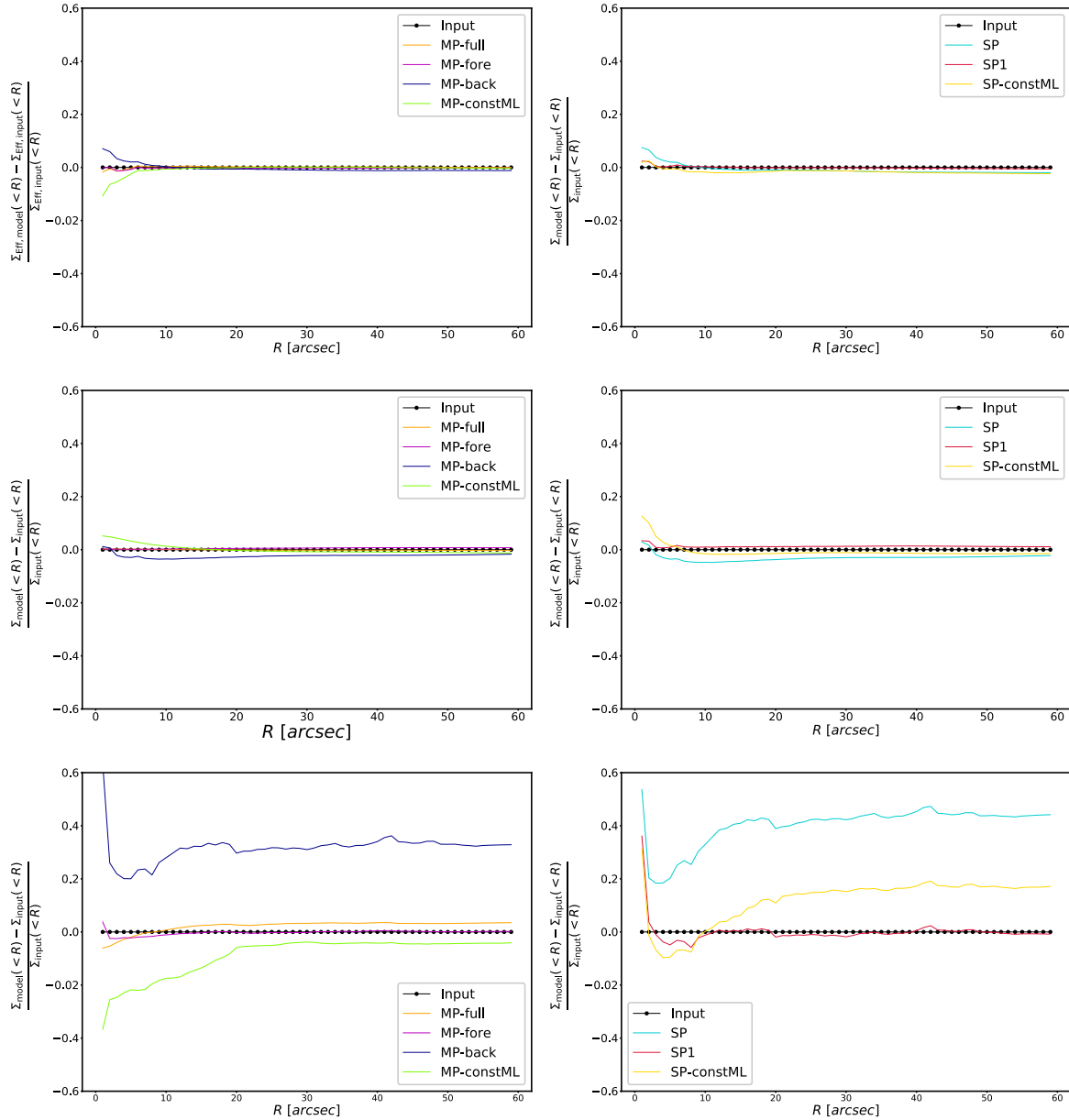


Fig. 18. Relative error with respect to the input mock MACS cluster on the average surface mass density $\Sigma(<R)$ as a function of radius (from G1) for the multi-plane MP (*left panels*) and single-plane SP (*right panels*) best-fit models of the mock MACS mass distribution compared to the input model (in dotted black). *Upper panels*: relative error on the $\Sigma_{\text{eff}}(<R)$ (for source at redshift $z_s = 3$) for the total cluster, i.e. halos, members and perturbers. *Central panels*: relative error on the average surface mass density of the halos for the different models. *Bottom panels*: relative error on the average surface mass density of the cluster members only. We note that in the total multi-plane configuration the $\Sigma_{\text{eff}}(<R)$ is relative to the total deflection angle, as explained in Sect. 2.1.

images of two source clumps of a galaxy at redshift $z_s = 3.2213$, which are located in the north-west region of the cluster (system 14 in [Caminha et al. 2017](#)). The multiple images of this source galaxy are located in the neighbourhood of two cluster members (cyan squares in Fig. 20), that act as strong lenses for that particular source. The images are highly magnified and form arcs around those members, and they are very sensitive to the shape of the critical curves due to the presence of those members. If we look at the left panel of Fig. 20 (corresponding to the input model), we see that a pair of images (30c, 31c) are predicted to be on different side of critical curves, with respect to the right panel (corresponding to the SP-constML model). Therefore, in the SP-constML model their predicted image positions coincide with those of, respectively, image 30b and 31b. If we compare the

input with the central panel (MP-constML model), we see that in this case the images 30c and 31c are closer to the critical curves, so they are predicted with a much higher magnification. Indeed, as already shown in Sect. 3.2.4, changing the mass-to-light ratio with which we scaled the Einstein and truncation radii of the cluster members, increases or decreases the mass of the members, and consequently changes the shapes of the critical curves for those members, affecting the relative position with respect to critical curves of images nearby. In the model of [Caminha et al. \(2017\)](#), these two cluster members were considered as free parameters instead of being scaled with the other members, since they are the main contributors to the creation of the multiple images of this system (system 14 in [Caminha et al. 2017](#)). In general, we suspect images close or on the critical curves need

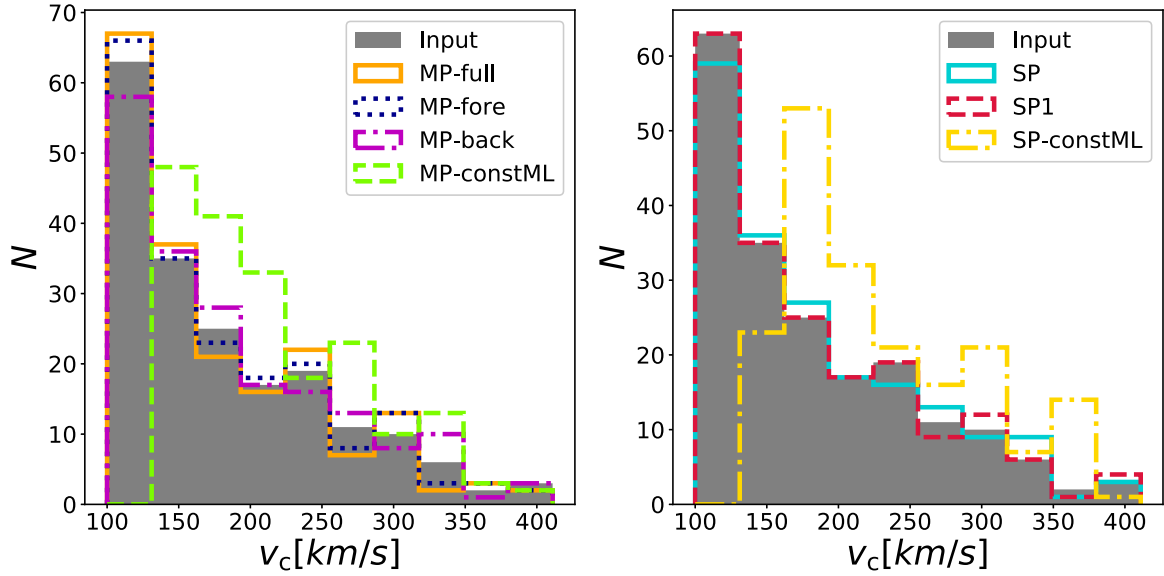


Fig. 19. Number of sub-halos as a function circular velocity for our multi-plane (*left*) and single plane (*right*) models within an aperture of ~ 400 kpc. Surprisingly, the model that reproduces better the input is the SP1 model, instead of the MP-full. However, these results are related to the best-fit models and we suspect this might be a coincidence specific to the configuration of MACS J0416.1–2403.

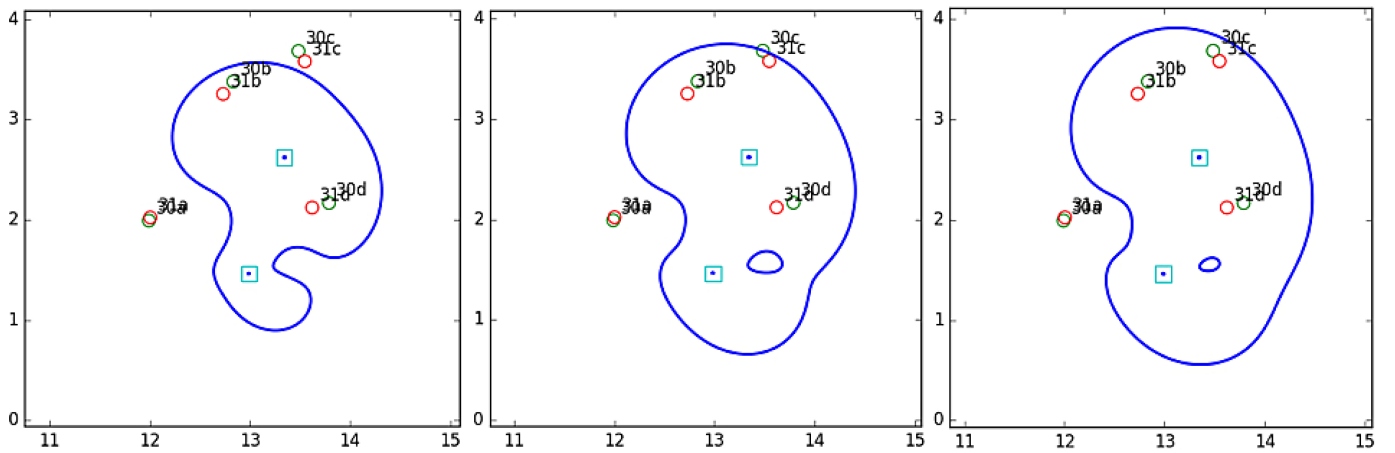


Fig. 20. Critical curves for the region of the multiple images of the source galaxy at $z_s = 3.2213$ of our mock MACS models: MP-full (*left*), the MP-constML (*centre*) and the SP-constML (*right*). The different clumps of the source are labelled with different numbers (30 and 31).

to be treated carefully when trying to reconstruct the intrinsic brightness of the background sources.

4.4.10. Mass reconstruction

Figure 18 shows the relative error on the average surface mass density for our models as compared to the input mock MACS cluster. We do not observe significant differences among the different models for the total $\Sigma(< R)$, which reaches a maximum relative error of 0.6 in the cluster core, probably since the number of the perturbers is much smaller than the total cluster mass, therefore they do not contribute significantly to the total mass load. Moreover, all the models agree quite well in the region of the Einstein radius of the cluster, since the mass within the Einstein radius is the quantity that lensing constraints tightly. If we look at the single matter components, that is, only the halos (central panel) and only the members (bottom panel), we find slight differences among the models, especially in the inner part of the cluster, while in the outskirts they all agree quite well. The models that are slightly different from the input are, among the

MP models, the MP-back and MP-constML, in both the halos and cluster members $\Sigma(< R)$. The model MP-back, as we already saw from the positional offset (Table 4 and Fig. 16), is the worst at reproducing the observables among the MP models. This tells us that including only the background perturbers can even perform similarly to the SP model, in which they are not included at all. This implies that the average surface mass density can be similar to that of the SP model (as shown in Fig. 18). Assuming a different mass-to-light relation makes the halos $\Sigma(< R)$ steeper within the inner part of the cluster, and the cluster members' one less peaky. We see the same trend in the SP case.

Despite these slight differences, the overabundance of galaxies at the centre of MACS0416 in observations compared to simulations noted in Grillo et al. (2015) is robust against the presence of LOS perturbers. Figure 19 shows the number of cluster members as a function of their circular velocity. If we compare it to the input, we see that most of the multi-plane and single plane models do not change substantially the overall velocity dispersion distribution of the members, showing that this is robust even though the model is not complete. However, we see that both in

the single and multi-plane case, if we assume a wrong scaling relation, the distributions tend to prefer more massive galaxies (around 100 km s^{-1} in the multi-plane case and even 200 km s^{-1} in the single plane case). Thus, the choice of scaling relation could potentially alleviate the tension in the disparate numbers of massive galaxies and sub-halos in the inner parts. Encouragingly, our simulations suggest that the χ^2 of the model fit could possibly be used to probe the underlying scaling relation. It is worth exploring further scaling relations of the cluster galaxies in future models of MACS J0416.1–2403.

5. Summary

In this work we explored the effects of the LOS galaxies in strong gravitational lensing modelling of galaxy clusters. We simulated different galaxy clusters and their environment, building models of increasing complexity and realism, that we used to simulate strong lensing observables. We then determined the lensing halos' and galaxies' parameters with different assumptions and compared to the input simulated cluster to assess the effects of the LOS perturbers.

The simulated system mock cluster mass distribution 2 is composed of a cluster at $z_c = 0.4$ with a halo and ten elliptical galaxies having different, realistic luminosities, axis ratios and orientations. We assumed a total mass-to-light ratio corresponding to the tilt of the fundamental plane. We added two foreground perturbers at $z_{fd} = 0.2$, one close and one far away in projection, and one close-in-projection background perturber at $z_{bd} = 0.6$. All the perturbers have equal, large mass, random ellipticity (between 0.6 and 1) and orientation. We used this configuration to simulate mock lensing data, and we obtained a set of 16 multiple image positions of three background sources. In this mock system we explored the effect of different mass-to-light relations and the spherical-elliptical galaxies assumption. We find that:

1. Isolated perturbers that are far-in-projection (ten times or more) from the multiple image region do not affect substantially the other parameters' values. Indeed their posterior probability is sampled as flat and they do not look correlated to other parameters.
2. In multi-plane modelling, assuming spherical galaxies (both members and perturbers) recovers different profile parameters than the input, and has generically lower magnification. Despite this, the multi-plane reconstruction has a rms very close to the observational uncertainty. Therefore, one needs to be careful in interpreting the reconstructed mass distribution and magnification, as the goodness of fit does not allow to distinguish between spherical and elliptical galaxies. However, the difference in the parameters is less prominent when only the cluster members are treated as spherical. This might be due to the fact that the perturbers of this model are very massive and very close to the cluster centre, therefore they have a large effect on the total mass reconstruction. Moreover, we suspect this effect might depend on the simplicity of our model and might be mitigated in more realistic clusters with higher number of cluster members.
3. Multi-plane models have a peakier average surface mass density than single plane models, which instead show a shallower profile. However, all the average surface mass density values match at $R \sim 10''$, which is the total Einstein radius of the cluster. This shows that the projected total mass enclosed within the Einstein radius - the quantity

that strong gravitational lensing constraints tightly - is well reconstructed by all these models.

The other simulated cluster we studied is a realistic model of the HFF cluster MACS J0416.1–2403, labelled mock MACS, which we built using 107 spectroscopically confirmed multiple images from 37 sources (with $1 < z_s < 6$) as constraints, and then modelling the three halos (two located in the NE, and one in the SW direction), the 193 spectroscopically confirmed cluster members, and the LOS galaxies' profile parameters. We assumed the cluster members and LOS galaxies to be spherical. We find that our mock of MACS J0416.1–2403 which includes 11 LOS perturber galaxies is able to reproduce the real cluster's observables with a rms of $\sim 0'.53$. We then modelled this simulated mock MACS cluster to assess the impact of LOS galaxies. Our results can be summarised as follows:

1. Neglecting the cluster's LOS environment leads to a rms of $\sim 0'.3$ in the offset distance between observed and modelled image positions.
2. The inclusion of LOS galaxies in the cluster modelling at the wrong redshift (i.e. at cluster redshift) reproduces the observed image positions with a rms of $\sim 0'.20$.
3. We find that foreground perturbers have a more significant effect than the background. However, since in this case the discrepancy is more prominent (including only the background is comparable to not including it at all), we ascribe it also to the particular configuration of the cluster.
4. Neglecting the lensing environment or assuming it at the wrong redshift does not affect the reconstruction of the magnification of background sources substantially (errors within $\sim 10\%$). However, apart from the full multi-plane reconstruction, all the other models tend to underpredict the magnification. We also find that systems with high magnification, or in general close to critical curves, are more sensitive to the different assumptions on the modelling. Indeed, since strong lensing is a highly non-linear effect in these regions, small changes in the parameter values can cause changes on the critical curves positions. These can lead to a much higher predicted magnification and may also cause a flip in image parity.
5. We do not observe a substantial difference in the average surface mass density of the cluster among the various mass models we have considered, probably because these perturbers are very small compared to the total cluster mass. Thus, the overabundance of galaxies at the centre of MACS0416 in observations compared to simulations noted in [Grillo et al. \(2015\)](#) seems robust against the presence of LOS perturbers. However, the mass function of the cluster members depends on the assumed scaling relations. The dependence can be partly mitigated by using the goodness of fit (χ^2) of the multiple image positions, since the 'true' scaling relations generally lead to lower χ^2 .
6. We find a correlation between the masses of the foreground LOS perturber galaxies and the halos' centroid co-ordinates. We also see that the Einstein radius of the different mass components are anti-correlated among each other, since a decrease of mass in one can be compensated by an increase in another component. Moreover, the Einstein radius and truncation radius of the galaxies are correlated with each other.
7. Assuming different scaling relations can lead to very different results for the mass of the members and of the halos, and therefore can change the substructure distribution quite significantly.

Finally, from our best-fit model of MACS J0416.1-2403 we find that, for this particular cluster, the total *rms* does not decrease substantially after the inclusion of LOS galaxies with respect to single plane models. This suggests that the dynamic and turbulent nature of this cluster is still probably not well depicted by, for example, the simplistic model of the halos. For future model improvements of this cluster those effects would need to be accounted for. We also find that using the correct scaling relations is crucial, especially for reconstructing the sub-halo mass function. Therefore, we conclude that the main contribution to the residual *rms* for this particular cluster is not due to LOS galaxy, and future models will have to move towards including more complex halo shapes, accounting for the ellipticity of galaxies and improving the scaling relations for the cluster members.

Acknowledgements. GC and SHS thank the Max Planck Society for support through the Max Planck Research Group for SHS. GC acknowledges Prof. Alexandre Refregier and Dr. Adam Amara for the helpful discussions. We thank the anonymous referee for the helpful comments that significantly contributed to improving the quality of the publication. CG acknowledges support by VILLUM FONDEN Young Investigator Programme grant 10123. GBC, AM and PR acknowledge financial support from PRIN-INAF 2014 1.05.01.94.02.

References

- Balestra, I., Mercurio, A., Sartoris, B., et al. 2016, *ApJS*, 224, 33
 Barkana, R. 1998, *ApJ*, 502, 531
 Bender, R., Burstein, D., & Faber, S. M. 1992, *ApJ*, 399, 462
 Blandford, R., & Narayan, R. 1986, *ApJ*, 310, 568
 Bonafede, A., Dolag, K., Stasyszyn, F., Murante, G., & Borgani, S. 2011, *MNRAS*, 418, 2234
 Bonamigo, M., Grillo, C., Ettori, S., et al. 2017, *ApJ*, 842, 132
 Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014, *ApJ*, 795, 126
 Caminha, G. B., Grillo, C., Rosati, P., et al. 2016, *A&A*, 587, A80
 Caminha, G. B., Grillo, C., Rosati, P., et al. 2017, *A&A*, 600, A90
 Coe, D., Zitrin, A., Carrasco, M., et al. 2013, *ApJ*, 762, 32
 Contini, E., De Lucia, G., & Borgani, S. 2012, *MNRAS*, 420, 2978
 D'Aloisio, A., Natarajan, P., & Shapiro, P. R. 2014, *MNRAS*, 445, 3581
 Diego, J. M., Broadhurst, T., Molnar, S. M., Lam, D., & Lim, J. 2015, *MNRAS*, 447, 3130
 Dressler, A. 1984, *ARA&A*, 22, 185
 Dunkley, J., Bucher, M., Ferreira, P. G., Moodley, K., & Skordis, C. 2005, *MNRAS*, 356, 925
 Elíasdóttir, A., Limousin, M., Richard, J., et al. 2007, ArXiv e-prints [arXiv:0710.5636]
 Faber, S. M., Dressler, A., Davies, R. L., Burstein, D., & Lynden-Bell, D. 1987, in *Nearly Normal Galaxies. From the Planck Time to the Present*, ed. S. M. Faber (New-York: Springer-Verlag), 175
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
 Gavazzi, R., Treu, T., Koopmans, L. V. E., et al. 2008, *ApJ*, 677, 1046
 Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, *ApJ*, 800, 38
 Harvey, D., Kneib, J. P., & Jauzac, M. 2016, *MNRAS*, 458, 660
 Hoag, A., Huang, K.-H., Treu, T., et al. 2016, *ApJ*, 831, 182
 Jauzac, M., Clément, B., Limousin, M., et al. 2014, *MNRAS*, 443, 1549
 Jauzac, M., Jullo, E., Eckert, D., et al. 2015, *MNRAS*, 446, 4132
 Johnson, T. L., Sharon, K., Bayliss, M. B., et al. 2014, *ApJ*, 797, 48
 Jullo, E., Natarajan, P., Kneib, J.-P., et al. 2010, *Science*, 329, 924
 Kassiola, A., & Kovner, I. 1993, *ApJ*, 417, 450
 Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016, *ApJ*, 819, 114
 Kravtsov, A. V., & Borgani, S. 2012, *ARA&A*, 50, 353
 Mann, A. W., & Ebeling, H. 2012, *MNRAS*, 420, 2120
 McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2014, *MNRAS*, 443, 3631
 McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2017, *ApJ*, 836, 141
 Natarajan, P., Chadayammuri, U., Jauzac, M., et al. 2017, *MNRAS*, 468, 1962
 Planck Collaboration XX 2014, *A&A*, 571, A20
 Richard, J., Jauzac, M., Limousin, M., et al. 2014, *MNRAS*, 444, 268
 Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, *ApJ*, 708, 645
 Schneider, P., Ehlers, J., & Falco, E. E. 1992, *Gravitational Lenses*, 112
 Springel, V., Frenk, C. S., & White, S. D. M. 2006, *Nature*, 440, 1137
 Suyu, S. H., & Halkola, A. 2010, *A&A*, 524, A94
 Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, *ApJ*, 750, 10
 Zitrin, A., Meneghetti, M., Umetsu, K., et al. 2013, *ApJ*, 762, L30

Appendix A: Mock cluster lensing mass distribution 1

A.1. Input

The simulated mock cluster lensing mass distribution 1 is that of a cluster at redshift $z_c = 0.4$ and two galaxies, a foreground galaxy at $z_{fd} = 0.2$ and a background galaxy at $z_{bd} = 0.6$, that are close in projection to the cluster and that we treat as perturbers. The cluster is composed of five galaxies and a dark matter halo. The input parameters are summarised in Table A.1. We assumed all the galaxies in the cluster to have the same luminosity, therefore same mass, for simplicity, and same truncation radius, that we fix to $15''$. We put three different point sources (S1, S2, S3) at different redshifts ($z_{s1} = 1.5$, $z_{s2} = 2$, $z_{s3} = 2.5$) and simulate their image positions in the case of the multi-plane configuration. We find that each source is mapped into five images, for a total of 15 images, and used these image positions as our observables. The set-up of this model is illustrated in Fig. A.1.

A.2. Full multi-lens-plane modelling

We optimised the free parameters of the multi-plane model (Einstein radius of the foreground, background and cluster galaxies, and all the parameters of the halo profile - Einstein radius, position, ellipticity, orientation, core radius and slope) by maximising the likelihood in the image plane. We used simulated annealing to find the global minimum and recover the best-fit parameter values. In this case, since we simulated the image positions with the multi-plane model, we recovered, within the uncertainties estimated by MCMC sampling, the initial parameters we have used to simulate, as shown in Table A.1. The modelled image positions and the magnifications are perfectly fitted, with a null total-rms offset. We find a strong correlation between the mass of the halo and its axis ratio and between the core radius of the halo and the slope of the halo profile. We also find a strong anti-correlation between the centroid position of the halo and the mass of the foreground galaxy, which is explained by the fact that the centroid position is lensed by the foreground galaxy, and between the mass of the cluster galaxies and the axis ratio of the halo. We also find that the mass of the background lens is strongly anti-correlated with the halo mass, core radius and slope, that could also be explained as a lensing effect by the cluster.

A.3. Single-plane modelling

We then removed the foreground and background lens and fit the same image positions with the cluster alone. We varied the cluster parameters and find, as shown in Table A.1, that the halo is still orientated along on the x direction, and the ellipticity is recovered within the errors. However, its centroid position is offset of $\sim 1''$ in both x and y direction, its mass is bigger and its core radius is larger by $4''$, making the profile less peaky in the centre. Moreover, we recover a smaller mass for the cluster galaxies. In this case we find an additional strong anti-correlation between the halo and the cluster galaxies Einstein radius, which was not so prominent in the multi-plane system. This could be explained as the removal of the two LOS perturbers influences the distribution of the mass between the two remaining mass components, the halo and the cluster galaxies. Moreover, we find degeneracies between the axis ratio and the Einstein radius of the cluster and between the latter and the cluster galaxies' Einstein radius, since these degeneracies would keep approximately the same

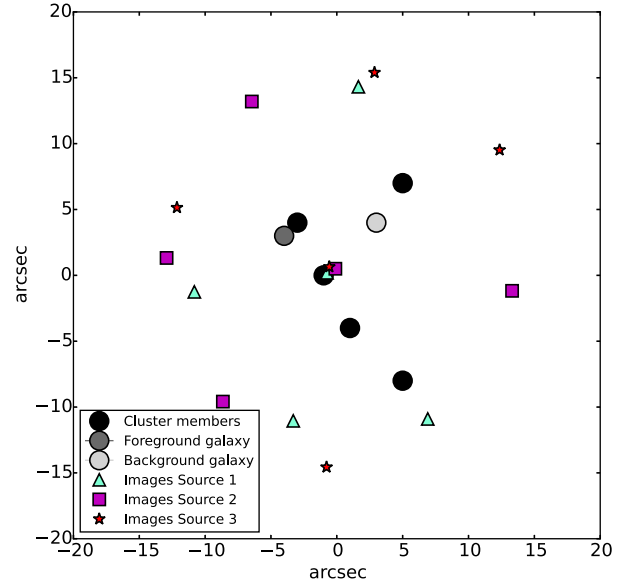


Fig. A.1. Mock cluster lensing mass distribution 1. The black circles represent the lenses (cluster members), the grey circles the foreground galaxy (darker grey) and the background galaxy (lighter grey). The cyan triangles, magenta squares and red stars represent the images of the three sources, respectively at $z_{s1} = 1.5$, $z_{s2} = 2.0$ and $z_{s3} = 2.5$.

total mass enclosed within the multiple images. This, indeed, is what strong lensing constrains tightly. As for the image positions, the single lens systems predicted images are all offset by $<0.8''$ in both x and y direction. In terms of total-rms, the image offset is $\sim 0.4''$. The magnification of the single plane configuration appears to be in general different, within a ratio of 0.8 to 1.8 with respect to the input.

A.4. Cluster and single perturber modelling

To investigate the effects of the single perturbers we model the same system with the cluster and the foreground lens only and the cluster and the background lens only, respectively. We find that the addition of the foreground perturber to the single-plane model allows us to recover parameters that are more similar to that of the multi-plane model. Indeed, as shown in Table A.1, the halo centroid position is offset by only $\sim 0.25''$, its mass is not as big as that of the single-plane model, and the halo slope is recovered within the error. If we include only the background perturber, instead, we get the same parameter values, within the errors, as that of the single plane model, confirming results from previous studies (e.g. McCully et al. 2014), which showed that foreground perturbers have a more significant effect on the modelling compared to the background ones. In terms of image positions rms, we find a total rms of $0.18''$ for the multi-plane model with only the foreground galaxy, while a total rms of $\sim 0.4''$ for the model with only the background galaxy, showing that the addition of the foreground perturber is more significant for a good fit of the image positions.

A.5. Effect of individual perturber

Before moving to a more realistic model, we performed a test to further investigate the effect of the individual perturbers along the LOS. We used the mock cluster lensing mass distribution 1 and we increase the mass of the background perturber such that the scaled Einstein radii of the two perturbers, for the intermediate source at $z_{s2} = 2$, is equal, namely

Table A.1. Constraints on lens parameters for different models of mock cluster lensing mass distribution 1.

Parameters	Input	MP-full	SP	MP-fore	MP-back	MP-wt
$\theta_{E, \text{fd}} ["]$	2.00	$1.99^{+0.07}_{-0.07}$	–	$1.60^{+0.19}_{-0.20}$	–	$2.30^{+0.08}_{-0.08}$
$x_{\text{halo}} ["]$	0.00	$0.01^{+0.05}_{-0.05}$	$0.98^{+0.08}_{-0.08}$	$0.21^{+0.11}_{-0.11}$	$1.15^{+0.09}_{-0.09}$	$-0.04^{+0.05}_{-0.05}$
$y_{\text{halo}} ["]$	0.00	$0.00^{+0.05}_{-0.05}$	$1.12^{+0.07}_{-0.06}$	$0.25^{+0.12}_{-0.12}$	$1.07^{+0.08}_{-0.07}$	$-0.04^{+0.05}_{-0.05}$
$\frac{b}{a}_{\text{halo}}$	0.80	$0.80^{+0.01}_{-0.01}$	$0.84^{+0.02}_{-0.03}$	$0.83^{+0.01}_{-0.02}$	$0.85^{+0.02}_{-0.02}$	$0.82^{+0.01}_{-0.01}$
$\theta_{\text{halo}} [\text{rad}]$	0.000	$0.000^{+0.004}_{-0.004}$	$3\pi^{+0.01}_{-0.01}$	$\pi^{+0.01}_{-0.01}$	$-0.97\pi^{+0.01}_{-0.01}$	$-0.004^{+0.003}_{-0.003}$
$\theta_{E, \text{halo}} ["]$	10.00	$10.00^{+0.04}_{-0.04}$	$14.10^{+0.73}_{-0.77}$	$12.80^{+0.46}_{-0.47}$	$14.10^{+0.86}_{-0.89}$	$11.40^{+0.31}_{-0.32}$
$r_{c, \text{halo}} ["]$	2.00	$2.00^{+0.35}_{-0.35}$	$6.10^{+0.93}_{-1.20}$	$3.40^{+1.60}_{-0.99}$	$6.60^{+1.10}_{-1.30}$	$2.40^{+0.39}_{-0.37}$
γ_{halo}	0.40	$0.40^{+0.02}_{-0.02}$	$0.70^{+0.07}_{-0.09}$	$0.49^{+0.10}_{-0.06}$	$0.69^{+0.08}_{-0.11}$	$0.41^{+0.02}_{-0.02}$
$\theta_{E, \text{g}} ["]$	2.00	$2.00^{+0.05}_{-0.05}$	$2.00^{+0.24}_{-0.20}$	$1.88^{+0.14}_{-0.12}$	$1.80^{+0.22}_{-0.19}$	$2.13^{+0.05}_{-0.05}$
$\theta_{E, \text{bd}} ["]$	2.00	$2.00^{+0.23}_{-0.24}$	–	–	$1.50^{+0.55}_{-0.61}$	$2.02^{+0.22}_{-0.20}$
$\text{rms}_{\text{best-fit}} ["]$		8×10^{-4}	0.41	0.18	0.39	0.028
$\chi^2_{\text{best-fit}}$		2×10^{-3}	587	116	532	3

Notes. The first column refers to the values used to simulate, the other columns refer to different models, in order, the full multi-plane, the single cluster-plane, the multi-plane with only foreground perturber, the multi-plane with only background perturber and the total multi-plane with the truncation radii of the galaxies fixed to the wrong value (10" instead of 15"). The values are the medians of the posterior probability distributions of the lens parameters together with their 1σ uncertainties. The orientation is measured counter clockwise from positive x -axis.

$$\frac{D_{\text{fd-s2}}}{D_{\text{s2}}} \theta_E^{\text{fd}} = \frac{D_{\text{bd-s2}}}{D_{\text{s2}}} \theta_E^{\text{bd}}. \quad (\text{A.1})$$

We resimulated the three sets of multiple image positions using Eq. (17). We then repeated the analysis from Sect. A.4 and we find that the perturber in the foreground is still more significant (rms $\sim 0''.32$) than the background (rms $\sim 0''.38$), but this effect is less prominent. This is due to two main factors: first that we have equated the strength of the two perturbers, but they are still at different locations with respect to the cluster centre; second is the lensing effect that the foreground has on the cluster. Moreover, in order for the scaled deflection angle of foreground and background galaxy to match, the background mass should be substantially larger than the foreground, meaning with an Einstein radius of $\sim 1''$ bigger than that of the foreground. So for typical scenarios where foreground galaxies have similar (or higher) masses as background galaxies, we should pay more attention to the foreground lenses.

A.6. Truncation radius

We tried to model the truncation radii of the galaxies, that is, roughly the half-mass radius (Elíasdóttir et al. 2007). At first we allowed the truncation radii of all the cluster galaxies (which we assumed to be equal) and of the two perturber galaxies to vary and we find that they are actually very difficult to constrain. We then fixed the truncation radii for the two perturber and model varying the truncation radius only for the cluster galaxies. In this case we find that the truncation radius posterior has a flat distribution. We therefore suspect that its value does not really affect the parameter modelling, as long as it is not significantly different from its true value. This is probably due to the limited amount of information and to the limited extent of the multiple image region ($\sim 15''$) of the system, since in this particular mock configuration we chose the truncation radius for cluster members to be comparable to the size of the multiple image region. To explore this further we modelled the multi-plane system keeping all the

truncation radii fixed to 10", instead of the input 15", value that is outside of the 1σ confidence intervals for the posterior probability distribution of the cluster members' truncation radius. We find that this wrong choice of truncation radii makes the estimation of the halo mass wrong by $\sim 1''.4$ and that of the cluster galaxies by $\sim 0''.13$, so non-negligible, as shown in Table A.1.

A.7. Conclusions

The simulated system mock cluster mass distribution 1 is composed of a cluster, at $z_c = 0.4$, with five spherical galaxies of equal luminosity and mass, and one dark matter halo with an Einstein radius of 10". Its environment is constituted by two close-in-projections LOS galaxies, one foreground $z_{\text{fd}} = 0.2$ and one background $z_{\text{bd}} = 0.6$, with equal, large mass. Our observables are a set of 15 images from three sources at redshifts, respectively $z_{\text{S1}} = 1.5$, $z_{\text{S2}} = 2.0$, $z_{\text{S3}} = 2.5$. For this cluster we studied the different effects of foreground and background perturbers. We conclude that:

1. Foreground perturbers have a more significant effect in the modelling than the background, for a given equal mass, and their inclusion reproduces the observed image positions more accurately. This might be due to their lensing effects on the observed image positions. Moreover, foreground perturbers have lensing effects on the halos, affecting their centroid position and shape. If we try to equate the scaled Einstein radius (Eq. (A.1)) of foreground and background perturbers we see that this trend is slightly attenuated, but still visible.
2. The truncation radius of the perturbers does not affect substantially the lensing model, since its extent is comparable to that of the multiple image region. Indeed its posterior distribution appears flat and does not appear to be correlated to the parameters of other profiles.

This is also confirmed by the mock MACS model, as discussed in Sect. 4.4.8.