Astronomy
&
Astrophysics

Cluster strong lensing with hierarchical inference

Formalism, functional tests, and public code release

P. Bergamini[1], A. Agnello[2], and G. B. Caminha[3,4]

[1] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy
[2] DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen Ø, Denmark
    e-mail: adriano.agnello@nbi.ku.dk
[3] Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands
[4] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

ABSTRACT

Context. Lensing by galaxy clusters is a versatile probe of cosmology and extragalactic astrophysics, but the accuracy of some of its predictions is limited by the simplified models adopted to reduce the (otherwise intractable) number of degrees of freedom.
Aims. We aim to explore cluster lensing models in which the parameters of all cluster member galaxies are free to vary around some common scaling relations with non-zero scatter and deviate significantly from these relations if, and only if, the data require this.
Methods. We devised a Bayesian hierarchical inference framework that enables the determination of all lensing parameters and the scaling relation hyperparameters, including intrinsic scatter, from lensing constraints and (if given) stellar kinematic measurements. We achieve this through BAYESLENS, a purpose-built wrapper around common parametric lensing codes that can sample the full posterior on parameters and hyperparameters; we release BAYESLENS with this paper.
Results. We ran functional tests of our code against simple mock cluster lensing datasets with realistic uncertainties. The parameters and hyperparameters are recovered within their 68% credibility ranges and the positions of all the "observed" multiple images are accurately reproduced by the BAYELENS best-fit model, without over-fitting.
Conclusions. We have shown that an accurate description of cluster member galaxies is attainable, despite a large number of degrees of freedom, through fast and tractable inference. This extends beyond the most updated cluster lensing models. The precise impact on studies of cosmography, galaxy evolution, and high-redshift galaxy populations can then be quantified on real galaxy clusters. While other sources of systematics exist and may be significant in real clusters, our results show that the contribution of intrinsic scatter in cluster member populations can now be controlled.

Key words. gravitational lensing: strong – methods: numerical – galaxies: clusters: general – cosmology: observations – dark matter

1. Introduction

Galaxy clusters (at $z \approx 0.3-0.9$) are ideal cosmic telescopes to study background galaxies out to $z \approx 7$ and they boost the lensing signal of their own galaxies. When galaxies reside in clusters, their lensing cross section is enhanced, allowing the study of galaxy populations at $z \approx 0.2-0.4$ over a wide mass range (down to $M_\star \approx 10^{9.5} M_\odot$; e.g., Keeton 2003; Grillo et al. 2014; Parry et al. 2016; Niemiec et al. 2017; Bergamini et al. 2019). If the total mass density distributions of lensing clusters are accurately reconstructed, cosmological parameters can also be inferred (e.g., Golse et al. 2002; Gilmore & Natarajan 2009; Caminha et al. 2016; Magaña et al. 2018; Grillo et al. 2018). Dedicated *Hubble* Space Telescope (HST) surveys, such as the Cluster Lensing And Supernova survey with *Hubble* (CLASH; Postman et al. 2012), the *Hubble* Frontier Fields program (HFF; Lotz et al. 2017), and the Re-ionization Lensing Cluster Survey (RELICS; Coe et al. 2019), have enabled the identification and multi-band characterization of tens, sometimes hundreds, of multiple images of background sources, and hundreds of cluster member galaxies per cluster so as to constrain cluster lens models (Priewe et al. 2017). Subsequent multi-slit spectroscopic follow-up observations, such as the CLASH-VLT

survey (Rosati et al. 2014) has gathered a wealth of multi-object spectroscopic data on cluster member galaxies and background sources, sampling light cones around 13 clusters ($\approx 20 \times 20$ arcmin^2). Owing to their extent and magnifying power, galaxy clusters have also been used to study faint and high-redshift sources (e.g., Ishigaki et al. 2015; Livermore et al. 2015; Kawamata et al. 2016a; Bouwens et al. 2017; Atek et al. 2018; Cerny et al. 2018; Hoag et al. 2018; Vanzella et al. 2017, 2019; Hashimoto et al. 2019). Since these studies may have implications on our understanding of star formation at high redshift and re-ionization, accurate magnification maps are required.

The advent of the Multi-Unit Spectroscopic Explorer (MUSE; Bacon et al. 2012) has enabled integral-field spectroscopy of galaxy clusters. In particular, in recent years numerous MUSE observations have targeted the cores of several massive clusters (e.g., Richard et al. 2014; Jauzac et al. 2015; Limousin et al. 2016; Kawamata et al. 2016b; Lagattuta et al. 2017; Caminha et al. 2016, 2017a,b; Vanzella et al. 2019). Every MUSE pointing provides a data cube with a field of view of 1 arcmin^2 and spatial sampling of 0.2″. The spectra cover $4750 < \lambda < 9350$ Å with a resolution of ~2.6 Å, almost constant along the whole wavelength range, and a spectral sampling of 1.25 Å pix^{-1} (the spectral resolution is robustly characterized; see

Guérou et al. 2017). Integral-field spectroscopy of cluster cores enabled a secure identification of tens of multiple images and their redshifts and complete and pure sets of member galaxies.

Besides the identification of cluster members and multiple images, spectroscopy provides further constraints on mass models through kinematics. The homogeneous spectroscopic coverage of CLASH-VLT means that velocities of individual cluster members can be used to aid the cluster profile reconstruction through dynamical modeling (Biviano et al. 2013; Sartoris et al. 2020). Similarly, integral-field follow-up with MUSE yields the internal stellar kinematics of cluster members and allows us to independently constrain their mass profiles.

Lensing clusters have undergone an extensive modeling effort by multiple independent teams. Up to now, the constraints consisted in the positions of tens of multiple images per cluster (e.g., Richard et al. 2014; Grillo et al. 2015; Jauzac et al. 2015; Limousin et al. 2016; Kawamata et al. 2016b; Caminha et al. 2016, 2017a,b, 2019; Lagattuta et al. 2017; Bonamigo et al. 2018; Bergamini et al. 2021). Mostly, lens models describe these clusters as a superposition of extended dark matter (DM) halos and more localized over-densities corresponding to cluster member galaxies. So-called nonparametric, grid-based models (e.g., Dye & Taylor 1998; Saha et al. 2001; Bradač et al. 2005; Diego et al. 2005) only add mass over-densities where they are required by the data. Alternatively, parametric models decompose the cluster potential in different (smooth) components that are expected to follow a physical hypothesis on how mass should be distributed in galaxies and clusters (see, e.g., Jullo et al. 2007, for an overview). In particular, Natarajan & Kneib (1997) argued that including cluster member galaxies explicitly, like individual components of cluster mass models, is crucial. This has also emerged from the HFF model comparison project (Meneghetti et al. 2017), in which parametric models routinely outperformed current versions of free-form models, both on semi-analytical and on fully simulated mocks. Toward the era of precision cosmology with galaxy clusters, free-form models may eventually replace parametric models, possibly resolving some of their built-in rigidity (e.g., Rodney et al. 2018). However, given the current performance of parametric and free-form models, we may safely adopt the working hypothesis that galaxy clusters can be described as a superposition of relatively simple galaxy and halo components.

The bottleneck in cluster lens models consists of $\approx 0.3''-0.6''$ errors in image-position reconstruction (e.g., Grillo et al. 2016), which is significantly higher than current positional uncertainties from HST data, and the appreciable discrepancies in magnification maps produced by different models on the same clusters (Meneghetti et al. 2017). One cause of such discrepancies may be the rigidity of the models used to reduce the degrees of freedom associated with cluster members, since galaxies are currently modeled as belonging to zero-scatter scaling relations, typically with hyperparameters that are imposed externally (e.g., Limousin et al. 2007). The freedom and accuracy in lensing models have gained importance in the era of high-accuracy cosmography (see, e.g., Treu et al. 2016), and on the reliability of magnification maps for the study of high-redshift galaxy populations (Bouwens et al. 2017). Up to now, departures of individual galaxies from prescribed scaling relations were determined heuristically on a galaxy-by-galaxy basis (Jauzac et al. 2018). Ideally, all galaxies should be let free to vary around some finite-scatter scaling relations, whose parameters (including scatter) are to be determined directly from the data on each given cluster.

The solution is a hierarchical Bayesian inference formalism, where each galaxy has its own associated parameters, and the parameters of all galaxies are posited to be drawn from common relations with hyperparameters to be determined through lens modeling and (if given) auxiliary kinematic information. In this paper, we illustrate this formalism and its application to lensing models of three simple mock clusters of increasing complexity with HFF and CLASH-VLT data quality, accounting for all observational constraints in a self-consistent manner. This ensures that cluster lens models are as flexible as possible, given the data, and that higher accuracy is reached in predicted image positions. Since only one of the contributions to the inference is from lensing and to ensure a fair comparison with the most current technology, we build our inference as a modular wrapper of common lens modeling codes. This also ensures that further constraints, such as time delays, flux ratios, and shapes of background sources, can be easily included in the inference. Moreover, we perform various functional tests, comparing a hierarchical approach with the most updated zero-scatter models, to assess which sources of systematics can be resolved by this method, and show that the hierarchical models do not over-fit.

This paper is structured as follows. Our hierarchical inference is detailed in Sect. 2. Section 3 covers some technicalities inherent to our code implementation of the models. In Sect. 4, we perform functional tests on three simple, but realistic, mock clusters. Results are discussed in Sect. 5, and we conclude in Sect. 6.

While a cosmological model is adopted to generate and fit mock clusters, the main results of this work are general and separate from the choice of cosmology. If needed, cosmological parameters may be sampled as additional hyperparameters in our inference scheme.

2. A hierarchical cluster lensing model

In this section, we generalize the formalism of parametric cluster lensing models to allow for hierarchical inference, in particular scaling relations with non-zero scatter. Given the abundance of symbols, in the following (Sects. 2.2 and 2.3) all model parameters and hyperparameters are marked by a "hat" symbol.

2.1. Cluster lensing components

In a parametric cluster lensing model, the total mass distribution is typically divided into a few component classes (Jullo et al. 2007). In this section, we indicate each class through its contribution, ϕ^{class}, to the lensing potential. A first, smooth, cluster-scale component (ϕ^{halo}) accounts for both the DM content of the cluster and the baryonic intra-cluster gas and light contributions. A second clumpy component describes the mass in cluster member halos (ϕ^{gal}), in DM and baryons. A third component accounts for the presence of massive structures in the outer cluster regions, and possibly additional massive halos along the line of sight ($\phi^{\mathrm{shear+los}}$). Within this model class, the total cluster potential is written as

$$\phi_{\mathrm{tot}} = \sum_{i=1}^{N^h} \phi_i^{\mathrm{halo}} + \sum_{j=1}^{N_{\mathrm{tot}}^{\mathrm{gal}}} \phi_j^{\mathrm{gal}} + \sum_{k=1}^{N^{\mathrm{sl}}} \phi_k^{\mathrm{shear+los}}, \qquad (1)$$

where N^h, $N_{\mathrm{tot}}^{\mathrm{gal}}$, N^{sl} are the number of cluster scale halos, cluster members, and shear plus line-of-sight contributions, respectively.

The number of constraints given by the observed multiple image positions is usually not sufficient to fit the parameters of mass profiles used to describe each individual galaxy in a given

cluster lens model. For this reason, cluster members are usually parameterized as circular dual pseudo-isothermal mass distributions (circular dPIEs; Limousin et al. 2005, Elíasdóttir et al. 2007), with negligible core radii, whose parameters are related to galaxy luminosities according to fixed scaling relations.

The circular dPIE is defined by the 3D mass density (Limousin et al. 2005)

$$\rho(r) = \frac{\rho_0}{(1 + r^2/r_{\text{core}}^2)(1 + r^2/r_{\text{cut}}^2)}, \tag{2}$$

with

$$\rho_0 = \frac{\sigma_0^2}{2\pi G} \frac{r_{\text{cut}} + r_{\text{core}}}{r_{\text{core}}^2 r_{\text{cut}}}. \tag{3}$$

In these equations, σ_0, r_{core}, and r_{core} are the central velocity dispersion, core radius, and truncation radius of the dPIE, respectively. For a vanishing r_{core}, r_{cut} encircles about half of the total 3D dPIE mass and 60% of the projected mass. In the limit $r_{\text{cut}} \to \infty$ and $r_{\text{core}} > 0$, the dPIE coincides with the pseudo-isothermal mass distribution defined in Kassiola & Kovner (1993) (PIEMD). In lensing models of galaxy clusters, the following scaling relations are commonly adopted for the central velocity dispersions (σ_0^{gal}), core radii ($r_{\text{core}}^{\text{gal}}$), and truncation radii ($r_{\text{cut}}^{\text{gal}}$) of cluster members:

$$\sigma_{0,i}^{\text{gal}} = \sigma_0^{\text{ref}} \left(\frac{L_i}{L_0}\right)^{\alpha} \ , \ r_{\text{core},i}^{\text{gal}} = r_{\text{core}}^{\text{ref}} \left(\frac{L_i}{L_0}\right)^{\beta_{\text{core}}} \ , \ r_{\text{cut},i}^{\text{gal}} = r_{\text{cut}}^{\text{ref}} \left(\frac{L_i}{L_0}\right)^{\beta_{\text{cut}}} \ , \tag{4}$$

without any intrinsic scatter. Under the hypothesis of a power-law scaling $M_{\text{tot},i}/L_i \propto L_i^{\gamma}$ of the total mass of a cluster member, and since the total dPIE mass is $M_{\text{tot}} = (\pi\sigma_0^2 r_{\text{cut}})/G$, the exponents are therefore related through

$$\beta_{\text{cut}} = \gamma - 2\alpha + 1 \ . \tag{5}$$

2.2. Measured kinematics

Spectroscopic information, when incorporated inside the lensing models, produces robust and accurate reconstructions of the projected cluster masses. However, strong lensing models still suffer from internal degeneracies between the parameters of their mass components (Eq. (1)). A degeneracy exists, for example, between the two normalizations, σ_0^{ref} and $r_{\text{cut}}^{\text{ref}}$, in Eq. (4). This degeneracy is easily understandable if we consider that the total mass of a dPIE, with negligible r_{core}, inside an aperture of radius R is (Jullo et al. 2007) written as

$$M(R) = \frac{\pi\sigma_0^2}{G} \left(R + r_{\text{cut}} - \sqrt{r_{\text{cut}}^2 + R^2}\right), \tag{6}$$

and that multiple images constrain only the total mass within their (projected) distance from the center of a galaxy. Other degeneracies also exist between the cluster-scale DM halo parameters and the clumpy subhalo components since the overall mass may be redistributed differently between the main halo and the subhalos, except where the multiple-image positions are strongly constraining.

Recently, velocity dispersions of cluster members (from fits to spectra with signal-to-noise ratio greater than ten) were used to break, or at least reduce, these internal degeneracies (Verdugo et al. 2007; Monna et al. 2015, 2017 and more extensively by Bergamini et al. 2019). Monna et al. (2015, 2017)

fixed the velocity dispersions of several cluster galaxies to their measured values, thereby breaking their model degeneracies by imposing a strong hypothesis on their mass content. In a more flexible approach, Bergamini et al. (2019) used high-quality MUSE spectra of cluster member galaxies to build priors on the parameters of the scaling relations in Eq. (4).

In this work, we combine both of the above to fully characterize the subhalo components of clusters. Our purpose-built BAYESLENS wrapper uses the available measured velocity dispersions of the cluster galaxies, $\sigma_{\text{m}}^{\text{gal}} \pm \delta\sigma_{\text{m}}^{\text{gal}}$, to infer the hyperparameters ($\hat{\sigma}^{\text{ref}}$, $\hat{\alpha}$) of the σ-mag scaling relation. A third hyperparameter, $\hat{\Delta\sigma}^{\text{ref}}$, quantifies the scatter of the measured galaxies around the σ-mag scaling relation. Gaussian priors, centered on the measured $\sigma_{\text{m}}^{\text{gal}}$ and with standard deviations equal to five times the measured errors $\delta\sigma_{\text{m}}^{\text{gal}}$, are adopted in the lens model for the dPIE velocity dispersions of galaxies with measured velocity dispersions. For galaxies without a measured velocity dispersion, we assumed Gaussian priors centered on the inferred σ-mag scaling relation and with a standard deviation equal to $\hat{\Delta\sigma}^{\text{ref}}$. We adopt a uniform prior for the reference cut radius of the r_{cut}-mag scaling relation, and β_{cut} is obtained from Eq. (5). Unless otherwise stated, all model hyperparameters are left free to vary to fully explore the model posterior probability as described below. All hat symbols are introduced more formally in the following subsection (Sect. 2.3).

2.3. Fitting it all together: the posterior

In our models, we use the measured velocity dispersions of $N_{\text{m}}^{\text{gal}}$ cluster galaxies, $\sigma_{\text{m}}^{\text{gal}} \pm \delta\sigma_{\text{m}}^{\text{gal}}$, together with the positions $\boldsymbol{x}_{\text{im}}$ of N^{im} multiple images, from N^{fam} different sources with positions $\boldsymbol{x}_{\text{sou}}$, as observational constraints to the lens model free-parameters. Hereafter, these free-parameters are marked with a hat symbol. To explore the lens models, we sample the total posterior probability function p_{tot}, expressed as the product of: the odds p_{g} of drawing each galaxy (independently) from a given scaling relation (with intrinsic scatter); the likelihood p_{sr} of the scaling relation hyperparameters, given the measured velocity dispersions and luminosities; the likelihood p_{im} of reproducing the observed image positions, given the cluster model (including the parameters of each cluster member galaxy); and a prior p_{h} on the parameters of the main cluster-scale halo(s). We also include a term p_{mg} that links the dPIE lensing velocity dispersions of the galaxies with measured kinematics to their stellar velocity dispersions, with large uncertainties. Although not formally needed, this term is only used as a loose regularization to ensure convergence.

In synthesis,

$$p_{\text{tot}}\left(\hat{\sigma}^{\text{ref}}, \hat{\alpha}, \hat{\Delta\sigma}^{\text{ref}}, \hat{r}_{\text{cut}}^{\text{ref}}, \hat{\sigma}_{\text{m}}^{\text{gal}}, \hat{\sigma}^{\text{gal}}, \hat{\phi}^h \mid \text{mag}^{\text{gal}} \sigma_{\text{m}}^{\text{gal}}, \delta\sigma_{\text{m}}^{\text{gal}}, \boldsymbol{x}_{\text{im}}\right)$$

$$\propto p_{\text{sr}}\left(\hat{\sigma}^{\text{ref}}, \hat{\alpha}, \hat{\Delta\sigma}^{\text{ref}}, \hat{r}_{\text{cut}}^{\text{ref}} \mid \text{mag}^{\text{gal}} \sigma_{\text{m}}^{\text{gal}}, \delta\sigma_{\text{m}}^{\text{gal}}\right)$$

$$\times p_{\text{mg}}\left(\hat{\sigma}_{\text{m}}^{\text{gal}} \mid \sigma_{\text{m}}^{\text{gal}}, \delta\sigma_{\text{m}}^{\text{gal}}\right) \times p_{\text{g}}\left(\hat{\sigma}^{\text{gal}} \mid \hat{\sigma}^{\text{ref}}, \hat{\alpha}, \hat{\Delta\sigma}^{\text{ref}}\right)$$

$$\times p_h\left(\hat{\phi}^h\right) \times p_{\text{im}}\left(\boldsymbol{x}_{\text{im}} \mid \hat{\sigma}^{\text{ref}}, \hat{\alpha}, \hat{\Delta\sigma}^{\text{ref}}, \hat{r}_{\text{cut}}^{\text{ref}}, \hat{\sigma}_{\text{m}}^{\text{gal}}, \hat{\sigma}^{\text{gal}}, \hat{\phi}^h, \boldsymbol{x}_{\text{sou}}\right). \tag{7}$$

Each of the five factors on the right-hand-side is discussed below. In the following, the quantities referring to cluster galaxies with measured velocity dispersions are marked with the subscript "m".

Some of the factors (e.g., the one on hyperparameters) can be interpreted as a posterior on some parameters given some observations, which (by the Bayes chain rule) can further be used as a prior for the full hierarchical posterior. To avoid confusion between those posteriors and the final posterior, we alternatively refer to them as term or odds in the following.

2.3.1. Odds on the scaling relation hyperparameters, p_{sr}

This factor is responsible for the σ-mag and r_{cut}-mag scaling relation hyperparameters, given the set of N_m^{gal} measured cluster galaxies. For the galaxy velocity dispersions, we consider a scaling relation of the same form of Eq. (4) parameterized by the reference measured velocity $\hat{\sigma}^{ref}$ and slope $\hat{\alpha}$ plus an intrinsic scatter $\hat{\Delta\sigma}^{ref}$ in measured velocity dispersions. Regarding the r_{cut}-mag scaling relation, in the current version of our models we only optimize the reference value $\hat{r}_{cut}^{ref}$ while the slope β_{cut} is determined using Eq. (5) from the inferred $\hat{\alpha}$ and assuming a fixed mass-to-light scaling for the cluster galaxies. No scatter around this relation is considered.

The term p_{sr} can be expressed as

$$
\begin{aligned}
p_{sr}&\left(\hat{\sigma}^{ref}, \hat{\alpha}, \hat{\Delta\sigma}^{ref}, \hat{r}_{cut}^{ref} \mid mag^{gal}, \sigma_m^{gal}, \delta\sigma_m^{gal}\right) \\
&\propto p_{0,sr}\left(\sigma_m^{gal} \mid mag^{gal}, \delta\sigma_m^{gal}, \hat{\sigma}^{ref}, \hat{\alpha}, \hat{\Delta\sigma}^{ref}\right) \\
&\times \varpi_{sr}\left(\hat{\sigma}^{ref}, \hat{\alpha}, \hat{\Delta\sigma}^{ref}, \hat{r}_{cut}^{ref}\right),
\end{aligned}
\tag{8}
$$

where the uninformative prior ϖ_{sr} is defined by

$$
\ln\left\{\varpi_{sr}\left(\hat{\sigma}^{ref}, \hat{\alpha}, \hat{\Delta\sigma}^{ref}\right)\right\} =
\begin{cases}
-\ln(\hat{\Delta\sigma}^{ref}), & \text{if } \sigma_{min}^{ref} < \hat{\sigma}^{ref} < \sigma_{max}^{ref} \\
& \text{and } \alpha_{min} < \hat{\alpha} < \alpha_{max} \\
& \text{and } \Delta\sigma_{min}^{ref} < \hat{\Delta\sigma}_m < \Delta\sigma_{max}^{ref}, \\
& \text{and } r_{cut,min}^{ref} < \hat{r}_{cut}^{ref} < r_{cut,max}^{ref} \\
-\infty, & \text{otherwise}
\end{cases}
\tag{9}
$$

and limits the σ-mag plus scatter scaling relation hyperparameters and the reference truncation radius, $\hat{r}_{cut}^{ref}$, to lie within suitably chosen boundaries $\sigma_{min(max)}^{ref}$, $\alpha_{min(max)}$, $\Delta\sigma_{min(max)}^{ref}$, and $r_{cut,min(max)}^{ref}$. The log-likelihood term

$$
\begin{aligned}
&\ln\left\{p_{0,sr}\left(\sigma_m^{gal} \mid mag^{gal}, \delta\sigma_m^{gal}, \hat{\sigma}^{ref}, \hat{\alpha}, \hat{\Delta\sigma}^{ref}\right)\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{N_m^{gal}}\left[\frac{\left(\sigma_{m,i}^{gal} - \hat{\sigma}_{m,i}^{sr}\right)^2}{\left(\delta\sigma_{m,i}^{gal}\right)^2 + \left(\hat{\Delta\sigma}^{ref}\right)^2} \right. \\
&\quad \left. + \ln\left\{2\pi\left[\left(\delta\sigma_{m,i}^{gal}\right)^2 + \left(\hat{\Delta\sigma}^{ref}\right)^2\right]\right\}\right]
\end{aligned}
\tag{10}
$$

quantifies the goodness of fit of scaling relation hyperparameters to the measured velocity dispersions. In Eq. (10), we define $\hat{\sigma}_{m,i}^{sr}$ as

$$
\hat{\sigma}_{m,i}^{sr} = \hat{\sigma}^{ref} 10^{0.4\,\hat{\alpha}\left(mag^{ref} - mag_i^{gal}\right)},
\tag{11}
$$

where mag^{ref} corresponds to the reference luminosity L_0 in Eq. (4).

2.3.2. Term on measured galaxies, p_{mg}

This term only applies to the N_m^{gal} galaxies with measured velocity dispersions. It allows for residual uncertainties (e.g., from mass-anisotropy degeneracy and asphericity) in converting measured velocity dispersions in dPIE $\hat{\sigma}_m^{gal}$, which in turn propagate on the scaling relations that all galaxies are drawn from. This is attained by linking the kinematic sigma to the lensing sigma with some tolerance. If this were not done, we might obtain biased results if the kinematic sigma-mag relation has some intrinsic scatter, but the lensing sigma-mag does not. This term is also needed to account for systematic uncertainties in the measured velocity dispersions, which in the high signal-to-noise ratio regime can dominate over the statistical uncertainties. Erring on the conservative side, we choose the conversion tolerance to five times the uncertainties in the measured velocity dispersion.

Therefore, p_{mg} alone consists of Gaussian priors centering the cluster member aperture-average velocity dispersions $\hat{\sigma}_m^{gal}$ on their kinematic values σ_m^{gal}. We choose the standard deviation of Gaussian priors equal to five times the error on the kinematic measurements $\delta\sigma_m^{gal}$. This term is given by

$$
\begin{aligned}
&\ln\left\{p_{mg}\left(\hat{\sigma}_m^{gal} \mid \sigma_m^{gal}, 5\,\delta\sigma_m^{gal}\right)\right\} \\
&= -\frac{1}{2}\sum_{i=0}^{N_m^{gal}}\left\{\frac{\left(\hat{\sigma}_{m,i}^{gal} - \sigma_{m,i}^{gal}\right)^2}{\left(5\delta\sigma_{m,i}^{gal}\right)^2} + \ln\left[2\pi\left(5\,\delta\sigma_{m,i}^{gal}\right)^2\right]\right\}.
\end{aligned}
\tag{12}
$$

In other words, this term attributes to galaxies their measured velocity dispersions, with very wide tolerance, unless a deviation from these values produces a significant improvement of the lensing model. We emphasize that this term is, strictly speaking, not necessary and formally it may prevent the finite-scatter models from reducing to zero-scatter models (if this is needed by the data). However, given the typical $\delta\sigma_m^{gal} \gtrsim 15\,km\,s^{-1}$ uncertainties on measured kinematics, the $5\,\delta\sigma_m^{gal}$ term ensures that this factor only acts as a loose regularization, preventing pathological solutions and aiding the convergence of the models.

2.3.3. Prior on unmeasured galaxies, p_g

This term is a collection of Gaussian priors on velocity dispersion values for the $N^{gal} = N_{tot}^{gal} - N_m^{gal}$ galaxies without kinematics measurements. Its form is such that the final posterior prefers lens models in which the unmeasured galaxies lie on the σ-mag scaling relation inferred by p_{sr}, unless otherwise required by the lensing data, as follows:

$$
\begin{aligned}
&\ln\left\{p_g\left(\hat{\sigma}^{gal} \mid \hat{\sigma}^{ref}, \hat{\alpha}, \hat{\Delta\sigma}^{ref}\right)\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{N^{gal}}\left\{\frac{\left(\hat{\sigma}_i^{gal} - \hat{\sigma}_i^{sr}\right)^2}{\left(\hat{\Delta\sigma}^{ref}\right)^2} + \ln\left[2\pi\left(\hat{\Delta\sigma}^{ref}\right)^2\right]\right\}.
\end{aligned}
\tag{13}
$$

The $\hat{\sigma}^{sr}$ are computed through Eq. (11), but now considering galaxies without measured velocity dispersions.

2.3.4. Prior on halo parameters, p_h

This term consists of flat priors on the smooth cluster-scale halo parameters collectively indicated as $\hat{\phi}^h$ and is written as

$$\ln\left\{p_h\left(\hat{\phi}^h\right)\right\} = \begin{cases} 0, & \text{if } \phi_{\min}^h < \hat{\phi}^h < \phi_{\max}^h \\ -\infty, & \text{otherwise} \end{cases} . \qquad (14)$$

If (as follows) these halos are parameterized as PIEMD profiles, p_h is a prior on the sky coordinates $\hat{x}^h$ and $\hat{y}^h$, on the ellipticity $\hat{e}^h$, position angle $\hat{\theta}^h$, core radius $\hat{r}_{\text{core}}^h$, and central velocity dispersion $\hat{\sigma}_0^h$.

2.3.5. Multiple-image likelihood, p_{im}

This final term is a likelihood function quantifying the agreement between observed and lens model predicted multiple-image positions. Given N^{fam} sources with N_i^{im} multiple images associated with the same i-th source (usually called a family), we follow Jullo et al. (2007) and express the likelihood as

$$p_{\text{im}}\left(\boldsymbol{x}_{\text{im}} \mid \hat{\sigma}^{\text{ref}}, \hat{\alpha}, \hat{\Delta}\sigma^{\text{ref}}, \hat{r}_{\text{cut}}^{\text{ref}}, \hat{\sigma}_{\text{m}}^{\text{gal}}, \hat{\sigma}^{\text{gal}}, \hat{\phi}^h, \boldsymbol{x}_{\text{ou}}\right)$$

$$= \prod_{i=0}^{N^{\text{fam}}} \frac{e^{-\chi_i^2/2}}{\prod_j \Delta x_{i,j}\sqrt{2\pi}}, \qquad (15)$$

where the χ_i^2 associated with the i-th family is

$$\chi_i^2 = \sum_{j=1}^{N_i^{\text{im}}} \frac{\left\|\boldsymbol{x}_{i,j}^{\text{obs}} - \boldsymbol{x}_{i,j}^{\text{pred}}\right\|^2}{\Delta x_{i,j}^2}, \qquad (16)$$

where $\boldsymbol{x}_{i,j}^{\text{obs}}$ are the observed positions of the multiple images on the lens plane, $\Delta x_{i,j}$ are the isotropic uncertainties on these positions, and $\boldsymbol{x}_{i,j}^{\text{pred}}$ are the model predicted positions, given the adopted cosmology and the following inferred set of model parameters: $\hat{\sigma}^{\text{ref}}, \hat{\alpha}, \hat{\Delta}\sigma^{\text{ref}}, \hat{r}_{\text{cut}}^{\text{ref}}, \hat{\sigma}_{\text{m}}^{\text{gal}}, \hat{\sigma}^{\text{gal}}, \hat{\phi}^h$.

3. Technicalities

To sample the complete posterior in Eq. (7), we use the Affine-Invariant sampling as originally introduced by Goodman & Weare (2010), which is especially suited to our highly dimensional (and possibly degenerate) parameter space. In particular, to enable full portability and reproducibility, we use the latest `python` release[1] of `emcee` (Foreman-Mackey et al. 2013).

The first four terms in Eq. (7) are directly implemented in our code, while to compute the multiple-images likelihood p_{im} we exploit the publicly available software LENSTOOL (Kneib et al. 1996; Jullo et al. 2007; Jullo & Kneib 2009). The synergy between our code and LENSTOOL requires some technicalities, as described in the following subsections. In any case, our code is fully modular, so that LENSTOOL can be replaced by any other parametric lensing software by changing only a few `python` lines.

3.1. Calls to LENSTOOL: Input and output files

For each parameters combination, corresponding to a given walker position inside the parameters space, BAYESLENS silently

---

calls LENSTOOL to compute the p_{im} term of the total posterior. Every LENSTOOL call needs a different input file generated in our code by a specific `python` function. Another function reads the resulting likelihood, computed by LENSTOOL, from an output file. Since all these output files are saved on the disk using the same name, we create folders with unique (random) names to differentiate each LENSTOOL call. These folders are deleted at the end of every p_{im} computation.

To sample the posterior p_{tot}, millions of walker positions are required. Thus, millions of input and output LENSTOOL files are quickly created and deleted on the disk during this process. To avoid the disk wear and the bottleneck represented by the process of writing and reading files, part of our computer RAM is reserved for creating a RAM disk, where the input and output files are temporarily saved and deleted.

3.2. From measured σ_m^{gal} to LENSTOOL fiducial σ_{LT}

The dPIE is implemented in LENSTOOL through a fiducial velocity dispersion σ_{LT} related to the 1D central velocity dispersion σ_0 (in Eq. (3)) by $\sigma_0 = \sqrt{3/2}\sigma_{\text{LT}}$. To convert the model predicted aperture-average line-of-sight velocity dispersions $\hat{\sigma}_m^{\text{gal}}$ and $\hat{\sigma}^{\text{gal}}$ to their fiducial LENSTOOL values ($\hat{\sigma}_{(m),\text{LT}}^{\text{gal}}$), we relate them through

$$\hat{\sigma}_{(m),\text{LT}}^{\text{gal}} = \hat{\sigma}_{(m)}^{\text{gal}}/c_p(R) , \qquad (17)$$

where R is the aperture radius chosen for the cluster member spectral extraction. Adopting orbital isotropy and a (spherical) galaxy surface brightness profile proportional to the dPIE matter density, the projection coefficient $c_p(R)$ is given by

$$c_p^2(R) = \frac{6}{\pi}\frac{r_c + r_t}{r_c^2 r_t}\left(\sqrt{r_c^2 + R^2} - r_c - \sqrt{r_t^2 + R^2} + r_t\right)^{-1}$$

$$\times \int_0^R \int_R^\infty R' \frac{r_t \arctan\left(\frac{r}{r_t}\right) - r_c \arctan\left(\frac{r}{r_c}\right)}{(1 + r^2/r_c^2)(1 + r^2/r_t^2)}\frac{\sqrt{r^2 - R'^2}}{r^2}\mathrm{d}r\mathrm{d}R',$$

$$(18)$$

with $r_c = r_{\text{core}}$ and $r_t = r_{\text{cut}}$ (see, Bergamini et al. 2019).

4. Functional tests

To test the ability of BAYESLENS in recovering the correct halos and subhalos mass parameters and in predicting the correct multiple-image positions, we developed three mock galaxy clusters, namely Cluster-1, Cluster-2, and Cluster-3, with an increasing degree of complexity. Cluster-1 is a simple and well-controlled cluster mock, while Cluster-2 and 3 have the main scope to analyze the role of low-mass cluster substructures and accuracy in the multiple-image positions. Although none of the models claim to reproduce the complexity of real galaxy clusters, they represent ideal preliminary tests for our code, leaving to future works its application to more complex simulations (see Meneghetti et al. 2017) and to real cluster observations. In particular, the mocks allow us to verify the flexibility of our hierarchical approach, their dependence on the lensing or kinematics constraints, and their robustness against over-fitting. Since our simple mocks are developed using LENSTOOL, the results we obtain are directly comparable to model inputs, thus avoiding the possibility of parameter misidentifications in the discussion of the results.

In this section, we describe the main characteristics of the simulated clusters and the settings adopted in the BAYESLENS

optimization. We emphasize that these are functional tests, that is, on the behavior of the method itself in recovering the input parameters of simple but realistic mocks. These are the very first tests that any new method undergoes before it is tested on less well-controlled mocks (e.g., from simulations) to ensure that it can properly reproduce its own input parameters. On fully simulated or real-life clusters, the performance of each separate method may worsen because the true underlying mass distribution is unknown a priori, so the only meaningful comparison is among different models on well-controlled mocks.

4.1. Cluster-1: the simplest toy cluster model

Cluster-1 is based on the best-fit LENSTOOL lens model developed by Caminha et al. (2017b) (hereafter C-17) for the CLASH cluster MACS J1206.2−0847 at redshift $z = 0.439$. With respect to C-17, the cluster-scale component of our simulated cluster is made by a single halo, parametrized through a PIEMD mass profile. The center of the halo has an offset of 0.92″ from the brightest cluster galaxy (BCG) reference position. The halo has a core radius $r_{core}^h = 3''$, while its LENSTOOL fiducial velocity dispersion has a value of $\sigma_{LT}^h = 1000 \, \mathrm{km \, s^{-1}}$. Its ellipticity $e = (a^2 - b^2)/(a^2 + b^2)$ is fixed to $e^h = 0.7$, with a position angle of $\theta^h = 19.14°$ counterclockwise from west.

The clumpy subhalo component of the mock cluster is composed by 138 cluster galaxies, selected from the C-17 members catalog to have magnitudes in the HST $F160W$ filter given by $m_{F160W} < 22$. All these galaxies are parametrized as circular dPIE profiles whose r_{core}^{gal}, r_{cut}^{gal}, and σ_{LT}^{gal} values are determined as follows. For r_{core}^{gal} and r_{cut}^{gal} we adopt the scaling relations in Eq. (4) with slopes $\beta_{core} = 0.5$ and $\beta_{cut} = 0.66$. The two normalizations, computed at the BCG luminosity L_0 ($m_{F160W}^{ref} = 17.19$), are $r_{core}^{ref} = 0.01''$ and $r_{cut}^{ref} = 5''$. We assign line-of-sight stellar velocity dispersions to the cluster member galaxies, averaged within a circular aperture, assuming a 15% Gaussian scatter around the scaling relation in Eq. (4) with $\alpha = 0.27$ and normalization equal to $350 \, \mathrm{km \, s^{-1}}$ at the BCG luminosity. To determine the LENSTOOL fiducial velocity dispersions, we deproject the aperture-averaged velocity dispersions using Eq. (17) and assuming apertures of radius $R = 0.8''$. No shear nor foreground structures are present in our Cluster-1 mock.

Given the total mass distribution for the mock cluster, we use LENSTOOL to ray trace the position of 15 background sources, randomly selected from the C-17 catalog, to their multiple images on the lens plane. The sources are within the redshift range $1.01 \le z \le 6.06$ and produce a total of 85 multiple images. Images are identified using an integer number and a letter in such a way that images with identical numbers in their ID belong to the same family. We exclude 26 of the 85 multiple images from the final sample, either because they are de-magnified or because they are too close to a cluster member to be detectable, in real clusters, owing to light contamination. We also exclude 1 of the 15 families of multiple images (i.e., the tenth family adopting image IDs in Fig. 1) since it is constituted by a single lensed image. Therefore, our final mock multiple images catalog consists of 58 multiple images, from 14 background sources, shown in Fig. 1. The final simulated cluster, despite a purposely simple mass distribution, contains a number of halos, subhalos, and multiple images comparable to those of some CLASH or HFF clusters.

In our tests, we consider isotropic statistical errors of 0.2″ on the multiple-image positions. This is a compromise choice between the sharper accuracies that can be achieved

(in principle) with most updated imaging and the tolerance that the models need to efficiently explore the full parameter space, avoiding narrow likelihood peaks around local minima. Choosing a 0.2″ tolerance on the image positions also allows us to keep conservative uncertainties on the posteriors in the hyperparameters. In practice, we may apply these models in subsequent iterations where the tolerance is gradually reduced to match the astrometric uncertainties.

We also suppose that the stellar velocity dispersions of 58 luminous cluster galaxies are measured within apertures of radius $R = 0.8''$. This is comparable to the number of velocity dispersions that Bergamini et al. (2019) measured from the MUSE data cube of MACS J1206.2−0847. To associate an error with these simulated measures, we use the following empirical relation derived from the measurements by Bergamini et al. (2019):

$$\frac{\delta \sigma_m^{gal}}{\sigma_m^{gal}} = 1.6 \times 10^{-3} \, m_{F160W}^3 - 8.53 \times 10^{-2} \, m_{F160W}^2$$
$$+ 1.509 \, m_{F160W} - 8.879. \tag{19}$$

4.2. Cluster-2 and Cluster-3: harder rungs with low-mass substructures and systematics in the astrometry of cluster galaxies and multiple images

Similar to Cluster-1, the cluster-scale component of both Cluster-2 and Cluster-3 mocks is described through a single elliptical PIEMD profile. The PIEMD has an offset of about 1.41″ from cluster BCG and is characterized by an ellipticity $e^h = 0.72$, a position angle of $\theta^h = 19.8°$, a core radius $r_{core}^h = 11''$, and a LENSTOOL fiducial velocity dispersion $\sigma_{LT}^h = 1080 \, \mathrm{km \, s^{-1}}$. The galaxy-scale component of the new mocks extends over a wider magnitude range from the BCG luminosity ($m_{F160W}^{ref} = 17.19$) down to $m_{F160W} = 24$, and amounts to 258 galaxies extracted from the cluster member catalog by C-17. Cluster galaxies are modeled as circular dPIEs whose r_{core}^{gal}, r_{cut}^{gal}, and σ_{LT}^{gal} are determined as in Sect. 4.1, but assuming $\alpha = 0.28$, $\sigma^{ref} = 295.5 \, \mathrm{km \, s^{-1}}$, $\beta_{cut} = 0.64$, and $r_{cut}^{ref} = 3.74''$. As in Cluster-1, a 15% Gaussian scatter around the $\sigma^{gal} - m_{F160W}^{gal}$ scaling relation is assumed for the aperture averaged (line-of-sight projected) stellar velocity dispersions of the galaxies. With this choice of dPIE parameter values, the cluster member population in Cluster-2 and Cluster-3 ranges from a total mass of $1.6 \times 10^{12} \, M_\odot$, for the cluster BCG, down to a total mass of $1.1 \times 10^9 \, M_\odot$, for the less massive galaxy of the cluster. In the following, we assume to know the measured stellar velocity dispersion and associated errors of 58 cluster galaxies (see Sect. 4.1). These velocities are given as inputs to BAYESLENS for the lens model optimization.

What differentiates Cluster-2 from Cluster-3 is that in the latter model, we introduce, in addition to cluster galaxies, a large population of low-mass subhalos inside the simulated cluster. Thus, Cluster-3 is mainly designed to test the impact of numerous, undetected, substructures on BAYESLENS performances. To assign a mass to the faint subhalos, we exploit the subhalo mass function fitted by Giocoli et al. (2010) and adopted by the MOKA algorithm for simulating the gravitational lensing by galaxy clusters (see Giocoli et al. 2012). In using this formula, we assume a virial mass of $M_{vir} = 1.59 \times 10^{15} \, M_\odot$ for the host cluster halo. This value corresponds to the total mass of MACS J1206-0847 (C-17) contained within a distance from the cluster center of $r_{200c} = 2.06 \, \mathrm{Mpc}$ (362.79″ at $z = 0.439$) that is expected to be
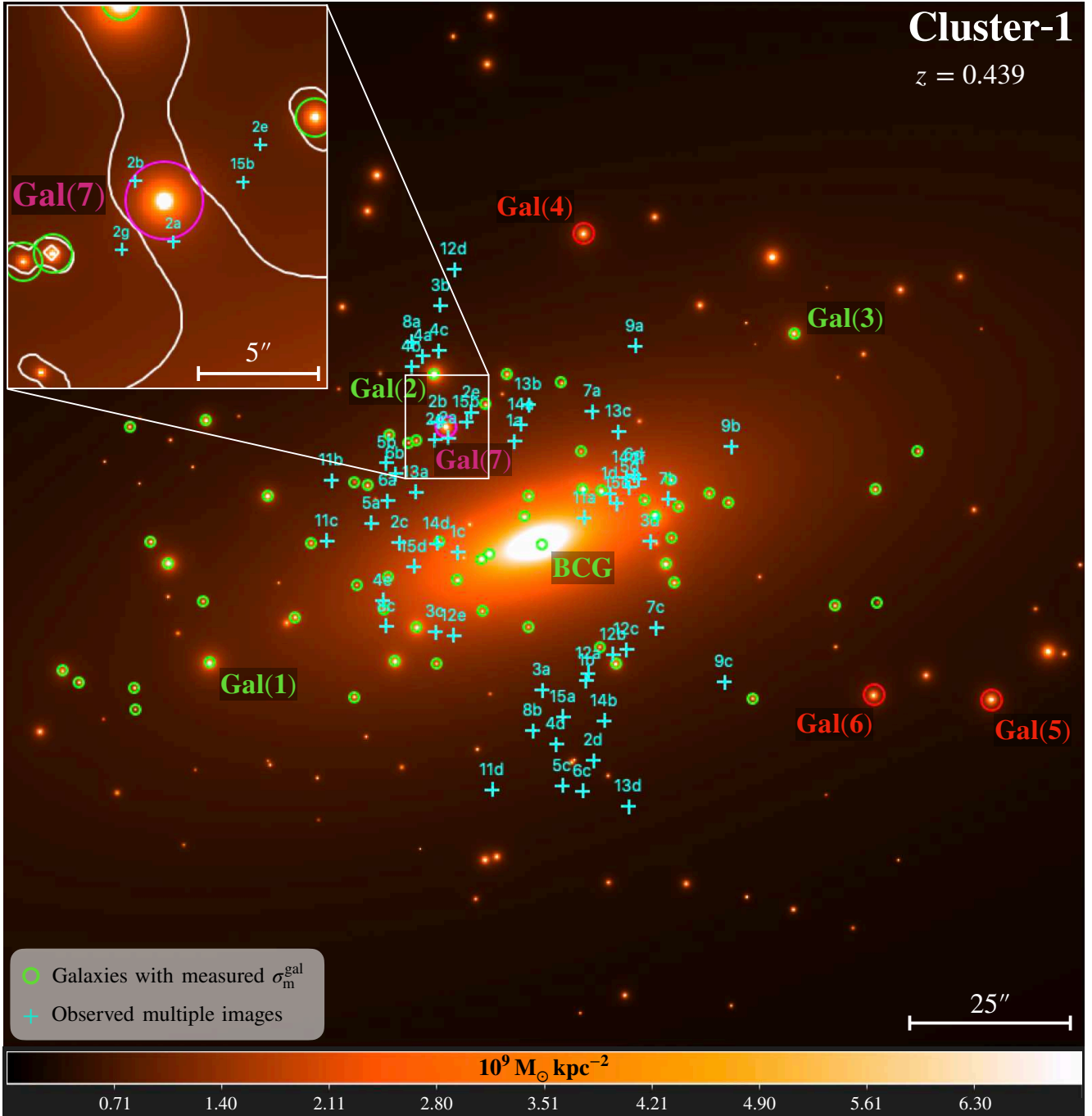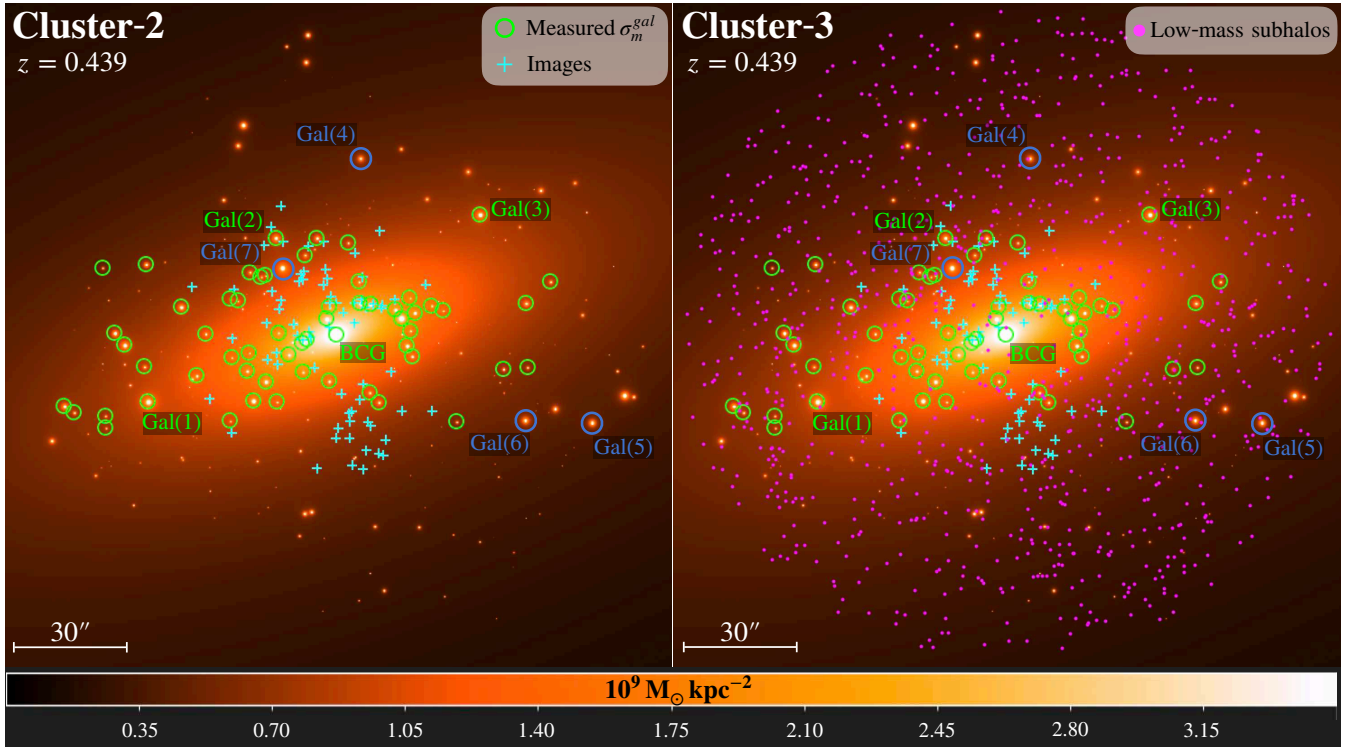
Fig. 1. Mass density distribution, color coded in $M_\odot\,\mathrm{kpc}^{-2}$, of the central region of the Cluster-1 mock used for our functional tests, loosely based on the Caminha et al. (2017b) lens model for the cluster MACS J1206.2−0847 at redshift $z = 0.439$. Green circles denote the cluster galaxies with measured velocity dispersion within apertures of radius $R = 0.8''$. The brightest galaxies with and without measured velocity dispersion are labeled in green and red, respectively. Cyan crosses denote the position of observed multiple images. In the inset, a galaxy-scale multiple-image system is shown around the cluster member Gal(7) (labeled in magenta). The critical lines computed for the redshift, $z_2 = 2.539$, of the second family of multiple images are shown in white.

close to the virial radius of the cluster (Umetsu et al. 2014). We note that a circularly symmetric PIEMD mass distribution with $r^h_{\mathrm{core}} = 11''$ and $\sigma^h_{\mathrm{LT}} = 1080\,\mathrm{km\,s^{-1}}$ has a mass, within a radius of $362.79''$, exactly equal to $1.59 \times 10^{15}\,M_\odot$.

Low-mass subhalos are spatially distributed according to the function derived by Gao et al. (2004) fitting the results of cosmological simulations. In particular, we adopt a concentration parameter $c_{200c} = r_{200c}/r_{\mathrm{s}} = 3.5$ for the host cluster halo and we assume $a = 1.944$ in Eq. (3) of Gao et al. (2004). Recent

work by Meneghetti et al. (2020) shows a detailed comparison between subhalos in current N-body and hydrodynamical simulations and observed subhalos from a sizable sample of cluster lens models.

In Cluster-3, we consider all the faint substructures within a distance of $89.65''$ from the cluster BCG; this is the distance of the most external galaxy considered in the model; with masses ranging between $1.56 \times 10^9\,M_\odot$, that is about the mass of the least massive cluster member, down to $0.93 \times 10^9\,M_\odot$. Under these

Fig. 2. Mass density distribution, color coded in $M_\odot\,\mathrm{kpc}^{-2}$, of the central region of the Cluster-2, and Cluster-3 mocks. As in Fig. 1, green circles denote the galaxies for which we have a measured velocity dispersion, while cyan crosses are the observed multiple images. The same set of measured and unmeasured galaxies selected in Fig. 1 are labeled in green and blue, respectively. *Right panel*: magenta data points are plotted showing the spatial distribution of the low-mass subhalo population of Cluster-3.

assumptions, the low-mass subhalo population of the Cluster-3 mock amounts to 910 profiles.

Each low-mass subhalo in the mock is parameterized as a circular dPIE, and to obtain a value for their σ_{LT} and r_{cut} parameters, we adopt the following procedure. Firstly, we compute the total mass for all the simulated cluster member galaxies ($M_{\mathrm{tot}}^{\mathrm{gal}}$). Then, we fit a $M_{\mathrm{tot}}^{\mathrm{gal}}-m_{F160W}^{\mathrm{gal}}$ relation between galaxy masses and magnitudes. This relation is adopted to associate a fictitious magnitude with each low-mass subhalo. Finally, we use scaling relations in Eq. (4) to derive velocity dispersions, core radii, and truncation radii for the simulated subhalos from their inferred magnitudes. As for cluster galaxies, we assume 15% Gaussian scatter around the $\sigma^{\mathrm{gal}}-m_{F160W}^{\mathrm{gal}}$ scaling relation for the low-mass subhalos as well.

Using Cluster-2 and Cluster-3 mocks, we map the position of 22 background sources, within the redshift range $1.012 \leq z \leq 5.793$ and randomly selected from the C-17 catalog, to their multiple images on the lens plane. Both Cluster-2 and Cluster-3 amount to 69 magnified multiple images, whose positions are used as constraints for the lens model optimizations. As in Cluster-1, we still assume an isotropic statistical error of $0.2''$ on multiple image positions, but an additional systematic uncertainty on the position of multiple images and cluster member galaxies is considered in the Cluster-2, and Cluster-3 mocks. To do so, we assign random Gaussian displacements to the true positions, with a dispersion of $0.01''$ in both directions. These are comparable to the residual uncertainties from the HST-like quality of imaging data.

In Fig. 2, we plot the total projected mass distribution of Cluster-2 and Cluster-3. Cluster member galaxies with measured velocity dispersions are encircled in green, while we indicate the

position of the observed multiple images using cyan crosses. The spatial distribution of the low-mass subhalo population included in the Cluster-3 mock is shown using magenta data points on the right panel of the figure.

In Table 1, we report a summary of the main proprieties of Cluster-1,2,3 mocks, such as the number of galaxies with and without a measured velocity dispersion ($N_{\mathrm{m}}^{\mathrm{gal}}$ and N^{gal}), the number of low-mass subhalos (N^{DH}), the statistical and systematic errors on galaxy and multiple images positions ($(\Delta x^{\mathrm{gal}})_{\mathrm{sys}}$, $(\Delta x^{\mathrm{im}})_{\mathrm{st}}$, $(\Delta x^{\mathrm{im}})_{\mathrm{sys}}$), and the number of observed multiple images ($N_{\mathrm{tot}}^{\mathrm{im}}$).

4.3. BAYESLENS on the mock clusters

In this subsection, we describe the parameter ranges adopted in BAYESLENS to probe the mass distribution of the Cluster-1, Cluster-2, and Cluster-3 mocks. The cluster-scale component of all three lens models is parametrized using a single PIEMD profile. The x, y coordinates of its center can vary within flat priors, $3''$ wide, centered on the BCG position, while for the ellipticity $\hat{e}^h$, position angle $\hat{\theta}^h$, and fiducial velocity dispersion $\hat{\sigma}_{\mathrm{LT}}^h$, we adopt uniform priors inside the following intervals, respectively: $0.0-0.9$, $5-35°$, and $700-1300\,\mathrm{km\,s}^{-1}$. Finally, for the PIEMD core radius $\hat{r}_{\mathrm{core}}^h$, we assume a uniform prior between $1''$ and $7''$ in the Cluster-1 lens model and between $3''$ and $15''$ for Cluster-2 and Cluster-3. The subhalo component of lens models is parametrized using the three scaling relations in Eq. (4) normalized at the BCG luminosity. In particular, we fix $\hat{r}_{\mathrm{core}}^{\mathrm{ref}} = 0.01''$ and $\hat{\beta}_{\mathrm{core}} = 0.5$, while the following flat priors are assumed on the others scaling relation parameters: Cluster-1 ($0.20 \leq \hat{\alpha} \leq 0.34$, $250\,\mathrm{km\,s}^{-1} \leq \hat{\sigma}^{\mathrm{ref}} \leq 450\,\mathrm{km\,s}^{-1}$, $1'' \leq \hat{r}_{\mathrm{cut}}^{\mathrm{ref}} \leq 11''$), Cluster-2, and Cluster-3 ($0.21 \leq \hat{\alpha} \leq 0.35$, $150\,\mathrm{km\,s}^{-1} \leq \hat{\sigma}^{\mathrm{ref}} \leq 350\,\mathrm{km\,s}^{-1}$,

Table 1. Main features of the Cluster-1, Cluster-2, and Cluster-3 mocks.

	Cluster proprieties			
	Cluster-1	Cluster-2	Cluster-3	
N_m^{gal}	58	58	58	SUBHALOS
N^{gal}(Out SR)	80 (80)	200 (22)	200 (22)	
N^{DH}	0	0	910	
$(\Delta x^{gal})_{sys}$ [″]	0.0	0.01	0.01	
N_{tot}^{im}	58	69	69	IMAGES
$(\Delta x^{im})_{st}$ [″]	0.2	0.2	0.2	
$(\Delta x^{im})_{sys}$ [″]	0.0	0.01	0.01	
χ_{tot}^2	8.66	12.53	13.75	RES.
Δ_{rms}^{tot} [″]	0.08	0.09	0.09	

Notes. The number of galaxies with and without a measured velocity dispersion is given by N_m^{gal} and N^{gal}, respectively. For the latter, the number of galaxies, in BAYESLENS, that in principle are free to scatter around the best-fit σ-mag scaling relation is quoted in brackets. The systematic uncertainty assumed on galaxy positions is $(\Delta x^{gal})_{sys}$. The quantity N_{tot}^{im} is the total number of observed multiple images, while $(\Delta x^{im})_{st}$ and $(\Delta x^{im})_{sys}$ are the statistical and systematic isotropic errors assumed on their positions. In the last two lines, the total chi-square and the rms displacement between the measured and model-predicted image positions are listed.

$1'' \leq \hat{r}_{cut}^{ref} \leq 9''$). We note that the slope of the r_{cut}−mag scaling relation, $\hat{\beta}_{cut}$, is not a lens model free-parameter since its value is derived from $\hat{\alpha}$ through Eq. (5) with $\gamma = 0.2$. In contrast to LENSTOOL, in our code, the σ-mag cluster member scaling relation refers to measured (line-of-sight) stellar velocity dispersions averaged within apertures of radius $R = 0.8''$.

The BAYESLENS model optimization is performed on the lens plane using the positions of the multiple images inside the simulated catalog and the 58 measured stellar velocity dispersions of the cluster galaxies. These velocities are used, on the one hand, to determine the best σ-mag scaling relation parameters, and on the other, to derive a Gaussian prior for each measured galaxy. Thanks to these priors, the lens model tends to prefer solutions with cluster members velocities that coincide with the measured values (see Eq. (12)), unless the lensing data require them to deviate. The presence of a non-zero scatter also enables the models to fairly sample the whole parameter space, thereby avoiding the underestimation of systematics.

In running BAYESLENS on the Cluster-1 mock, we leave all galaxies free to scatter around the fitted σ-mag scaling relation, while in Cluster-2 and Cluster-3 lens models, only the 58 galaxies with a measured velocity dispersion, and the 22 most luminous unmeasured galaxies are left free to vary. The number of free-to-scatter galaxies is one of the inputs of our code, and in general, a lower number of free galaxies sensibly speed up the optimization of the lens model.

The total posterior probability distribution (p_{tot}) of the Cluster-1 lens model is sampled using 298 walkers, while we use 500 walkers for the other two models. The initial walker positions are randomly distributed within the flat priors of the cluster-scale parameters (ϕ^h), but they are initialized inside a Gaussian hyper-sphere, around an estimated maximum of the likelihood (in p_{sr}, p_{mg}, p_g), in the subspace of the other free-parameters. In the following, we derive the marginalized posterior distribution of model free-parameters by flattening the final Monte Carlo Markov chain (MCMC) of the walkers after remov-

ing a sufficiently large burn-in phase. Moreover, to determine the model-predicted multiple image positions and the rms displacement, Δ_{rms}^{tot}, we use the best-fit BAYESLENS model. This best-fit model is defined as the set of free-parameter values that maximize the total posterior p_{tot}.

5. Results and discussion

In this section, we describe the main results obtained from an application of BAYESLENS on the three simulated clusters, namely Cluster-1, Cluster-2, and Cluster-3.

5.1. BAYESLENS results for Cluster-1

In Fig. 3, we show in black the marginalized BAYESLENS posterior distributions for the free-parameters of the Cluster-1 lens model. The true parameter values assumed in the mock cluster are denoted with solid red lines. Panels a and b show the posterior distributions of the scaling relation and cluster-scale halo parameters, respectively. In panel c, we plot the marginalized posterior distributions for the stellar velocity dispersions of the four brightest measured cluster galaxies: $\hat{\sigma}_m^{BCG}$, $\hat{\sigma}_m^{Gal(1)}$, $\hat{\sigma}_m^{Gal(2)}$ and $\hat{\sigma}_m^{Gal(3)}$. Similarly, panel d shows the velocity dispersions of the three brightest cluster galaxies without measured kinematics ($\hat{\sigma}^{Gal(4)}$, $\hat{\sigma}^{Gal(5)}$ and $\hat{\sigma}^{Gal(6)}$) and of the lens galaxy, $\hat{\sigma}^{Gal(7)}$, forming the galaxy scale strong lensing system zoomed in Fig. 1. In these last two panels, we also plot blue solid lines showing the value of galaxy stellar velocity dispersions, as predicted by the best-fit σ^{gal}−m_{F160W}^{gal} scaling relation. The true and BAYESLENS optimized values of the lens model free-parameters are reported in Table 2. The latter correspond to the medians of parameter marginalized posterior distributions, while the 16th and 84th percentiles are quoted as an error. Thanks to the measured cluster members stellar velocity dispersions and to the observed multiple-image positions, BAYESLENS recovers, well within the 1σ uncertainty, all of the mock true hyperparameters of the scaling relations and the cluster-scale halo parameters. To quantify the accuracy of BAYESLENS in determining the correct stellar velocity dispersions of the cluster members, we plot in Fig. 4 the aperture average velocity dispersions within aperture of $0.8''$radius, σ^{gal}, as a function of the galaxy magnitudes, m_{F160W}. All of the galaxies lie on their measured velocity dispersions, and the zero-scatter solution still exists within the sampled parameter space. This shows that the loose prior on measured galaxies still allows the models to generalize the zero-scatter ansatz. The bottom panel shows the mock and posterior dispersions for the unmeasured galaxies. Most of these are compatible with their true values (red dots) within the uncertainties. Most notably, one galaxy-scale strong lensing system (denoted as Gal(7), see Fig. 1 closeup) automatically departs from the zero-scatter, best-fit scaling relation and has a dispersion that is very well constrained by the lensing data. The compatibility of mock and posterior dispersions for all galaxies (within the intrinsic scatter) is represented in Fig. 5, showing that there is no appreciable systematic mismatch trend between the true and inferred dispersions.

Figure 6 shows the cumulative mass profile of the whole cluster, projected along the line of sight, for the mock and the best-fit (maximum a posteriori) model. Even though the total mass of the mock cluster is redistributed in a slightly different way between the cluster halo and the member galaxies, its value at large radii is accurately recovered by the lens model. At the distances where most of the multiple images form, yielding more
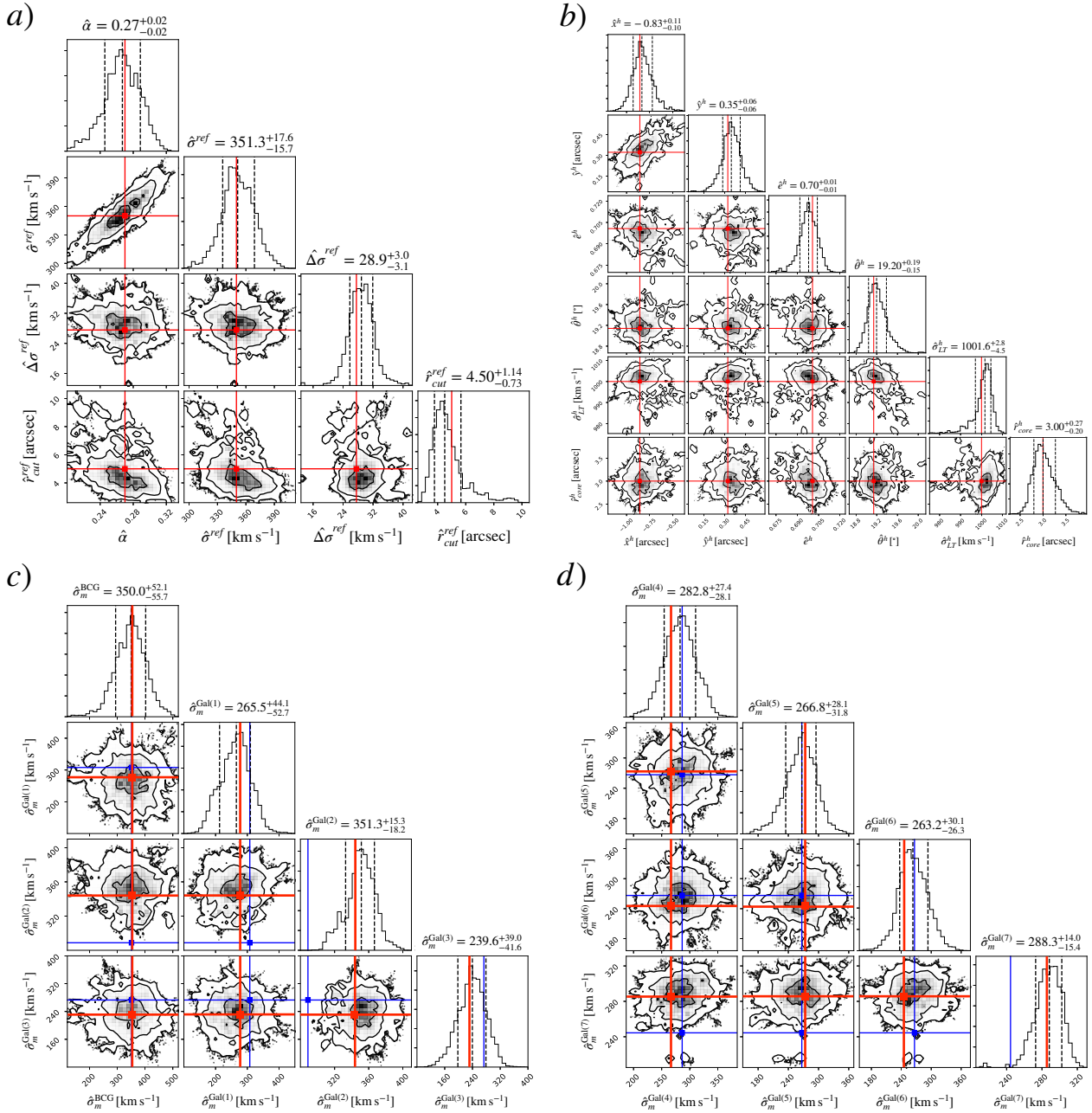
Fig. 3. BAYESLENS marginalized posterior distributions on the free parameters of the Cluster-1 mocks. The contours limit the 1σ, 2σ, and 3σ regions. The black dashed vertical lines in the histograms correspond to the 16th, 50th, and 84th percentiles of the marginalized distributions (these values are reported in the titles). The solid red lines are the true values of the parameters of the mock cluster. *Panel a*: scaling relation hyperparameters. In this plot, $\hat{\beta}_{cut}$ is omitted because its value is not optimized in BAYESLENS, but is directly derived from Eq. (5) assuming $\gamma = 2$. *Panel b*: posterior distributions for the cluster-scale halo parameters, while *panel c*: average velocity dispersions within apertures of radius $R = 0.8''$ of the four brightest measured galaxies (BCG, Gal(1), Gal(2), and Gal(3)). *Panel d*: average velocity dispersions within apertures of radius $R = 0.8''$ of the three brightest galaxies without a measured σ: Gal(4), Gal(5), and Gal(6). In the plot, the galaxy Gal(7) forming a galaxy-scale strong lensing system (see text) is included. *Panels c and d*: blue vertical lines indicate the stellar, aperture-averaged, velocity dispersion of cluster members predicted by the best-fit $\sigma^{gal} - m_{F160W}$ scaling relation (zero-scatter solution).

constraints to the lens model, the model predicts a correct value for the mass profiles within 1%. This is simply because the projected mass $M(< R_E) = \pi \Sigma_{cr} R_E^2$ within the Einstein radius of a lens does not depend on its density profile.

To relate the goodness of fit of the models to a commonly used benchmark, we examine the rms offset of observed versus inferred positions of the multiple images (Δ_{rms}^{tot}). Our best-fit hierarchical model for Cluster-1 can reproduce all images with

$\Delta_{rms}^{tot} = 0.08''$, while the lens model total chi-square is $\chi_{tot}^2 = 8.66$ (see Table 1). This non-zero Δ_{rms}^{tot} shows that our hierarchical model does not over-fit while also reproducing all observables well within their uncertainties.

The robustness against over-fitting can be understood in two ways. First, if we rerun the models on a mock with $0.05''$ statistical uncertainties on the image positions, the maximum a posteriori parameters do not change, and the rms offset is

Table 2. Comparison between the true input values of the Cluster-1, Cluster-2, and Cluster-3 mocks and the BAYESLENS output results.

	Scaling relation hyperparameters ($m_{F160W}^{ref} = 17.19$)			
	$\hat{\alpha}$	$\hat{\sigma}^{ref}$ [km s^{-1}]	$\hat{\Delta\sigma}^{ref}$ [km s^{-1}]	$\hat{r}_{core}^{ref}$ [arcsec]
True (Cluster-1)	0.27	350.0	27.6	5.00
BL (Cluster-1)	$0.27^{+0.02}_{-0.02}$	$351.3^{+17.6}_{-15.7}$	$28.9^{+3.0}_{-3.1}$	$4.50^{+1.14}_{-0.73}$
True (Cluster-2,3)	0.28	295.5	22.7	3.74
BL (Cluster-2)	$0.28^{+0.02}_{-0.02}$	$289.9^{+11.5}_{-11.2}$	$19.3^{+2.2}_{-2.1}$	$3.91^{+0.78}_{-0.61}$
BL (Cluster-3)	$0.28^{+0.02}_{-0.02}$	$290.1^{+10.8}_{-12.0}$	$19.3^{+2.6}_{-2.1}$	$3.93^{+0.77}_{-0.64}$

	Halo parameters					
	$\hat{x}^h$ [arcsec]	$\hat{y}^h$ [arcsec]	$\hat{e}^h$	$\hat{\theta}^h$ [°]	$\hat{\sigma}_{LT}^h$ [km s^{-1}]	$\hat{r}_{core}^h$ [arcsec]
True (Cluster-1)	-0.86	0.32	0.70	19.14	1000.0	3.00
BL (Cluster-1)	$-0.83^{+0.11}_{-0.10}$	$0.35^{+0.06}_{-0.06}$	$0.70^{+0.01}_{-0.01}$	$19.20^{+0.19}_{-0.15}$	$1001.6^{+2.8}_{-4.5}$	$3.00^{+0.27}_{-0.20}$
True (Cluster-2,3)	-1.40	0.14	0.72	19.80	1080.0	11.00
BL (Cluster-2)	$-1.33^{+0.16}_{-0.16}$	$0.16^{+0.07}_{-0.07}$	$0.72^{+0.004}_{-0.004}$	$19.77^{+0.12}_{-0.11}$	$1079.0^{+2.5}_{-3.0}$	$11.00^{+0.18}_{-0.17}$
BL (Cluster-3)	$-1.34^{+0.17}_{-0.16}$	$0.17^{+0.08}_{-0.07}$	$0.72^{+0.004}_{-0.005}$	$19.76^{+0.11}_{-0.12}$	$1079.3^{+2.7}_{-2.8}$	$11.02^{+0.17}_{-0.17}$

	Galaxies with measured σ [km s^{-1}]			
	$\hat{\sigma}_m^{BCG}$ ($m_{F160W} = 17.19$)	$\hat{\sigma}_m^{Gal(1)}$ ($m_{F160W} = 17.72$)	$\hat{\sigma}_m^{Gal(2)}$ ($m_{F160W} = 17.98$)	$\hat{\sigma}_m^{Gal(3)}$ ($m_{F160W} = 18.21$)
True (Cluster-1)	354.0 ± 12.9	278.2 ± 10.4	344.1 ± 12.9	230.8 ± 8.8
BL (Cluster-1)	$350.0^{+52.1}_{-55.7}$	$265.5^{+44.1}_{-52.7}$	$351.3^{+15.3}_{-18.2}$	$239.6^{+39.0}_{-41.6}$
True (Cluster-2,3)	292.9 ± 10.6	261.2 ± 9.7	202.0 ± 7.6	233.8 ± 8.9
BL (Cluster-2)	$290.5^{+15.6}_{-15.8}$	$252.3^{+44.5}_{-42.8}$	$208.3^{+16.8}_{-16.4}$	$233.1^{+41.6}_{-43.1}$
BL (Cluster-3)	$292.1^{+15.2}_{-16.4}$	$256.8^{+46.2}_{-45.6}$	$209.4^{+16.8}_{-17.1}$	$234.9^{+41.2}_{-42.4}$

|	Galaxies without measured σ [km s^{-1}]			
	$\hat{\sigma}^{Gal(4)}$ ($m_{F160W} = 18.01$)	$\hat{\sigma}^{Gal(5)}$ ($m_{F160W} = 18.30$)	$\hat{\sigma}^{Gal(6)}$ ($m_{F160W} = 18.31$)	$\hat{\sigma}^{Gal(7)}$ ($m_{F160W} = 18.66$)
True (Cluster-1)	266.6	274.1	244.9	285.5
BL (Cluster-1)	$282.8^{+27.4}_{-28.1}$	$266.8^{+28.1}_{-31.8}$	$263.2^{+30.1}_{-26.3}$	$288.3^{+14.0}_{-15.4}$
True (Cluster-2,3)	174.3	217.6	210.4	256.8
BL (Cluster-2)	$231.6^{+20.9}_{-20.6}$	$218.9^{+20.3}_{-20.8}$	$218.4^{+19.0}_{-19.7}$	$216.1^{+20.4}_{-20.3}$
BL (Cluster-3)	$233.3^{+19.4}_{-20.5}$	$217.8^{+20.2}_{-19.9}$	$217.1^{+20.8}_{-19.9}$	$225.6^{+16.6}_{-16.3}$

Notes. The 50th, 16th, and 84th percentiles computed from the marginalized posteriors are quoted. The first table refers to scaling relations hyperparameters, including the scatter of cluster galaxies around the inferred σ-mag relation. In the second table, the cluster-scale halo parameters are shown. The third table shows the results for the four brightest galaxies with a measured velocity dispersion: BCG, Gal(1), Gal(2), and Gal(3). Finally, the last table shows the results for the galaxy Gal(7) (see text) and for the three brightest galaxies without a measured velocity dispersion, i.e., Gal(4), Gal(5), and Gal(6). The galaxy magnitudes are shown within round brackets.

still $\approx 0.08''$, while only the confidence intervals in the inferred parameters are shrunk. This is because all image positions in this first mock are kept on their true locations, and the statistical uncertainties are simply a tolerance parameter that enables a smooth exploration of parameter space. Second, from the posterior we see that only about ten galaxies deviate by more than 1σ from the backbone of the scaling relation; if we rerun a LENSTOOL model with only these galaxies free, and all other galaxies on the zero-scatter scaling relation, the χ^2 improves by ≈ 33, that is three times the change in degrees of freedom.

5.2. BAYESLENS results for Cluster-2 and Cluster-3

Table 2 gives a comparison between the true values of the Cluster-2 and Cluster-3 parameters with BAYESLENS results. We note this table shows the same set of free-parameters for the three mocks, however, their true values are different in the Cluster-1 and Cluster-2,3 mocks.
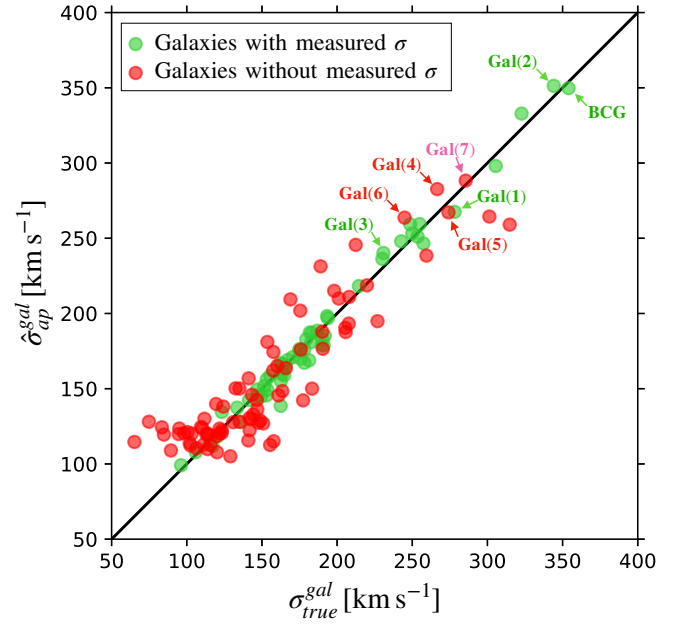
Despite the increasing complexity of the last two mock clusters, BAYESLENS recovers well within 1σ all cluster-scale halo and scaling relations parameters. Moreover, BAYESLENS finds
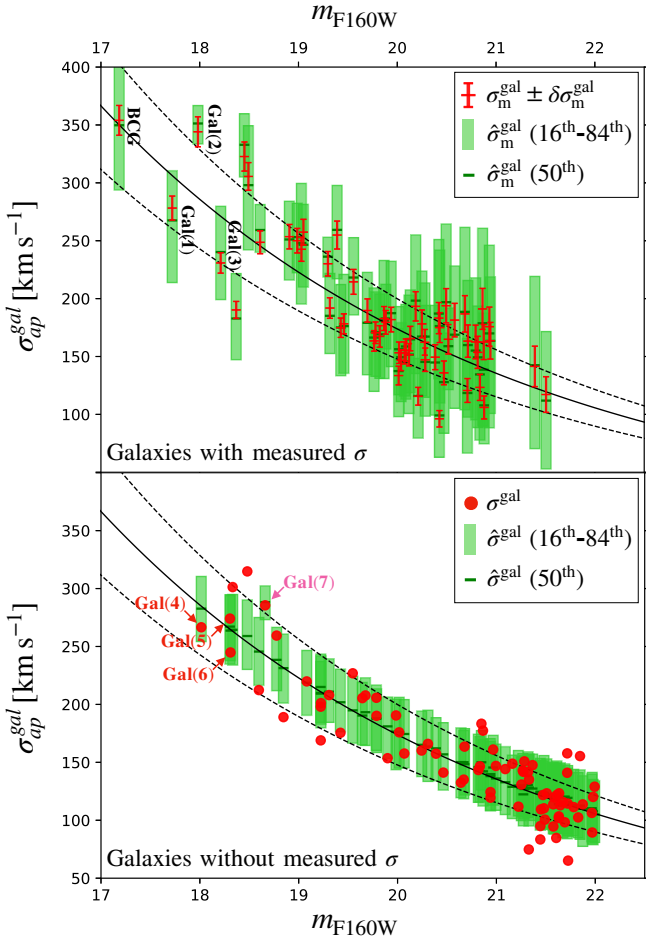
Fig. 4. Velocity dispersions of the Cluster-1 member galaxies as a function of their magnitudes. In the mock cluster, a 15% of Gaussian scatter (black dashed lines) is assumed for the galaxy velocity dispersion around the scaling relation plotted as a solid black line. *Top panel*: measured velocity dispersions with their errors are plotted in red. Similarly, the true velocity dispersions of the galaxies without a measured velocity dispersion are denoted with red dots in the *bottom panel*. The green rectangles in the plots are bounded by the 16th and 84th percentiles of the marginalized posterior distribution for each galaxy, while the 50th percentiles are the small green bars. The most luminous galaxies with and without measured velocity dispersion are labeled in black and red, respectively. The cluster member *Gal*(7) forming the galaxy-scale strong lensing system shown in the cut-out of Fig. 1 is labeled in magenta.

Fig. 5. Velocity dispersions of cluster member galaxies from the BAYESLENS lensing model of Cluster-1 (*y*-axis) vs. their true values in the mock (*x*-axis). The most luminous galaxies with and without a measured velocity dispersion are labeled in green and red, respectively. In magenta the cluster member *Gal*(7) responsible for the creation of the galaxy-scale strong lensing system shown in the cut-out of Fig. 1.
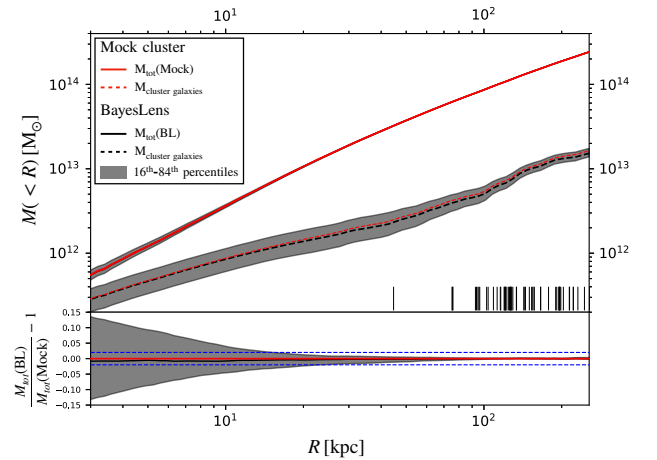
Fig. 6. *Top*: cumulative projected total mass profiles for the Cluster-1 mock as a function of the projected distance from the BCG. The red line corresponds to the true mass profile. The solid black line is obtained from the best-fit BAYESLENS model. The projected distances of the multiple images from the BCG center are shown as vertical black lines. *Bottom*: relative variation of the BAYESLENS total mass profile with respect to the true mass profile of the mock cluster. The dashed blue line in the *bottom panel* indicates a 2% relative difference between the mock and model mass profiles. In both panels, the gray shaded stripes delimit the 16th–84th percentiles of the mass profiles with parameters drawn from the posterior.

similar values for the stellar, aperture-average, velocity dispersions of measured, and unmeasured cluster galaxies in the two mocks.

In Fig. 7, we plot marginalized posterior distributions, obtained from the Cluster-3 MCMC chains, for the same set of model parameters reported in Table 2. The figure shows that the four measured galaxies BCG, Gal(1), Gal(2), Gal(3), and the two unmeasured galaxies Gal(5), Gal(6) have marginalized posterior distributions that pick very close to the true mock stellar velocity dispersions. Conversely, we observed large deviations from true values for Gal(7) and even more for Gal(4). The reason for these deviations is at the basis of the hierarchical approach implemented in BAYESLENS, and it is due to the interplay between the p_g and p_{im} terms of the total posterior in Eq. (7). The unmeasured galaxy Gal(4) has marginalized posterior distribution centered on the stellar velocity dispersion value predicted by the

best-fit $\sigma^{gal}-m_{F160W}$ scaling relation (denoted with a blue line in Fig. 7). In the absence of constraints from the position of the observed multiple images, the p_g term encourages a zero-scatter lens model solution with all unmeasured galaxy velocity dispersions that lie on the best-fit scaling relation. Similarly, the displacement of the $\hat{\sigma}^{Gal(7)}$ posterior distribution from the scaling relation toward the true mock value is possible because the

increase of the $p_{\rm im}$ term (owing to a lower rms displacement of the images close to Gal(7)) overcome the decrease of $p_{\rm g}$ that penalizes scattered solutions.

The best-fit Cluster-2 lens model reproduces the position of multiple images with a $\Delta_{\rm rms}^{\rm tot} = 0.09''$, while the total chi-square is $\chi_{\rm tot}^2 = 12.53$. As shown in Table 1, the addition of the low-mass subhalo population in the Cluster-3 mock produces a small increase, equal to 1.22, of the lens model total chi-square, however, the final $\Delta_{\rm rms}^{\rm tot}$ of multiple images remains essentially unchanged.

5.3. Testing against main halo mismatch

Different mass models, related through unobservable source-position transformations, may be able to reproduce the same image positions, even though these degeneracies are more severe for systems with less symmetry (Schneider & Sluse 2014; Unruh et al. 2017). This occurrence is common to all parametric lensing codes. In this section, we check whether BAYESLENS uses the non-zero scatter to over-fit models that are different from the mocks.

To test whether this happens, we also considered one last simulation whose main halo has a Navarro-Frenk-White (hereafter NFW) density profile, while the BAYESLENS model still adopt a PIEMD to parameterize the main cluster halo. The cluster-scale NFW halo has a total mass $M_{200c} = 1.59 \times 10^{15}\, M_\odot$ within a radius $r_{200c} = 2.06\,{\rm Mpc}$. These values correspond to those measured by Umetsu et al. (2014), using weak lensing techniques, for the cluster MACS J1206.2−0847. Assuming a concentration $c_{200c} = 3.5$, we derive a scale radius $r_{\rm s} = 588.6\,{\rm kpc}$. The center position of the NFW profile has an offset of $1.4''$ from the cluster BCG, an ellipticity of 0.4, and a position angle of $19.8°$ counterclockwise from west. The subhalo component of the cluster and the position of the lensed sources are assumed the same as the Cluster-2 simulation. The lens model for this cluster is constrained by 69 observed multiple images, and systematic isotropic uncertainties of $0.01''$ are assumed on galaxies and image positions (see Cluster-2 model for more details).

The rms displacement of the best-fit BAYESLENS model is $0.40''$, that is, about $0.32''$ higher than the $\approx 0.08''$ of a PIEMD model on a PIEMD mock. Some of the observed multiple images are not produced by the best-fit model, whereas the PIEMD models on PIEMD mocks (Cluster-1,2,3) were all able to produce the whole set of images. We rerun BAYESLENS while also penalizing models for each multiple image that cannot be produced by the model[2]. In this case, the best-fit rms displacement is just $0.35''$, and all the images are reproduced well by the model.

The slightly higher rms displacement, as well as the new issues in producing all multiple images, indicate that a main-halo mismatch may be diagnosed when modeling real clusters. BAYESLENS does not seem to over-fit the most evident systematics, despite higher freedom in the cluster member galaxies.

6. Conclusions

We have shown that, despite the (nominally) large number of degrees of freedom in cluster lensing, the exploration of flexible models is feasible thanks to tractable, hierarchical Bayesian inference. Unlike conventional models, where galaxies are placed on a razor-thin scaling relation, and some are

---

[2] We add -100 to the log-likelihood for each multiple image that is not produced by the lens model.

freed ad hoc, we populate a scaling relation with non-zero scatter, with hyperparameters that are inferred directly from the lensing constraints and (when available) stellar kinematic data. Our tests on a realistic, albeit simplified, mock cluster (namely Cluster-1) show that the improvement over conventional models may be significant: the velocity dispersions of all the galaxies of our mock cluster are reliably recovered, as are the scaling relation hyperparameters, and the rms displacement of multiple images between measurements and model predictions can decrease appreciably (in the simplest mocks, from $0.40''$ to $0.08''$) unless there are other major sources of systematics. As shown by the tests in Sect. 5.3, the impact of the non-zero scatter is mainly in the deflections from the cluster member galaxies and cannot correct for a mismatch between the true density profile of the main cluster halo and that used in the models. This also has some practical implications when modeling real-life clusters: if the rms displacements do not change appreciably from zero-scatter to hierarchical models, this would indicate that other major sources of systematics must be analyzed (e.g., the density profile of the main halo or major mass contributions along the line of sight). We also remark that these models do not make any claim on the internal mass content of the galaxies themselves, but just on the fact that they do not necessarily have simple PIEMD profiles, that there is a real and measured scatter in their properties, and that the information from lensing is complementary to that from the measured stellar kinematics (i.e., the hyperparameters are free to differ from those from stellar kinematics).

We have also tested the robustness of the hierarchical models against two systematic effects: dark subhalos and small systematic offsets in the measured image and cluster member positions (Cluster-2 and Cluster-3 mocks). Within realistic regimes, compatible with current (HST) or upcoming (ELT) depth and image quality, the model performance on the mock cluster does not degrade appreciably.

A crucial consequence of our hierarchical approach is that it also resolves a false dichotomy between well-fitting cluster models and realistic population properties of galaxies. Zero-scatter models around a kinematic prior have realistic galaxy populations but a higher rms offset than models without any kinematic prior, whose scaling relations are significantly discrepant from the true input relation. In zero-scatter models, there is typically a trade-off between the rms and recovering realistic galaxy populations (see, e.g., Bergamini et al. 2019). Our hierarchical models have realistic galaxy populations and produce a small rms offset (on our mock cluster). With respect to zero-scatter models, BAYESLENS automatically identifies the galaxy-scale systems where more freedom from the scaling relation backbone is needed to reproduce the positions of multiple images around those galaxies with high accuracy. This makes the hierarchical models robust against over-fitting since in our mocks only about ten galaxies deviate from the scaling relation backbone by more than 1σ while the chi-square improves by $\Delta\chi^2 \approx 33$.

We should emphasize that this is a functional test on mocks, and additional effects may play a role in real-life systems, such as massive substructure, deviations from simple geometry of the main DM halo(s) and cluster members, and additional contributions along the line of sight. However, our hierarchical models can eliminate part of the systematic source of uncertainties and show that internal systematics can be controlled.

The freedom in the parameters of each cluster member galaxy, within the intrinsic scatter of their parent population, has multiple implications. First, the data themselves dictate which galaxy should be freed and deviate significantly from a baseline scaling relation. Second, intrinsic scatter is a hyperparameter
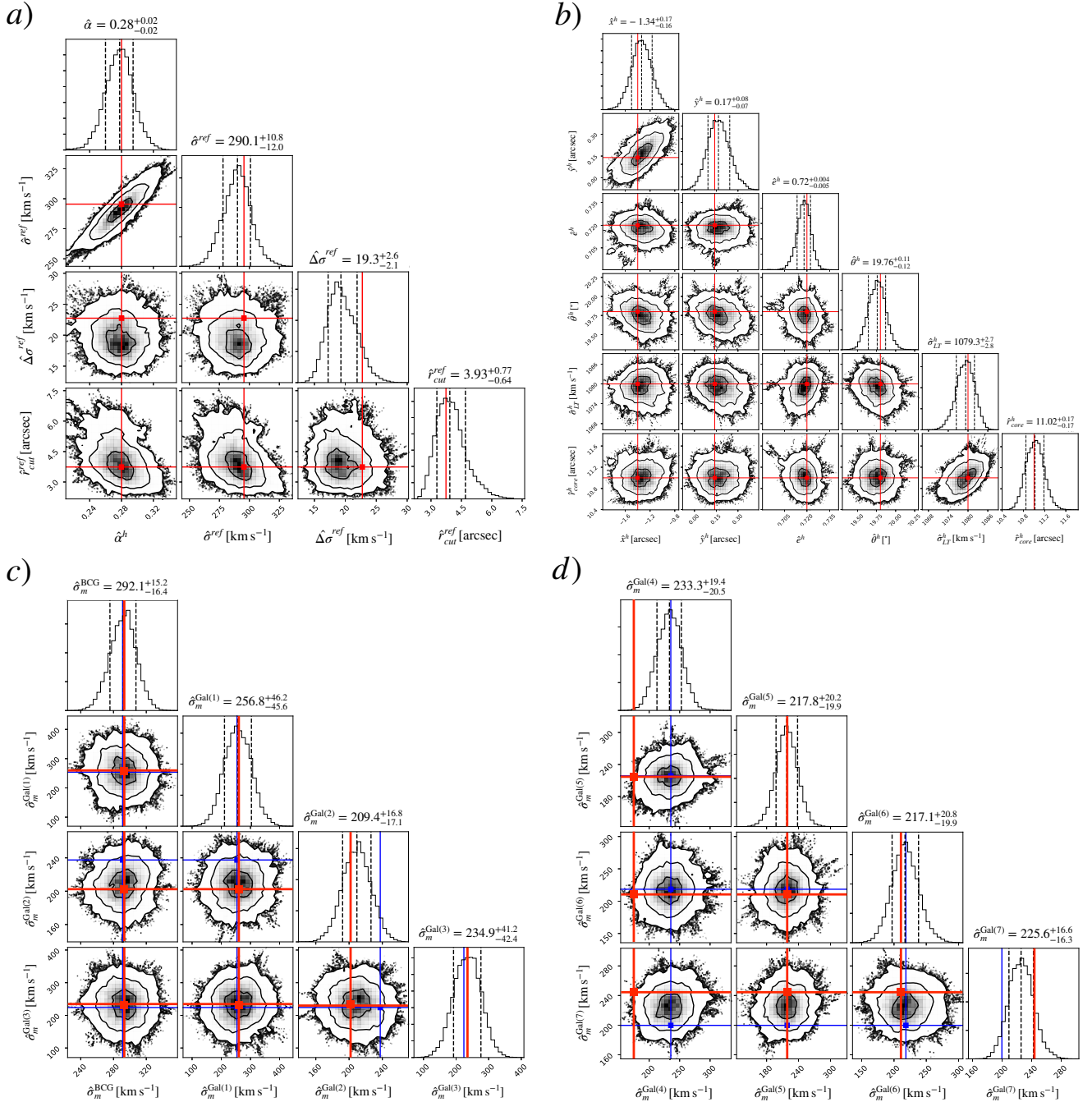
Fig. 7. Same as Fig. 3 but for the Cluster-3 mock.

that is left free in the modeling inference, and this allows for a direct determination of galaxy-population properties, whence accurate studies of galaxy formation and evolution. In particular, accurate mass functions of cluster member galaxies can be compared to predictions from cosmological simulations. Also, the freedom in cluster member parameters allows (in principle) for the quantification of environmental effects on galaxies at various locations across the cluster, which up to now has been possible only on stacked weak-lensing measurements (Niemiec et al. 2017).

An important feature of our inference is that BAYESLENS is fully modular. By changing specific Python functions within our code, in theory it is possible to allow for calls to any chosen parametric lensing software (besides LENSTOOL, used here as

the benchmark) and to implement different mass profiles. Moreover, different prescriptions to relate the stellar velocity dispersions and lensing parameters of cluster member galaxies can be used. The currently used scaling relations may be easily replaced with fundamental-plane relations (with free hyperparameters) so as to study the evolution of the fundamental plane across redshift and the environment. Its modular structure also enables the use of additional constraints from, for example, flux ratios or extended-source reconstruction. Samples of magnification maps may be produced simply through an other module, which may enable accurate studies of high-redshift galaxy populations.

This is not central to this first paper, so for the sake of brevity, in Fig. 8, we show simply a comparison of magnification maps from the best-fit models of the three mock clusters
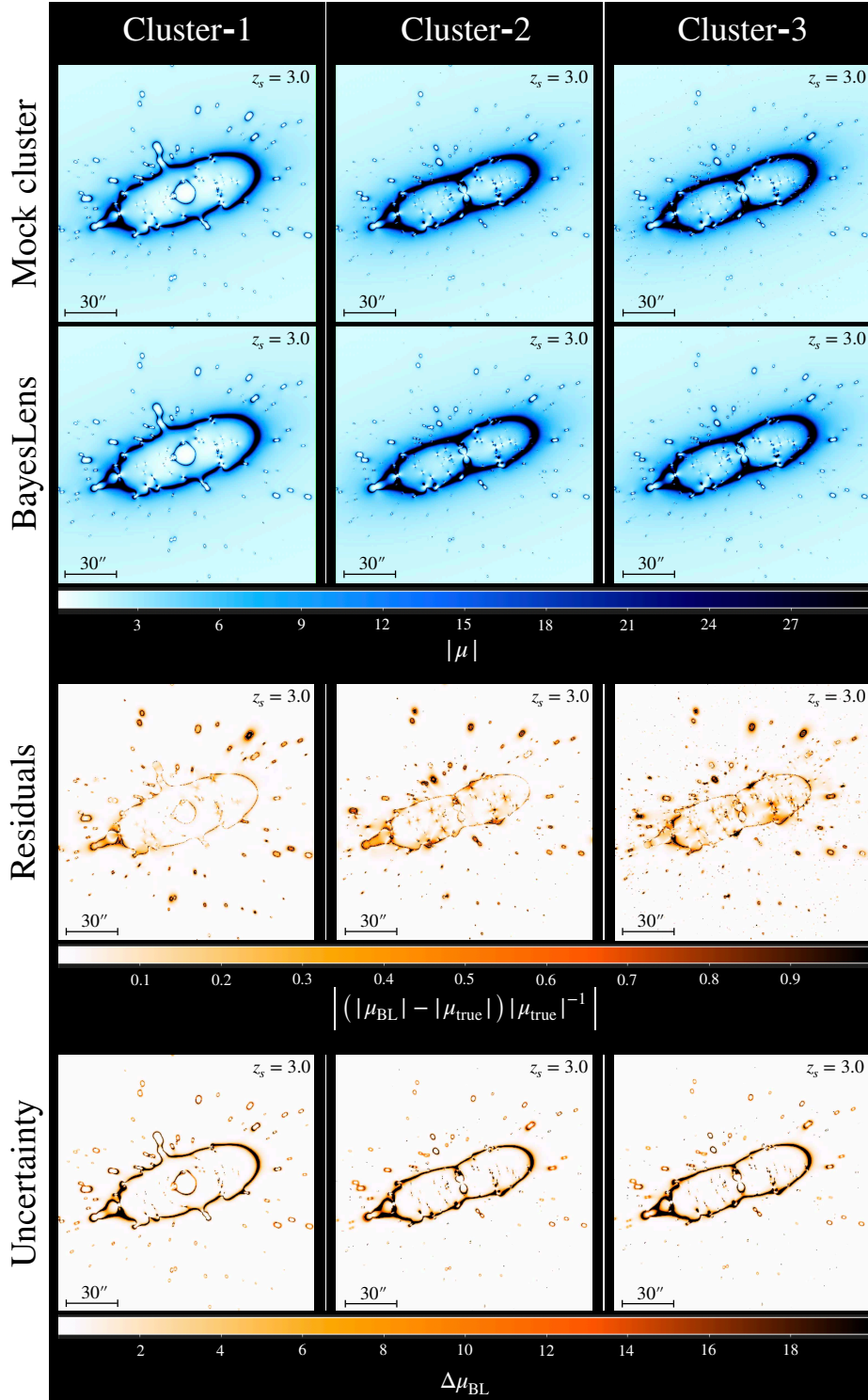
Fig. 8. Maps of absolute magnification for the three simulated clusters computed for a source redshift $z_s = 3.0$. In the *first line*, the magnification maps for the true Cluster-1, Cluster-2, and Cluster-3 mocks are plotted. The maps in the *second line* are obtained from BAYESLENS best-fit lens models of the clusters, while in the *third line* the normalized residuals between BAYESLENS and true magnifications, $\left\| |\mu_{BL}| - |\mu_{true}|/|\mu_{true}| \right\|$, are shown. In the *last line*, the absolute uncertainties associated with BAYESLENS magnification maps are plotted. These are computed considering 100 realizations of the lens models by randomly extracting 100 parameter samples from the MCMC chains and taking half of the difference between the 84th and 16th percentiles of magnification distributions in each pixel.

(namely Cluster-1,2,3). The magnification maps are well reproduced, and as can be expected, the appreciable differences happen only across the critical lines and in high magnification regions ($|\mu| \gtrsim 20$), where the posterior also predicts a wider variance in the magnification. We remark that these models are already an extension over the most current models, and the availability of samples from the posterior allows us to evaluate magnification uncertainties at all points.

While our set of hyperparameters is currently limited to those of the scaling relations (intercepts, slopes, scatter widths),

this can be in principle extended to cosmological parameters, enabling (fully self-consistent) cosmographic measurements. Our code is released publicly and documented[3].

References

Atek, H., Richard, J., Kneib, J.-P., & Schaerer, D. 2018, MNRAS, 479, 5184
Bacon, R., Accardo, M., Adjali, L., et al. 2012, The Messenger, 147, 4
Bergamini, P., Rosati, P., Mercurio, A., et al. 2019, A&A, 631, A130
Bergamini, P., Rosati, P., Vanzella, E., et al. 2021, A&A, 645, A140
Biviano, A., Rosati, P., Balestra, I., et al. 2013, A&A, 558, A1
Bonamigo, M., Grillo, C., Ettori, S., et al. 2018, ApJ, 864, 98
Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, ApJ, 843, 129
Bradač, M., Schneider, P., Lombardi, M., & Erben, T. 2005, A&A, 437, 39
Caminha, G. B., Grillo, C., Rosati, P., et al. 2016, A&A, 587, A80
Caminha, G. B., Grillo, C., Rosati, P., et al. 2017a, A&A, 600, A90
Caminha, G. B., Grillo, C., Rosati, P., et al. 2017b, A&A, 607, A93
Caminha, G. B., Rosati, P., Grillo, C., et al. 2019, A&A, 632, A36
Cerny, C., Sharon, K., Andrade-Santos, F., et al. 2018, ApJ, 859, 159
Coe, D., Salmon, B., Bradač, M., et al. 2019, ApJ, 884, 85
Diego, J. M., Protopapas, P., Sandvik, H. B., & Tegmark, M. 2005, MNRAS, 360, 477
Dye, S., & Taylor, A. 1998, MNRAS, 300, L23
Elíasdóttir, , Limousin, M., Richard, J., et al. 2007, ArXiv e-prints [arXiv:0710.5636]
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gao, L., White, S. D. M., Jenkins, A., Stoehr, F., & Springel, V. 2004, MNRAS, 355, 819
Gilmore, J., & Natarajan, P. 2009, MNRAS, 396, 354
Giocoli, C., Tormen, G., Sheth, R. K., & van den Bosch, F. C. 2010, MNRAS, 404, 502
Giocoli, C., Meneghetti, M., Bartelmann, M., Moscardini, L., & Boldrin, M. 2012, MNRAS, 421, 3343
Golse, G., Kneib, J. P., & Soucail, G. 2002, A&A, 387, 788

Goodman, J., & Weare, J. 2010, Commun. Appl. Math. Comput. Sci., 5, 65
Grillo, C., Gobat, R., Presotto, V., et al. 2014, ApJ, 786, 11
Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, ApJ, 800, 38
Grillo, C., Karman, W., Suyu, S. H., et al. 2016, ApJ, 822, 78
Grillo, C., Rosati, P., Suyu, S. H., et al. 2018, ApJ, 860, 94
Guérou, A., Krajnović, D., Epinat, B., et al. 2017, A&A, 608, A5
Hashimoto, T., Inoue, A. K., Mawatari, K., et al. 2019, PASJ, 71, 71
Hoag, A., Bradač, M., Brammer, G., et al. 2018, ApJ, 854, 39
Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2015, ApJ, 799, 12
Jauzac, M., Richard, J., Jullo, E., et al. 2015, MNRAS, 452, 1437
Jauzac, M., Harvey, D., & Massey, R. 2018, MNRAS, 477, 4046
Jullo, E., & Kneib, J.-P. 2009, MNRAS, 395, 1319
Jullo, E., Kneib, J.-P., Limousin, M., et al. 2007, New J. Phys., 9, 447
Kassiola, A., & Kovner, I. 1993, ApJ, 417, 450
Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016a, ApJ, 819, 114
Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016b, ApJ, 819, 114
Keeton, C. R. 2003, ApJ, 584, 664
Kneib, J.-P., Ellis, R. S., Smail, I., Couch, W. J., & Sharples, R. M. 1996, ApJ, 471, 643
Lagattuta, D. J., Richard, J., Clément, B., et al. 2017, MNRAS, 469, 3946
Limousin, M., Kneib, J.-P., & Natarajan, P. 2005, MNRAS, 356, 309
Limousin, M., Kneib, J. P., Bardeau, S., et al. 2007, A&A, 461, 881
Limousin, M., Richard, J., Jullo, E., et al. 2016, A&A, 588, A99
Livermore, R. C., Jones, T. A., Richard, J., et al. 2015, MNRAS, 450, 1812
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Magaña, J., Acebrón, A., Motta, V., et al. 2018, ApJ, 865, 122
Meneghetti, M., Natarajan, P., Coe, D., et al. 2017, MNRAS, 472, 3177
Meneghetti, M., Davoli, G., Bergamini, P., et al. 2020, Science, 369, 1347
Monna, A., Seitz, S., Zitrin, A., et al. 2015, MNRAS, 447, 1224
Monna, A., Seitz, S., Geller, M. J., et al. 2017, MNRAS, 465, 4589
Natarajan, P., & Kneib, J.-P. 1997, MNRAS, 287, 833
Niemiec, A., Jullo, E., Limousin, M., et al. 2017, MNRAS, 471, 1153
Parry, W. G., Grillo, C., Mercurio, A., et al. 2016, MNRAS, 458, 1493
Postman, M., Coe, D., Benítez, N., et al. 2012, ApJS, 199, 25
Priewe, J., Williams, L. L. R., Liesenborgs, J., Coe, D., & Rodney, S. A. 2017, MNRAS, 465, 1030
Richard, J., Jauzac, M., Limousin, M., et al. 2014, MNRAS, 444, 268
Rodney, S. A., Balestra, I., Bradac, M., et al. 2018, Nat. Astron., 2, 324
Rosati, P., Balestra, I., Grillo, C., et al. 2014, The Messenger, 158, 48
Saha, P., Williams, L. L. R., & Abdelsalam, H. M. 2001, in Gravitational Lensing: Recent Progress and Future Go, eds. T. G. Brainerd, & C. S. Kochanek, ASP Conf. Ser., 237, 279
Sartoris, B., Biviano, A., Rosati, P., et al. 2020, A&A, 637, A34
Schneider, P., & Sluse, D. 2014, A&A, 564, A103
Treu, T., Brammer, G., Diego, J. M., et al. 2016, ApJ, 817, 60
Umetsu, K., Medezinski, E., Nonino, M., et al. 2014, ApJ, 795, 163
Unruh, S., Schneider, P., & Sluse, D. 2017, A&A, 601, A77
Vanzella, E., Calura, F., Meneghetti, M., et al. 2017, MNRAS, 467, 4304
Vanzella, E., Calura, F., Meneghetti, M., et al. 2019, MNRAS, 483, 3618
Verdugo, T., de Diego, J. A., & Limousin, M. 2007, ApJ, 664, 702

---

[3] https://github.com/pietrobergamini89/BayesLens