Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **508,** 5587–5601 (2021)
Advance Access publication 2021 October 6

https://doi.org/10.1093/mnras/stab2857

A new framework for understanding systematic errors in cluster lens modelling – I. Selection and treatment of cluster member galaxies

Catie A. Raney, Charles R. Keeton ★ and Dhruv T. Zimmerman

Department of Physics and Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Road, Piscataway, NJ 08854, USA

Accepted 2021 September 24. Received 2021 August 31; in original form 2021 August 31

ABSTRACT

With high-quality data from programs like the *Hubble* Frontier Fields, cluster lensing has reached the point that models are dominated by systematic rather than statistical uncertainties. We introduce a Bayesian framework to quantify systematic effects by determining how different lens modelling choices affect the results. Our framework includes a new two-sample test for quantifying the difference between posterior probability distributions that are sampled by methods like Monte Carlo Markov chains. We use the framework to examine choices related to the selection and treatment of cluster member galaxies in two of the Frontier Field clusters: Abell 2744 and MACS J0416.1–2403. When selecting member galaxies, choices about depth and area affect the models; we find that model results are robust for an *I*-band magnitude limit of $m_{\rm lim} \geq 22.5$ mag and a radial cut of $r_{\rm lim} \geq 90$ arcsec (from the centre of the field), although the radial limit likely depends on the spatial extent of lensed images. Mass is typically assigned to galaxies using luminosity/mass scaling relations. We find that the slopes of the scaling relations can have significant effects on lens model parameters but only modest effects on lensing magnifications. Interestingly, scatter in the scaling relations affects the two fields differently. This analysis illustrates how our framework can be used to analyse lens modelling choices and guide future cluster lensing programs.

Key words: gravitational lensing: strong – methods: statistical – galaxies: clusters: individual: Abell 2744 – galaxies: clusters: individual: MACS J0416.1–2403.

1 INTRODUCTION

Gravitational lensing by galaxy clusters has become a powerful tool to study the origin and evolution of galaxies. A galaxy cluster acting as a lens can be used as a 'cosmic telescope' to magnify galaxies from the first billion years of cosmic history, at redshifts $z \sim 10$–11 (e.g. Zheng et al. 2012; Coe et al. 2013; Bouwens et al. 2014; Salmon et al. 2018, 2020) and help improve constraints on the high-redshift luminosity function (e.g. McLeod, McLure & Dunlop 2016; Bouwens et al. 2017). However, in order to estimate a lensed galaxy's intrinsic luminosity, a magnification must be determined in order to correct the observed luminosity. Since errors in magnification can thus propagate into other results, it is important to understand them and how they may be affected by modelling choices.

Cluster lens models are quite complex, as they describe mass distributions that are themselves very complex (see e.g. Kneib & Natarajan 2011). Modelling is made more complicated by the fact that clusters known for their lensing capabilities are often disturbed systems, because having multiple group- or cluster-sized haloes can significantly enhance the cross-section for strong lensing (Wong et al. 2012; French et al. 2014). In order to describe these complex systems, lens mass models often have 20+ parameters, and errors are often found using Markov chain Monte Carlo (MCMC) sampling of the parameter space. However, there are also many choices that have to be made when constructing a model, which do not explicitly appears as model parameters and thus are hidden from an MCMC analysis. This means that uncertainties revealed by an MCMC analysis, which are statistical in nature, will underestimate the true errors due to other systematic effects. Since the problem is a systematic one, it will not necessarily be solved by better data (e.g. Johnson & Sharon 2016; Meneghetti et al. 2017; Priewe et al. 2017).

The need to understanding model errors has influenced the design of programs like the *Hubble* Frontier Fields (HFF; Lotz et al. 2017). Many teams were invited to create lens models using their own methodologies, with the idea that the resulting ensemble of models would sample the range of systematic effects. In Raney et al. (2020b), we compared the most recent models of the six HFF clusters and found that systematic differences in magnifications were indeed larger than the statistical errors. While the HFF lens modelling program has provided a major advance in exploring systematic uncertainties, it was not a controlled experiment so it did not clearly reveal what aspects of lens modelling contribute most to differences in models.

We have begun a project to build an error budget for cluster lens modelling that quantifies the effects of different modelling choices. In this paper, we lay out a new statistical framework that can be used to more fully estimate model errors (Section 2), and then apply it to choices related to the treatment of galaxies within the cluster, namely cluster member selection (Section 3) and the scaling relations used to assign galaxy masses (Section 4). We discuss our results and place them in context with previous work in Section 5, and summarize our conclusions in Section 6. In a companion paper, we use our new framework to quantify systematic effects associated with the choice

★ E-mail: keeton@physics.rutgers.edu

of images used as constraints on lens models (Zimmerman, Keeton & Raney 2021).

2 METHODOLOGY

2.1 Model components

We use a parametric modelling technique where mass is assigned using density profiles described by a given set of parameters. We begin by outlining the different mass components that comprise our models. Specifically, we do this in terms of the deflection angle $\hat{\boldsymbol{\alpha}}$, which describes how much light is bent as it passes by the component. For ease in reading, we will more often refer to the reduced deflection angle $\boldsymbol{\alpha} = (D_{\rm ls}/D_{\rm s})\hat{\boldsymbol{\alpha}}$, which is scaled by $D_{\rm ls}$ and $D_{\rm s}$, the angular diameter distances between the lens and a given source, and the observer and that source, respectively.

The first component includes large-scale haloes that characterize the dark matter and/or intracluster medium. These haloes are described by the parameters $\boldsymbol{q}_{\rm halo}$ and contribute a deflection $\boldsymbol{\alpha}_{\rm halo}$. We use softened isothermal elliptical mass distributions for these haloes, but the framework presented here is general.

The cluster member galaxies are described by a pseudo-Jaffe model for the mass distribution. These galaxies are not all free to vary, but are instead assigned mass and radius using scaling relations based on their luminosity in relation to a reference galaxy. In this work, we use the brightest cluster galaxy (BCG) as the reference galaxy. The scaling relations can then be described by $\boldsymbol{q}_{\rm scale} = (\eta_E, \eta_t)$ such that the nominal Einstein radius and truncation radius for galaxy i are given by

$$\bar{b}_i = b_{\rm gal}\left(\frac{L_i}{L_{\rm bcg}}\right)^{\eta_E}, \tag{1}$$

$$\bar{a}_i = a_{\rm gal}\left(\frac{L_i}{L_{\rm bcg}}\right)^{\eta_t}, \tag{2}$$

where L_i and $L_{\rm bcg}$ refer to the $F814W$ luminosity for galaxy i and the BCG, respectively. However, given scatter in the scaling relations we should treat the actual parameters b_i and a_i as random variables; we use lognormal distributions with means $\bar{b}_i$ and $\bar{a}_i$ and scatter σ_b and σ_a that are derived from observed scaling relations (e.g. Brimioulle et al. 2013).

We use $\boldsymbol{q}_{\rm gal} = (a_{\rm gal}, b_{\rm gal})$ to describe the zero-points of the scaling relations. The parameters used to determine which galaxies are treated as cluster members, e.g. a magnitude or radial limit, are described by $\boldsymbol{q}_{\rm mem}$. We note that $\boldsymbol{q}_{\rm mem}$ and $\boldsymbol{q}_{\rm scale}$ are usually fixed when modelling, while $\boldsymbol{q}_{\rm gal}$ is either optimized or marginalized, although there are some exceptions (see e.g. Kawamata et al. 2016). Because of the scatter in the scaling relations, the deflection due to the cluster members is a random variable $\boldsymbol{\alpha}_{\rm mem}$ drawn from a probability distribution given by $P_{\rm mem}(\boldsymbol{\alpha}_{\rm mem}|\boldsymbol{q}_{\rm mem}, \boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm scale})$.

The mass along the line of sight (LOS) must be handled slightly differently. Specifically, we must consider the foreground structure that lies between the observer and the cluster differently than the total LOS structure. This is because mass in the foreground will affect the positions at which deflections in the cluster lens plane must be calculated. The deflection term for the foreground mass is a sum of the deflections of all perturbers at a redshift a less than that of the lens, specified as $a < l$, scaled by their respective distance scalings:

$$\boldsymbol{\alpha}_{\rm fg} = \sum_{a<l}\frac{D_{\rm al}}{D_{\rm l}}\hat{\boldsymbol{\alpha}}_a, \tag{3}$$

where $D_{\rm al}$ and $D_{\rm l}$ refer to the angular diameter distances between the perturber and lens and between the observer and lens, respectively. Similarly, the total LOS deflection is a sum of all perturbers at redshift b between the observer and the source specified as $b < s$:

$$\boldsymbol{\alpha}_{\rm los} = \sum_{b<s}\frac{D_{\rm bs}}{D_{\rm s}}\hat{\boldsymbol{\alpha}}_b, \tag{4}$$

where $D_{\rm bs}$ refers to the angular diameter distance between the perturber and the source. Note that the foreground mass appears in both $\boldsymbol{\alpha}_{\rm fg}$ and $\boldsymbol{\alpha}_{\rm los}$, but with different weightings.

In this work we only consider mass along the LOS in the form of galaxies, which are described similarly to the cluster member galaxies; we use $\boldsymbol{q}_{\rm los}$ to characterize them. Thus, the deflections are also random variables, now drawn from the probability distribution $P_{\rm los}(\boldsymbol{\alpha}_{\rm fg}, \boldsymbol{\alpha}_{\rm los}|\boldsymbol{q}_{\rm los}, \boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm scale})$. This is a joint probability distribution since foreground perturbers are found in both sums, so $\boldsymbol{\alpha}_{\rm los}$ and $\boldsymbol{\alpha}_{\rm fg}$ will be correlated. We note that the LOS galaxies can still be anchored to the BCG as long as we account for the redshift scaling in the zero-points of the scaling relations.

Our complete set of model parameters is then given as $\boldsymbol{q} = \{\boldsymbol{q}_{\rm halo}, \boldsymbol{q}_{\rm mem}, \boldsymbol{q}_{\rm gal}, \boldsymbol{q}_{\rm scale}, \boldsymbol{q}_{\rm los}\}$, and we can use these to draw $\boldsymbol{\alpha}_{\rm mem}$, $\boldsymbol{\alpha}_{\rm los}$, and $\boldsymbol{\alpha}_{\rm fg}$ from their probability distributions.

2.2 Statistical framework

In a Bayesian framework, we can write down a likelihood for the data given a model, which explicitly depends on $\boldsymbol{\alpha}_{\rm mem}$, $\boldsymbol{\alpha}_{\rm los}$, and $\boldsymbol{\alpha}_{\rm fg}$ through the lens equation. If those random variables are Gaussian, we can marginalize over them as described in Appendix A. In particular, equation (A14) gives our final expression for the marginalized posterior distribution for the parameters given the data. That equation provides the formal tool we use to evaluate different models.

2.3 Modelling procedure

We begin by determining which constraints to use on the model. Our sample is the same one we have used previously (see Raney, Keeton & Brennan 2020a), which primarily consists of spectroscopically confirmed images. This is a conservative sample, which we feel is prudent to test our new framework. Future work will examine the choice of image constraints. We also note that for simplicity here we use the 2D approximation for LOS galaxies that we employed in that previous work, so we ignore any $\boldsymbol{\alpha}_{\rm fg}$ terms.

We then must determine which observed galaxies to treat as members of the cluster, i.e. we must specify $\boldsymbol{q}_{\rm mem}$ as discussed above. In this work, we consider two limits to determine cluster membership: magnitude limit and radial cut. After a cluster member sample is found, we determine how mass is assigned to the galaxies via $\boldsymbol{q}_{\rm scale}$, which is another thing we explore in this work. Again, we include scatter in the scaling relations through σ_b and σ_a.

To find the galaxy deflections, we run a large suite of Monte Carlo simulations. Each galaxy is assigned a mass drawn randomly from its allowed range. We note that we explicitly run simulations with different values for the radius parameter zero-point ($a_{\rm gal}$); we do not need to do this for the mass parameter zero-point as the deflection is directly proportional to $b_{\rm gal}$ so rescaling to different values is trivial. We sum the deflections from all galaxies to calculate $\boldsymbol{\alpha}_{\rm mem}$. We repeat this for the LOS galaxies to calculate $\boldsymbol{\alpha}_{\rm los}$. From this, we can characterize $P(\boldsymbol{\alpha}_{\rm mem})$ and $P(\boldsymbol{\alpha}_{\rm los})$. As we will show in the next section, these distributions appear Gaussian, and thus can be summarized by mean vectors and covariance matrices, e.g. $\bar{\boldsymbol{a}}_{\rm mem}$, $\bar{\boldsymbol{a}}_{\rm los}$, and $\boldsymbol{C}_{\rm mem}$, $\boldsymbol{C}_{\rm los}$.

With this, we can use equation (A14) to calculate our full posterior probability distribution, $P_{\rm marg}(\boldsymbol{q})$, for a given set of parameters $\boldsymbol{q}$ and

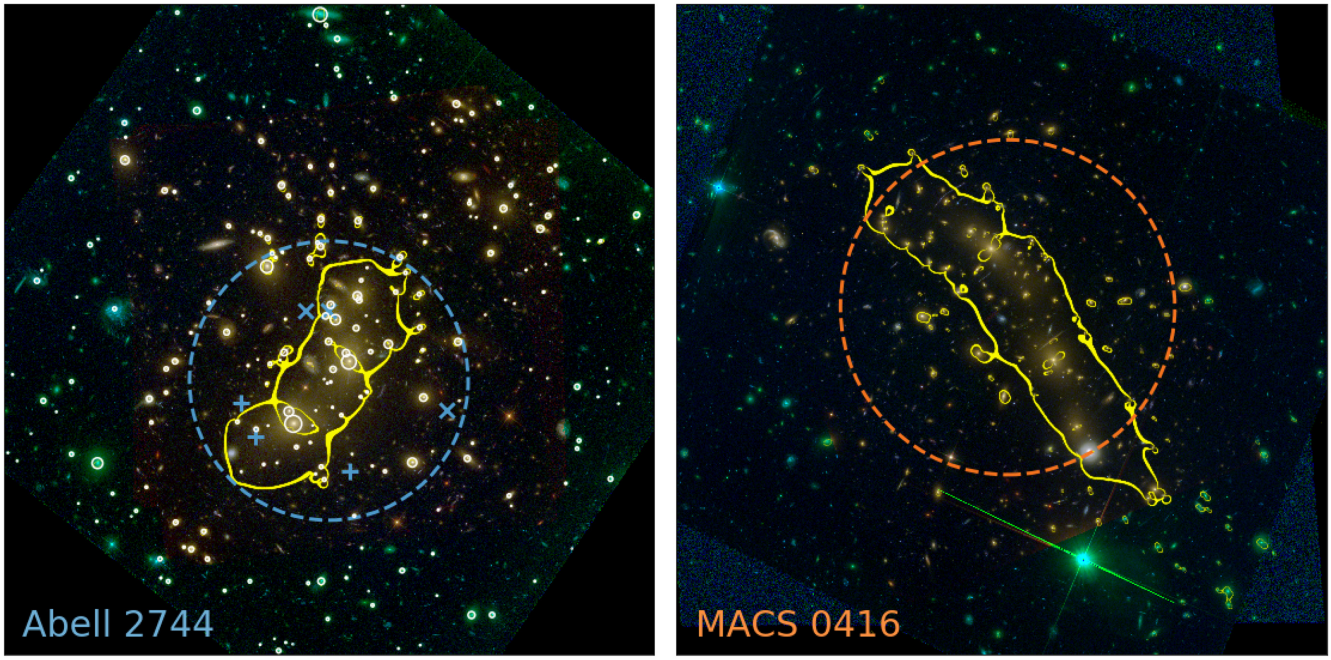

Figure 1. *HST* multiband colour images (produced using Trilogy, Coe et al. 2012) of the two fields we consider in this work. The critical curves for a source redshift of $z = 9$ from our fiducial 2D model as presented in Raney et al. (2020a) are shown in yellow. The dashed circles represent the extent of the images used as constraints and are centred at the image epicentre; they are 45 and 54 arcsec for Abell 2744 and MACS 0416, respectively. For Abell 2744, we show the galaxies used for the fiducial model with white circles whose size is proportional to luminosity. We also include three images from source 1 (+) and source 18 (×). Each panel is 3.5 arcmin on a side.

use MCMC methods to sample it. The samples obtained from the MCMC analysis can then be used to generate surface density or magnification maps, or any other lensing quantities. We can test different model choices by changing $\boldsymbol{q}$ and rerunning this procedure.

2.4 The galaxy clusters

In this work, we apply our methodology to two clusters: Abell 2744 and MACS J0416.1–2403, shown in Fig. 1. These are both fields that were observed during the *Hubble* Frontier Fields program, which used *HST* in conjunction with other space- and ground-based telescopes to create some of the deepest observations of lensing clusters. Specifically, six clusters were chosen, all based on their known lensing capabilities, and the goal was to use these fields to find and study high redshift galaxies. In addition, many teams were invited and/or funded to model these fields, which allowed the errors to be better estimated. Abell 2744 and MACS 0416 were the first two clusters observed by *HST* for this program.

Abell 2744 ($z = 0.308$), as indicated by its name, is part of the South sky Abell galaxy cluster catalogue (Abell, Corwin & Olowin 1989). It is a very large and complex cluster as it includes many systems that are in the process of merging or have recently merged. This is evidenced by the fact that there are five possible brightest cluster galaxies (BCG) in the field (Mann & Ebeling 2012); optical and X-ray studies also suggest the cluster has had two mergers in the recent past (Kempner & David 2004; Owers et al. 2011; Merten et al. 2011). A spectroscopic survey by Mahler et al. (2018) vastly increased the number of confirmed lensed images in the field. In this work, we use $N_{\rm img} = 71$ of these images from $N_{\rm src} = 24$ sources as constraints on our models.

MACS 0416 ($z = 0.396$) is a field from the Massive Cluster Survey (MACS; Ebeling, Edge & Henry 2001). It too is likely undergoing a merger; there are two clear BCGs in the field, and the mass distribution itself is elongated between the two galaxies. The X-ray map has two peaks as well (Mann & Ebeling 2012). This field has the most spectroscopically confirmed images of the six Frontier Fields; we use $N_{\rm img} = 95$ images from $N_{\rm src} = 35$ sources as constraints. The specific list of images for both fields can be found in the appendix of Raney et al. (2020a).

For both fields, our models consist of three large-scale haloes, in addition to the cluster members and LOS galaxies. We fit for the parameters of the large-scale haloes (mass, position, ellipticity and position angle, and radius), as well the zero-points for the member galaxies. We also include shear to account for any asymmetries in the mass distribution on large scales.

2.5 Distance metric

Since the results of our MCMC modelling have both large volume and high dimensionality, it is not trivial to compare two posterior distributions in order to see how similar they are. Thus, we introduce a distance metric to quantify these differences. Let us consider samples from two distributions: $\boldsymbol{x}_i$ for $i = 1,..., n_X$ from distribution X and $\boldsymbol{y}_i$ for $i = 1,..., n_Y$ from distribution Y. The root-mean-square (rms) distance between points in X is D_X such that

$$D_X^2 = \frac{1}{n_X(n_X - 1)} \sum_{i=1}^{n_X} \sum_{j=1}^{n_X} ||\boldsymbol{x}_i - \boldsymbol{x}_j||_C^2, \tag{5}$$

where $|| - ||_C$ denotes the distance between points computed with a given metric $\boldsymbol{C}$. For example, Euclidian distance would be calculated using the identity matrix as $\boldsymbol{C}$, whereas the Mahalanobis distance would use the covariance matrix. The rms distance between points in Y can be written similarly. The rms distance between points in

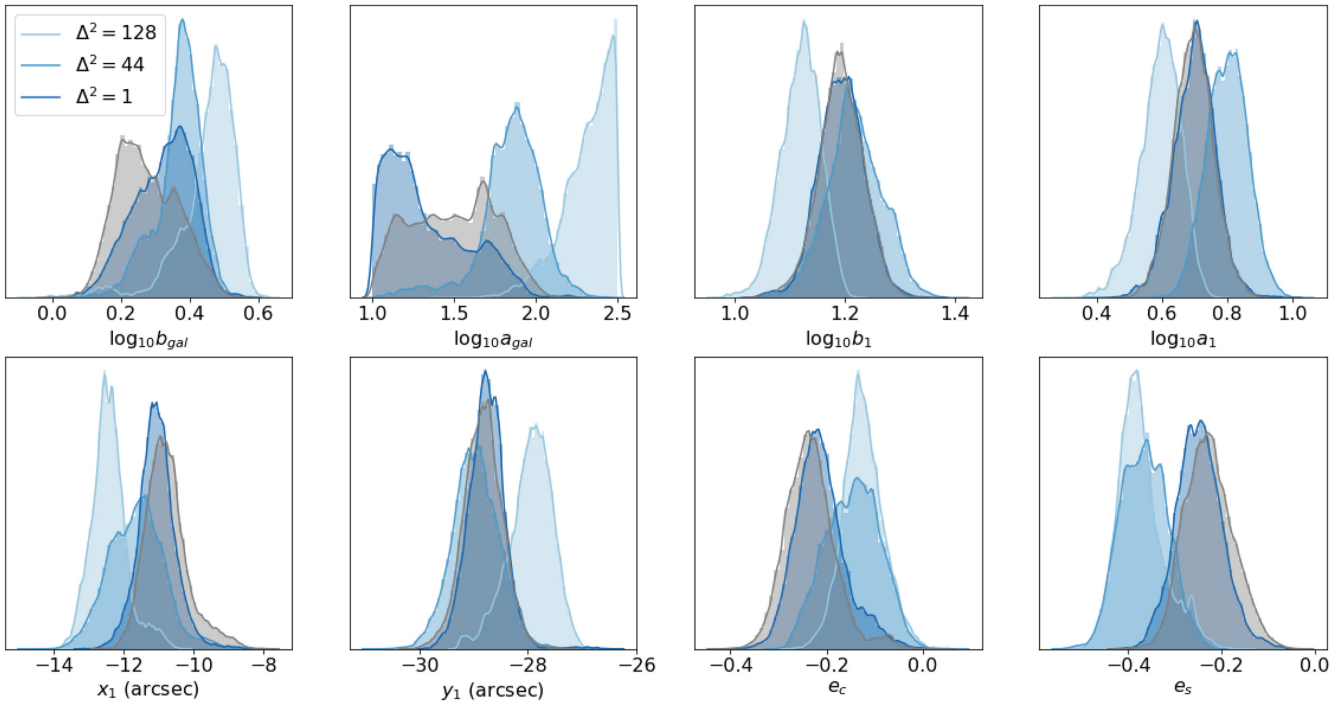

Figure 2. An example to illustrate how differences in parameter distributions translate to values of the distance metric Δ^2. Specifically, we show distributions for the galaxy mass and radius zero-points ($b_{\rm gal}$, $a_{\rm gal}$), mass and core radius parameters of the first dark matter halo (b_1, s_1), the spatial position of that halo as measured in arcsec from the image epicentre (x_1, y_1), and terms to describe the ellipticity of the halo (given in quasi-Cartesian coordinates $e_c = e\cos 2\theta_e$ and $e_s = e\sin 2\theta_e$). The different shades of blue represent three models of the field Abell 2744, chosen for their low, medium, and high Δ^2 values, measured with respect to the fiducial model shown in grey. The model with the lowest value of Δ^2 (dark blue) is most similar to the fiducial model, although differences exist; the other two models are increasingly different. We note that Δ^2 is calculated in the full parameter space.

different samples can be written as

$$D^2_{XY} = \frac{1}{n_X n_Y} \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} ||\boldsymbol{x}_i - \boldsymbol{y}_j||^2_C. \tag{6}$$

We can then define the difference between *distributions* as

$$\Delta^2 = D^2_{XY} - D_X D_Y. \tag{7}$$

Appendix B shows that this distance metric can be simplified in terms of the means $\boldsymbol{m}_X, \boldsymbol{m}_Y$ and covariance matrices $\boldsymbol{C}_X, \boldsymbol{C}_Y$ for the two samples as

$$\Delta^2 = (\boldsymbol{m}_X - \boldsymbol{m}_Y)^T \boldsymbol{C}^{-1} (\boldsymbol{m}_X - \boldsymbol{m}_Y) + \left[\sqrt{\mathrm{tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X)} - \sqrt{\mathrm{tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_Y)} \right]^2, \tag{8}$$

where $\boldsymbol{C}$ denotes the pooled covariance $\boldsymbol{C} = (\boldsymbol{C}_X + \boldsymbol{C}_Y)/2$. The first term measures a shift in the means, and is identical to Hotelling's t^2 statistic. The second term measures a shift in the shapes of the distributions. If two distributions are identical, the metric will return zero, whereas very different distributions will return a high value.

As a way to build intuition for the distance metric, we demonstrate how different Δ^2 values correspond to changes in parameter and deflection distributions. Fig. 2 shows distributions for a subset of our model parameters for four different models of the field Abell 2744. Our fiducial model is shown in grey, whereas the other three models were picked to represent a wide range in Δ^2 values measured in relation to the fiducial model. (Note that Δ^2 is calculated in the full parameter space, even though the figure shows 1D histograms for simplicity.) We see that the model with the smallest Δ^2 metric is also most similar to the fiducial model, as expected, although they are not identical. The other two models become increasingly different from the fiducial model.

We also show the cluster member galaxy deflection distribution $\boldsymbol{\alpha}_{\rm mem}$ for these three models in Fig. 3. Specifically, we show the deflections for two sets of images in Abell 2744, as indicated in Fig. 1. We now do not show the fiducial model, but represent the mean of its deflection distribution with the black $\times$. We see that the model with the lowest Δ^2 statistic and thus the most similar parameters lies closer to the fiducial mean.

We also use this to illustrate variations in the deflection distributions. For example, the outer images of source 1, namely images 1.1 and 1.3, are relatively far from any galaxies; image 1.2 is close to a galaxy, but that galaxy is relatively small. For source 18, we see that the right image 18.3 is far from the cluster core and thus its distributions look similar to those of source 1. However, the other two images are closer to the cluster core and 18.2 specifically is close to three galaxies. We see that the deflection distributions for 18.1 and 18.2 thus look different, with more overlap between two of the models.

3 GALAXY MEMBER SELECTION

3.1 Magnitude limit

Let us first examine the magnitude limit $m_{\rm lim}$. Our fiducial models used an $F814W$ magnitude limit of 23.5 when selecting cluster members, but now we consider changing that limit (while holding all of the other member selection criteria fixed).

We measure the effect of the magnitude limit on model parameters first by employing the Δ^2 statistic, as shown in Fig. 4. We see that only including the very brightest galaxies, i.e. using a low magnitude limit,

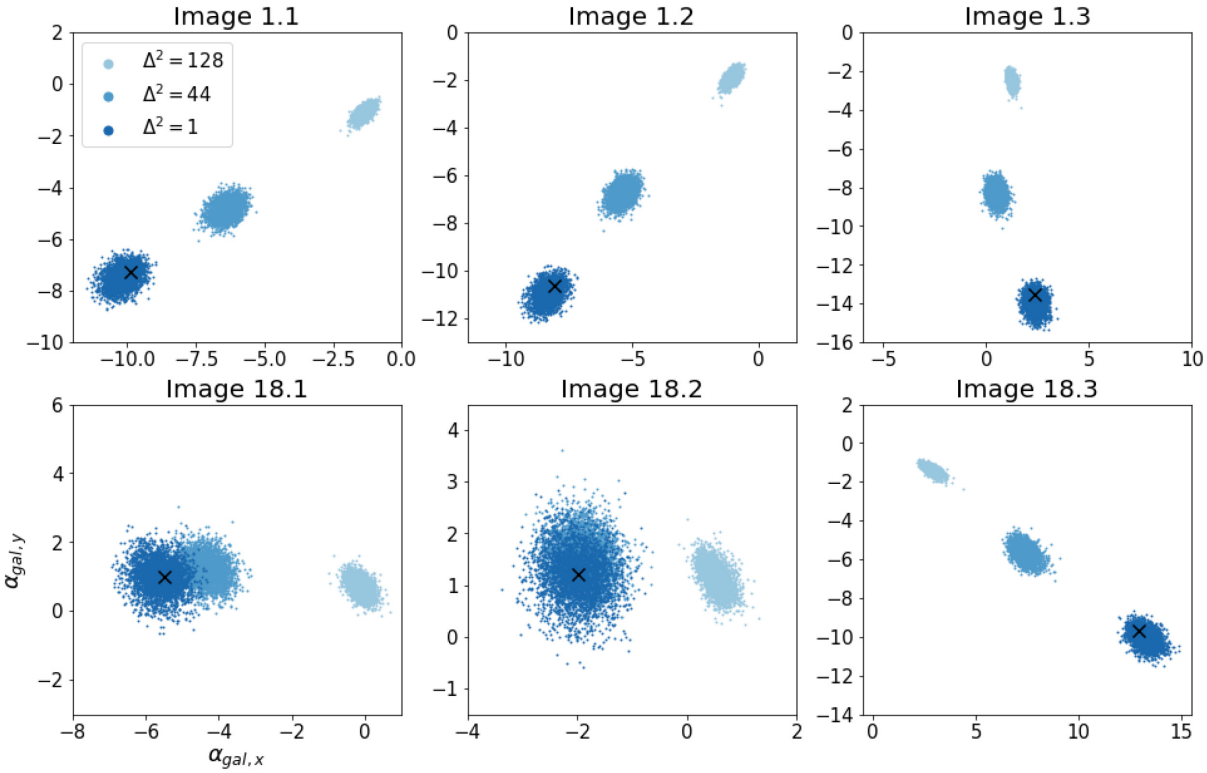

Figure 3. Samples from the distributions of cluster member galaxy deflections $\boldsymbol{\alpha}_{\rm mem}$ for two image sets in the field Abell 2744, taken from the models shown in Fig. 2; the means of the fiducial model are marked with $\times$. These are the same six images labelled in Fig. 1, with images from source 1 labelled as $+$ and images from source 18 labelled as $\times$. We see that sometimes there are distinct differences between the distributions, while other times there is significant overlap.

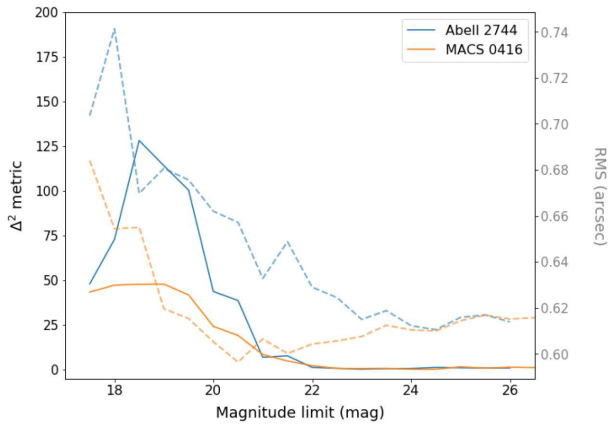

Figure 4. Distance metric Δ^2 as a function of magnitude limit used during cluster member selection for Abell 2744 (blue) and MACS 0416 (orange). The rms is also shown via the dashed lines with values as indicated by the right-hand labels.

produces a very different model than the fiducial, thus producing a large Δ^2 statistic. The differences start to decrease past a limit of 20 mag, and by a limit of 22 mag the differences are quite small. This is an expected result: ignoring more luminous, and thus more massive, galaxies limits the model's ability to fit the image positions on smaller scales. This result is true for both fields, although the differences are larger in Abell 2744.

As another point of comparison, we also include the rms, which is the square root of the mean squared distance between predicted and observed image positions. It is worthwhile to note that some models have similar rms values, but very different parameters as indicated by a higher Δ^2 value: for example, comparing $m_{\rm lim} = 19$ to $m_{\rm lim} = 26$ in MACS 0416. This is a known limitation of rms and something that has been seen previously, but the metric can still offer important insight (see e.g. Johnson & Sharon 2016; Priewe et al. 2017; Remolina González, Sharon & Mahler 2018).

We examine the effect of the magnitude limit on magnifications in terms of both accuracy and precision via a conditional probability distribution $P(\mu_x|\mu_{\rm ref})$. Specifically, we can choose a certain reference model, say $m_{\rm lim} = 17.5$, and find all of the pixels that equal a certain value, say $\mu = 2$, in the magnification map produced by that model. We then create a distribution $P(\mu_x|\mu_x)$ from those pixels across all of the realizations of that model. These realizations are created from parameters resulting from the MCMC analysis; they also include scatter in the galaxy mass–luminosity scaling relations. This would then test the precision, i.e. the scatter in the magnification maps of a given model. Similarly, we can also explore the accuracy of a given model, i.e. how well it can predict the magnifications of a different model. This new distribution $P(\mu_x|\mu_{\rm fid})$ is created by using the fiducial model as our reference and comparing all models against that one.

The results of this analysis are shown in Fig. 5 for both fields, where we plot the medians of the distributions previously described for a variety of magnitude limits as indicated by the colour of the lines. We plot the results of our fiducial model in black for easy comparison; the black lines for each field should be similar between panels since they are both $P(\mu_{\rm fid}|\mu_{\rm fid})$. For both fields, we see that the curves move away from the one-to-one line and towards lower values as magnification increases. This shift towards lower values is to be expected, and results from the non-linear quality of a magnification map; specifically, areas of high magnification are more rare than

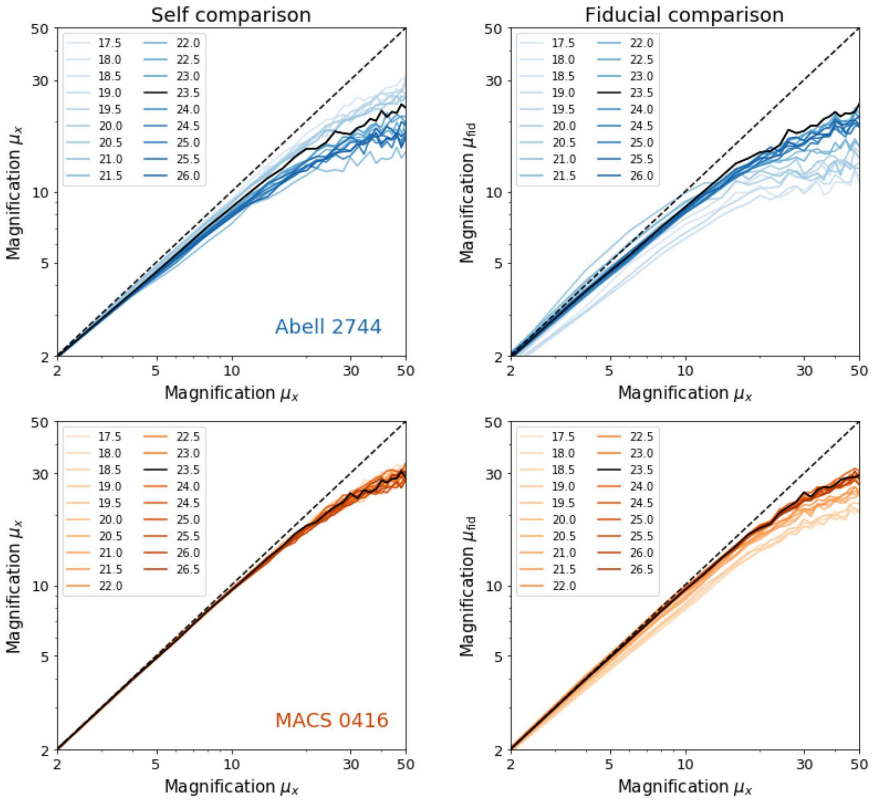

Figure 5. *Left*: Medians of the conditional probability distribution $P(\mu_x|\mu_x)$ where x denotes the specific magnitude limit of the model, as indicated by the colour. Results for Abell 2744 are shown on top, while MACS 0416 are on bottom. *Right*: Similar to the left-hand panels, but now comparing a model x with a given magnitude limit to the fiducial model with $m_{\rm lim} = 23.5$. Note: the black curve indicates the fiducial model, and thus has similar values between panels of a specific field.

those of lower magnification, so if model parameters shift, a given pixel will more likely have a smaller magnification than larger (for more detail, see appendix A in Raney et al. 2020b).

We see that the outcome is quite different between the two fields. In Abell 2744, there is a clear gradient in the self comparison panel, such that models with a brighter magnitude limit, and thus fewer galaxies, have a higher precision than those with a higher limit. Since we include scatter in the scaling relations used to assign mass to the galaxies, this is perhaps not surprising: there would be less scatter in magnifications if there are fewer galaxies being varied. However, this cannot be the whole story, as made clear by the self comparison results of MACS 0416 which are very similar across the different magnitude limits. This is something we will explore later in Section 5.

The fiducial comparison results are quite different, where a gradient is seen for both fields, but going in the opposite direction than in the self comparisons. The brighter magnitude limit models which had higher precision are seen to have lower accuracy, i.e. they are poor predictors of the fiducial model. Conversely, models that have a magnitude limit similar to that of the fiducial model are more accurate. Although this holds true for both fields, we see that there is a wider range in accuracies among the models of Abell 2744 than MACS 0416.

3.2 Radial limit

We also analyse the effect of a radial limit on our model. Specifically, we only include galaxies in the model if they fall within $r_{\rm lim}$ of the centre of the field, as defined by the epicentre of the images we use as constraints. We consider a range for $r_{\rm lim}$ from 10 to 150 arcsec for Abell 2744 and to 180 arcsec for MACS 0416; the difference in upper limit arises due to a disparity in survey coverage for the two fields. The resulting differences in parameter distributions as measured by Δ^2 is shown in Fig. 6. We also include the radial limits of the image constraints, denoted by the vertical lines in the figure, which are also shown as the dashed circles in Fig. 1. As in Fig. 4, we include rms values for reference.

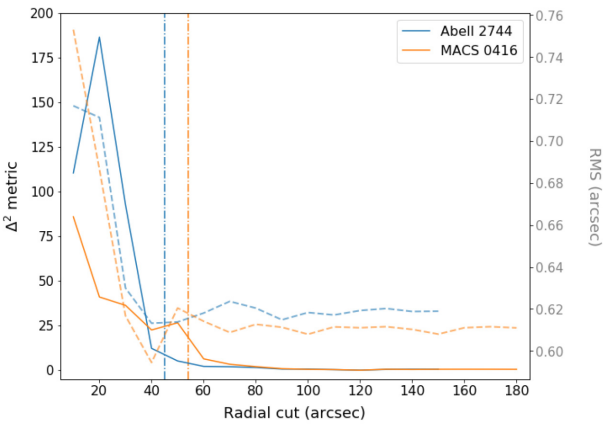

Figure 6. Distance metric Δ^2 as a function of radial cut used during cluster member selection for Abell 2744 (blue) and MACS 0416 (orange). The rms is also shown via the dashed lines with values as indicated by the right-hand labels. The vertical dot–dashed lines indicate the radial limit of the images used to constrain the model for each of the fields.

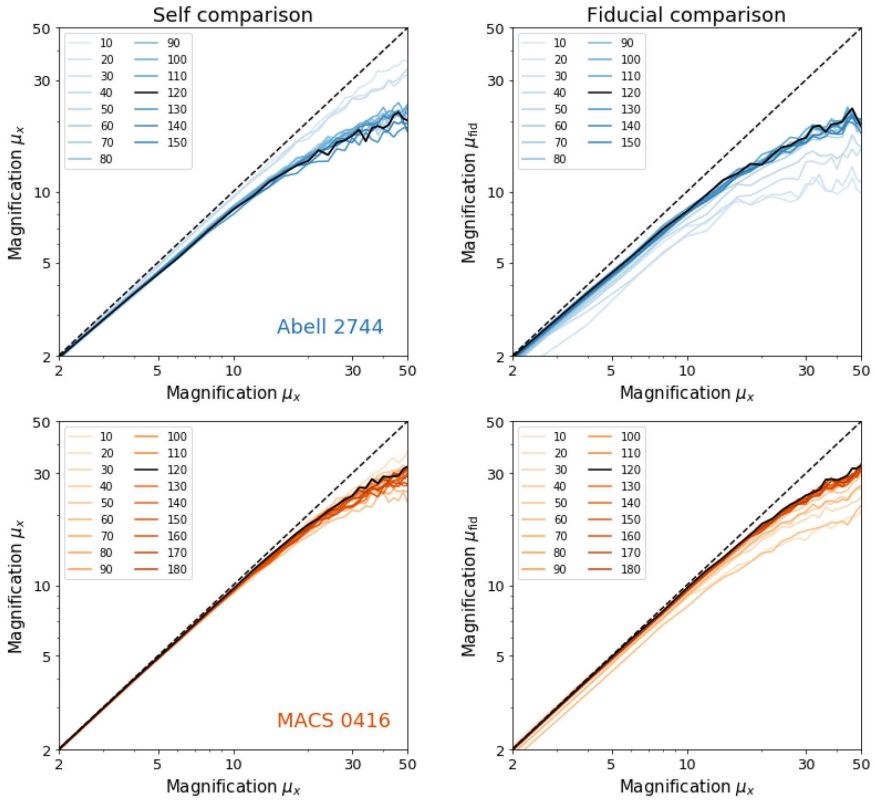

Figure 7. Similar to Fig. 5, but varying the radial limit. The legend indicates the radial cut in arcsec.

In Abell 2744, we see a sharp decline in Δ^2 values which then flattens out around $r_{\rm lim} = 40$ arcsec, behaviour that is also seen in the rms values. This may be due to the fact that images only extend to 45 arcsec from the epicentre, and thus the majority of galaxies near the images would be included at 40 arcsec. MACS 0416 also shows a decrease in Δ^2 values, although not one as dramatic as Abell 2744. Indeed, we again see that the highest values are found when $r_{\rm lim}$ is smaller than the radial extent of the image constraints. This indicates that small-scale structure outside the range of images is not very important, and the model is able to compensate in other ways (e.g. through shear). We do note that our image constraint selection is fairly conservative, and thus the images used here cover a smaller area than if we included all candidate images.

As in our magnitude limit analysis, we also consider the effect the radial limit has on the magnification of the models. We first examine the effects of varying $r_{\rm lim}$ on the accuracy of the models using a self-comparison, as shown in the left-hand panels of Fig. 7. The results are somewhat different than what we saw before. Specifically, for Abell 2744 the gradient we saw before is gone; instead, we see a few models with small radial cuts that have high precision, and then the rest of the models all have about the same results. Conversely, all of the models of MACS 0416 are clustred together, although with slightly more spread than when the magnitude limit was varied.

The results for the fiducial comparison conditional probability distribution, which tests a model's ability to predict the magnifications of the fiducial model, is shown in the right-hand panels of Fig. 7. As we saw previously, models of Abell 2744 with high precision, i.e. those with a smaller radial limit, have less accuracy when predicting the magnifications of the fiducial model. Similar results are seen in MACS 0416, including a cluster of models which are able to predict the fiducial magnifications well.

4 SCALING RELATIONS AND SCATTER

Once the cluster member galaxies are chose, they must be assigned mass as described in Section 2. Since there is scatter in the mass–luminosity relation, scatter can also be added. We note that galaxy kinematics, specifically measured velocity dispersions of cluster members, can be used to further constrain the galaxy parameters (see e.g. Monna et al. 2017; Bergamini et al. 2019), but we do not include such analysis in this work.

We consider two sources of possible systematic error: the slopes of the scaling relations and the scatter in galaxy mass and radius. Our fiducial model had scaling relation slopes of $\boldsymbol{q}_{\rm scale} = (\eta_E, \eta_t) = (0.5, 0.4)$. Other teams modelling the *Hubble* Frontier Fields used different values, such as $\boldsymbol{q}_{\rm scale} = (0.5, 0.5)$ (Johnson et al. 2014; Richard et al. 2014), $\boldsymbol{q}_{\rm scale} = (0.7, 0.5)$ (Caminha et al. 2017), and $\boldsymbol{q}_{\rm scale} = (0.54, 0.66)$ (Bergamini et al. 2019), the latter of which used galaxy kinematic information. In addition, our fiducial models included scatter of 0.1 dex in $\log_{10}R_E$ and 0.03 dex in $\log_{10}r_t$ for the member galaxies, and scatter of 0.3 dex in $\log_{10}R_E$ for the LOS galaxies.[1] Our choices were chosen based on observational results presented in Brimioulle et al. (2013).

4.1 Scaling relation slopes

To explore the systematic effects of changing $\boldsymbol{q}_{\rm scale}$, we considered a grid of η_E, η_t values ranging from 0.4 to 0.8 for both variables. These values were the only things that changed in the modelling process; the cluster member selection used fiducial values and was not varied.

[1]In our models, the LOS galaxies are assumed to be in the field and thus untruncated. In practice we set their truncation radius to a fixed, large value.

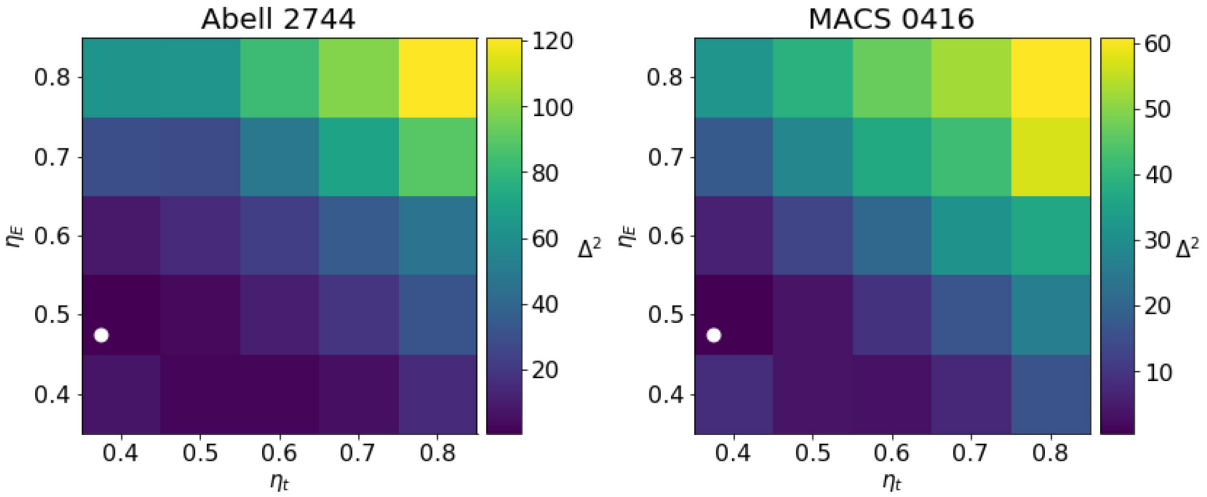

Figure 8. Distance metric Δ^2 for various values of η_E and η_t for Abell 2744 (left) and MACS 0416 (right). These values are calculated with reference to our fiducial values $\eta_E, \eta_t = (0.5, 0.4)$, denoted by the white circles.

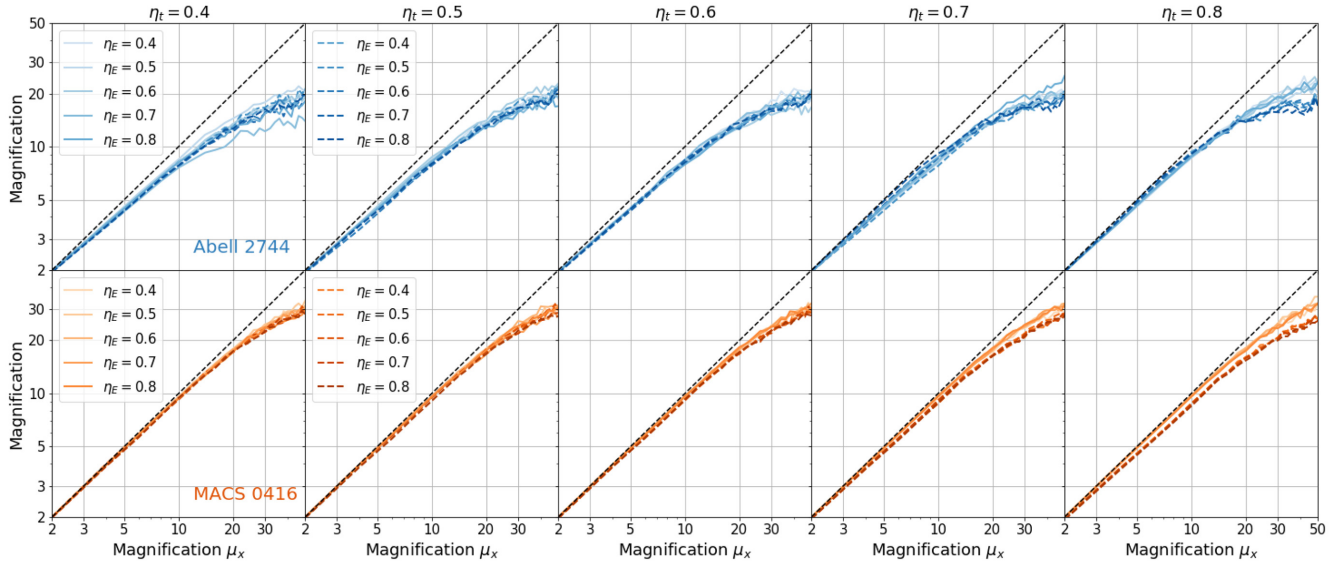

Figure 9. Similar to Fig. 5 for varying values of η_E, η_t. Specifically, panels indicate a fixed value of η_t as indicated, while the colours indicate the value of η_E; colours for a given η_E value are the same across panels of each field. Line style and brightness indicates whether the conditional probability distribution for a given model was calculated with respect to itself (lighter; solid) or with respect to the fiducial model (darker; dashed).

We first ask how this change affects the model parameters. Specifically, we calculate the distance metric between the new models and our fiducial model, as shown in Fig. 8. We see that, for both fields, Δ^2 increases in conjunction with both η_E and η_t, indicating that the model parameters become more different as a function of both variables. However, there seems to be more change as a result of varying η_E, as indicated by the fact that the top left corner has a higher value than the bottom right. We also note the difference in colour bar values for the fields: there is an increased difference among the Abell 2744 models than those of MACS 0416.

We also looked at the magnification maps for all of the grid models and performed a similar analysis to the conditional probability distributions shown previously. The results are shown in Fig. 9. Here, we have collapsed our results vertically, such that each panel has a fixed value of η_t, while the curves indicate varying values of η_E. We include both the results of comparing a model with itself (lighter, solid curves) and with the fiducial model (darker, dashed curves).

In Abell 2744, there are a few trends that can be seen. Among the self-comparisons, models with a lower η_t have a higher spread in precisions as η_E is changed, while models with a higher η_t have smaller variations and higher precision at greater magnification. Interestingly, both accuracy and precision of the models at lower magnifications increase as η_t is increased, as evidenced by the curves shifting towards the one-to-one line. The accuracy at larger magnifications does not vary with η_t or η_E.

For MACS 0416, there is less variation between the models as compared with Abell 2744, as expected given that the Δ^2 values are smaller. Specifically, there is almost no change as η_E increases. However, we do see that increasing η_t leads to a decrease in model accuracies across the range of magnifications.

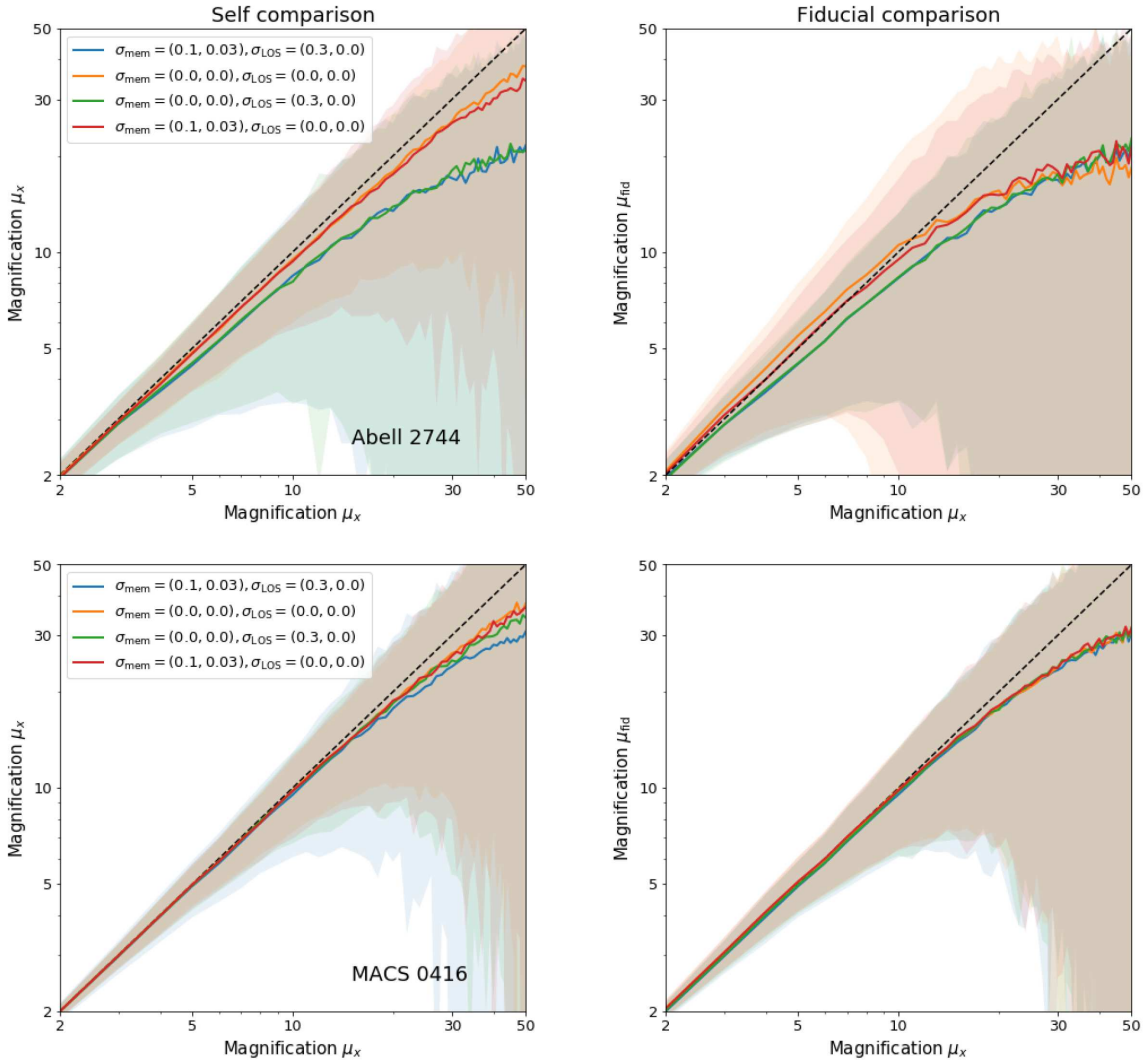

Figure 10. Similar to Fig. 5, but varying the scatter in ($\log_{10}R_E$, $\log_{10}r_t$) for cluster member and LOS galaxies (note that values are given in dex). We also include the 1σ error bars for each model.

4.2 Scaling relation scatter

Finally, we ask how including scatter in the mass and radius scaling relations affects model parameters and magnification. Specifically, we allow the member and LOS galaxies to have either the fiducial or zero scatter. We also test the cases where scatter is added to only the cluster member galaxies while LOS galaxies are held fixed, and vice versa as a way to test the relative importance of these populations.

For Abell 2744, we find that the differences between the model with fiducial scatter and no scatter in both cluster member and LOS galaxy populations are relatively small, as evidenced by a Δ^2 value of 27; for MACS 0416, the difference is even smaller at 13. If there is scatter in the cluster members but not the LOS galaxies, the differences decrease with values of $\Delta^2 = 14$ and 7 for Abell 2744 and MACS 0416, respectively. The smallest difference is seen when the cluster members do not include scatter, but the LOS do: $\Delta^2 = 4.5$ for Abell 2744 and 3.4 for MACS 0416.

We also consider the magnification distributions of these four models. We perform a similar analysis to the previous sections, where we consider the self and fiducial comparisons, as shown in Fig. 10. For the self-comparisons of Abell 2744, we see a clear difference between the models with scatter in the LOS galaxies (blue and green) and those without (orange and red), where the medians of the probability distributions of the former two models are very similar to one another, as are those of the latter two. The 1σ error bars for the model with no scatter in either galaxy populations is smaller than those of the model with cluster member scatter. The error bars for the two models with LOS galaxy scatter are both very similar and very large. For the fiducial comparison, we see that the two models without LOS galaxy scatter are relatively unchanged as compared to the self comparison panel, which is unsurprising. The models with LOS galaxy scatter have much larger error bars, but medians that are quite close to the one-to-one line, at least at lower magnifications.

In MACS 0416, the results are quite different. There is a slight gradient in the medians of the self-comparisons going from no scatter in either galaxy population, to scatter in only the cluster members population, to the scatter in only the LOS population, with the fiducial model having the least precision. However, the differences are much less than what we saw in Abell 2744. Further, the error bars also show that there is less scatter within the magnifications themselves, and again a smaller difference in magnification scatter is seen between the models. For the fiducial comparison, all models are indistinguishable, indicating that they all have similar accuracies.

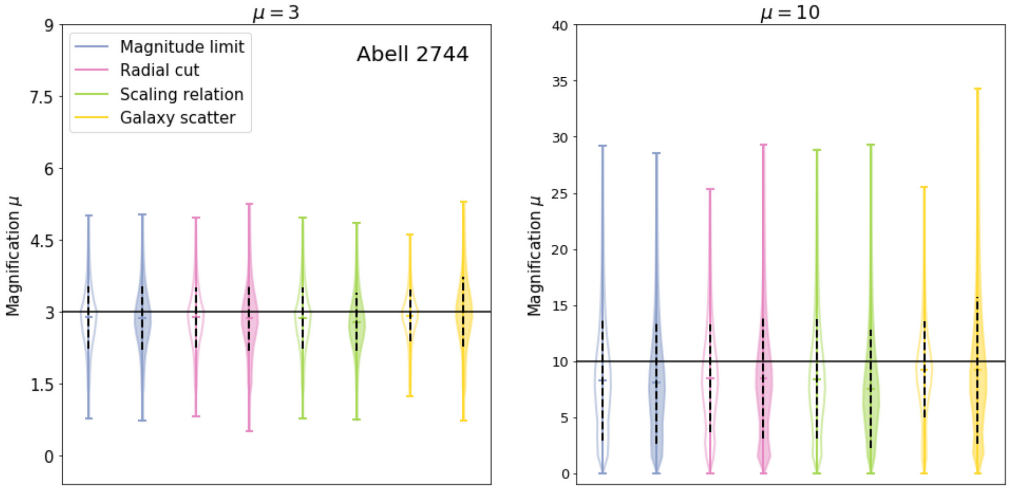

Figure 11. We show the full distributions calculated from the various modelling choices discussed in this work. Specifically, for magnitude limit and radial cut, we use only models that have reasonable limit values, i.e. $m_{\rm lim} \geq 22.5$ mag and $r_{\rm lim} \geq 90$ arcsec; for galaxy scatter and scaling relation distributions, all models are used. Unfilled violin bodies represent the statistical scatter in the models and are created using the self-comparisons previously described. Filled violin bodies compare every model in a given model choice grouping to every other model in that grouping, thus probing systematic error introduced by a given modelling choice. We do this analysis for a reference magnification of three and 10. Dashed black lines show the 1σ error.

5 DISCUSSION

We now seek to draw some general lessons from our examination of various model choices. While the HFF program and subsequent surveys provided a wealth of data on these two fields, studies of other clusters may have (much) less extensive data. Identifying choices that yield reliable results may therefore be helpful in guiding observing and modelling programs for cluster lensing.

We find that, for both fields, a magnitude limit of 22.5 mag and radial cut of 90 arcsec are lower limits past which the lens models become very similar, both in terms of model parameters and magnifications. The magnitude limit is a relatively robust result, but we do note that the radial limit likely depends on the extent of the lensed images; for Abell 2744, which has images in a smaller area, a smaller radial cut could be used.

We can draw such conclusions about parameters determining cluster member population as it is logical to assume that including all possible galaxies would give the best description of the true mass distribution. Drawing conclusions about the scaling relation and scatter parameters is not as straight forward since we do not know the true mass of each galaxy. However, our results show that there is a small bias introduced in the magnifications when a model with low η_t is compared to a model with a larger value.

In order to draw conclusions about the relative importance of each modelling choice, we consider the conditional probability distributions for the reasonable set of models. Specifically, we perform an analysis similar to what was done in Raney et al. (2020b). In this case, we do not compare models from different terms, but instead models with differing values for a given modelling choice.

We show the results of this analysis for Abell 2744 in Fig. 11. To create our distributions, we only include the models with $m_{\rm lim} \geq 22.5$ mag and $r_{\rm lim} = 90$ arcsec, corresponding to the values we determined to be reasonable choices. The violins with unfilled bodies show the statistical errors, i.e. the results of the self-comparison analysis, while those with filled bodies compare the results from each model in that group to all of the other models in that group, and thus correspond to a systematic error.

We see that in the $\mu = 3$ panel, the unfilled and filled violins for the three choices except galaxy scatter are all very similar; this indicates that there is no systematic error introduced in the magnifications if two models with varying model choices in a given group are compared. This trend continues in the $\mu = 10$ panel for scaling relation, but the difference between systematic and statistical error grows for magnitude limit and radial cut, suggesting that systematic errors may arise at larger magnifications due to these modelling choices.

The largest difference between statistical and systematic errors is seen with galaxy scatter. This is unsurprising given the results shown in Fig. 10, where we saw that statistical error for the models without LOS galaxy scatter was much smaller than those with the scatter, although the scatter when compared to the fiducial model was high. This is a somewhat surprising result given that we only include six LOS galaxies in our model, yet scatter in the LOS galaxies seems to be much more important than scatter within the cluster member galaxies. This is likely due to the placement of the LOS galaxies: two are in close proximity to more than 10 images, which also happen to be all of the constraints we use in the northern part of the field.

The results of this analysis are shown for MACS 0416 in Fig. 12. We see that the medians of the distributions are closer to the expected values of three or ten, which is expected given the previous results showing higher precision and accuracy than what was seen in Abell 2744. We also see that the distributions themselves are smaller, along with the 1σ error bars, which is in line with Fig. 10 where the scatter in magnifications is smaller even for the fiducial case. While we once again see that some modelling choices produce higher systematic error, specifically at $\mu = 10$, the differences are quite small and thus we conclude that none of the modelling choices introduce particularly important systematic errors into the magnifications.

In this work we do not endeavor to explain the differences in results between Abell 2744 and MACS 0416, but we will briefly offer our thoughts. The most obvious explanation as to why the magnifications errors in MACS 0416 are smaller than those of Abell 2744 would be

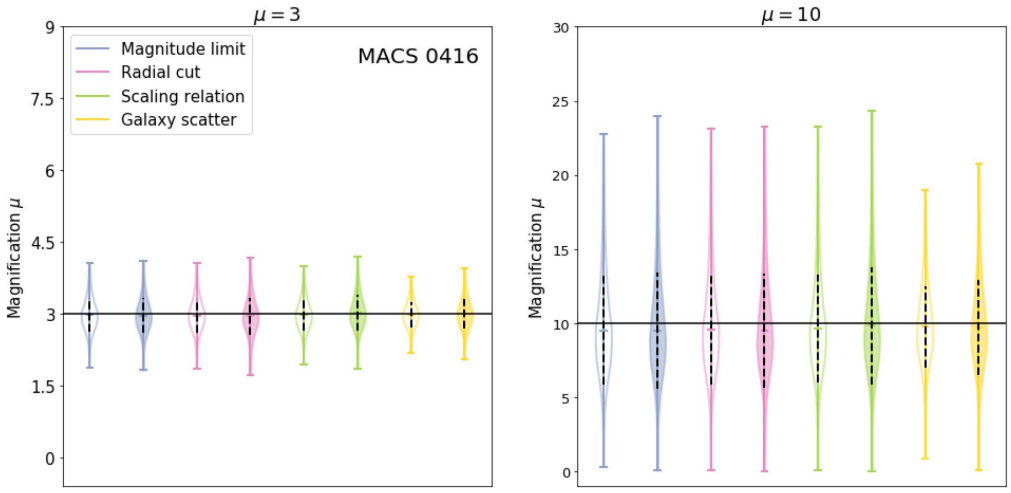

Figure 12. Similar to Fig. 11 for MACS 0416.

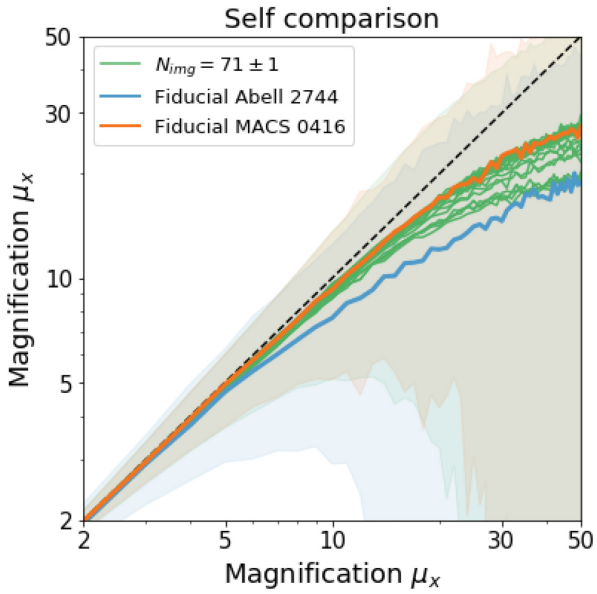

Figure 13. Medians and 1σ scatter of conditional probability distributions created by comparing realizations of a given model with itself. Here, we include the fiducial models of Abell 2744 (blue) and MACS 0416 (orange), along with 20 models of MACS 0416 created by subsetting the image constraints such that the number of images used is consistent with the number used to constrain Abell 2744 (green).

that the number of images used as constraints is higher: recall, we use $N_{\rm img} = 71$ for Abell 2744 and $N_{\rm img} = 95$ for MACS 0416.

We explore this by using only a subset of the fiducial constraints such that the number of images in the subset is 71 ± 1. We then rerun our analysis and calculate the magnifications. We show the self-comparisons for 20 different subsets in Fig. 13, along with the results from our fiducial models of Abell 2744 and MACS 0416. We see that the bulk of the models do appear to have curves similar to the fiducial analysis for MACS 0416, with only a few deviating to have similar values as Abell 2744. However, even models that do have similar medians at higher magnifications show a higher precisions at magnifications from $\mu = 6$–30.

Thus, it does not seem likely that the differences between the two fields are driven simply by the number of image constraints. It is possible that the arrangement of the images may be a factor. As we saw previously, Abell 2744 is more sensitive to scatter in the LOS galaxies likely because all of the image sets in the northern region have images near LOS galaxies. MACS 0416 has images fairly evenly spread across the field, and thus does not have such a configuration, which also may be why getting rid of 25 or so images is less impactful on the model. Zimmerman et al. (2021) take a closer look at how the choice of image constraints affects lens models.

6 CONCLUSIONS

In this work, we aimed to analyse some of the possible modelling choices that could be causing the systematic errors in magnifications we found in Raney et al. (2020b). In order to do this, we presented a modelling methodology using a Bayesian framework with which we analysed the effects of four different modelling choices: two affecting cluster member selection (magnitude limit and radial cut), and two affecting how mass is assigned to galaxies (scaling relation slopes and scatter).

For the cluster member selection, we found that there are clear differences in model parameters and magnifications if certain values are chosen. Specifically, we saw that a model with a smaller number of galaxies, e.g. a bright magnitude limit or very small radial cut, will produce magnifications that are very precise but cannot accurately predict magnifications of a model with a higher magnitude or radial limit. We found that for $m_{\rm lim} \geq 22.5$ mag and $r_{\rm lim} \geq 90$ arcsec, models had similar parameter and magnification distributions, although we note the radial limit likely depends on the extent of the images used as constraints.

For the scaling relation slopes, we found that model parameters may show significant differences as both η_E and η_t increase, but this results in only a modest change in magnifications. Unlike the other model choices, we found that the results for the galaxy scatter were quite different between the two fields. Specifically, scatter in the LOS galaxies greatly increased the scatter seen in the magnifications of Abell 2744. In MACS 0416, only a small change in model precision was seen when galaxy scatter was varied, and model accuracy did not change at all.

In general, our results show that MACS 0416 is better constrained than Abell 2744; this is seen in the smaller Δ^2 values and more accurate and precise magnifications. The differences are not fully explained by the fact that MACS 0416 has more image constraints than Abell 2744. It appears that each cluster lens field has some unique properties based on the configuration of images and the complexity of the mass distribution. Thus, our new framework for understanding systematic errors in cluster lens models allows us not only to understand the relative importance of different modelling choices, but also to identify the fields that are most robust to systematic effects.

ACKNOWLEDGEMENTS

We thank everyone who has contributed data and analysis for the *Hubble* Frontier Fields program. We thank Alyson Brooks, Eva Halkiadakis, Jack Hughes, and Keren Sharon for comments on the manuscript. We acknowledge support from the US National Science Foundation through grant AST-1909217.

DATA AVAILABILITY

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

REFERENCES

Abell G. O., Corwin H. G., Jr, Olowin R. P., 1989, ApJS, 70, 1
Baringhaus L., Franz C., 2004, J. Multivariate Anal., 88, 190
Bergamini P. et al., 2019, A&A, 631, A130
Bouwens R. J. et al., 2014, ApJ, 795, 126
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Brimioulle F., Seitz S., Lerchster M., Bender R., Snigula J., 2013, MNRAS, 432, 1046
Caminha G. B. et al., 2017, A&A, 600, A90
Coe D. et al., 2012, ApJ, 757, 22
Coe D. et al., 2013, ApJ, 762, 32
Ebeling H., Edge A. C., Henry J. P., 2001, ApJ, 553, 668
French K. D., Wong K. C., Zabludoff A. I., Ammons S. M., Keeton C. R., Angulo R. E., 2014, ApJ, 785, 59
Johnson T. L., Sharon K., 2016, ApJ, 832, 82
Johnson T. L., Sharon K., Bayliss M. B., Gladders M. D., Coe D., Ebeling H., 2014, ApJ, 797, 48
Kawamata R., Oguri M., Ishigaki M., Shimasaku K., Ouchi M., 2016, ApJ, 819, 114
Keeton C. R., 2010, Gen. Relat. Gravit., 42, 2151
Kempner J. C., David L. P., 2004, MNRAS, 349, 385
Kneib J.-P., Natarajan P., 2011, A&A Rev., 19, 47
Lotz J. M. et al., 2017, ApJ, 837, 97
Mahler G. et al., 2018, MNRAS, 473, 663
Mann A. W., Ebeling H., 2012, MNRAS, 420, 2120
McLeod D. J., McLure R. J., Dunlop J. S., 2016, MNRAS, 459, 3812
Meneghetti M. et al., 2017, MNRAS, 472, 3177
Merten J. et al., 2011, MNRAS, 417, 333
Monna A. et al., 2017, MNRAS, 465, 4589
Owers M. S., Randall S. W., Nulsen P. E. J., Couch W. J., David L. P., Kempner J. C., 2011, ApJ, 728, 27
Priewe J., Williams L. L. R., Liesenborgs J., Coe D., Rodney S. A., 2017, MNRAS, 465, 1030
Raney C. A., Keeton C. R., Brennan S., 2020a, MNRAS, 492, 503
Raney C. A., Keeton C. R., Brennan S., Fan H., 2020b, MNRAS, 494, 4771
Remolina González J. D., Sharon K., Mahler G., 2018, ApJ, 863, 60
Richard J. et al., 2014, MNRAS, 444, 268
Salmon B. et al., 2018, ApJ, 864, L22
Salmon B. et al., 2020, ApJ, 889, 189
Wong K. C., Ammons S. M., Keeton C. R., Zabludoff A. I., 2012, ApJ, 752, 104
Zheng W. et al., 2012, Nature, 489, 406
Zimmerman D., Keeton C. R., Raney C. A., 2021, MNRAS, Available at: https://doi.org/10.1093/mnras/stab2858

APPENDIX A: DERIVATION OF POSTERIOR PROBABILITY DISTRIBUTION

Let us begin with the lens equation relating a source position $\boldsymbol{y}$ with the image positions $\boldsymbol{x}$ and the deflection $\boldsymbol{\alpha}$ at those positions

$$\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{\alpha}_{\rm clus}(\boldsymbol{x}_i - \boldsymbol{\alpha}_{\rm fg}(\boldsymbol{x}_i)) - \boldsymbol{\alpha}_{\rm los}(\boldsymbol{x}_i), \tag{A1}$$

where $\boldsymbol{\alpha}_{\rm clus}$ represents the net deflection from all of the mass in the cluster. We note that the differing treatments of foreground versus total LOS mass is seen here, where only the foreground mass impacts the positions at which the cluster deflection must be calculated.

Equation (A1) fully describes the lensing along a 3D line of sight towards a source. However, we can simplify this equation by assuming that the foreground deflection is small when compared to that of the cluster. In doing so, we can make a Taylor series expansion:

$$\boldsymbol{\alpha}_{\rm clus}(\boldsymbol{x}_i - \boldsymbol{\alpha}_{\rm fg}(\boldsymbol{x}_i)) \approx \boldsymbol{\alpha}_{\rm clus}(\boldsymbol{x}_i) - \boldsymbol{\Gamma}_{{\rm clus},i}\boldsymbol{\alpha}_{\rm fg}(\boldsymbol{x}_i), \tag{A2}$$

where $\boldsymbol{\Gamma}_{{\rm clus},i} = \partial\boldsymbol{\alpha}_{\rm clus}/\partial\boldsymbol{x}_i$ is a 2×2 matrix. Using this, as well as expanding $\boldsymbol{\alpha}_{\rm clus}$ into $\boldsymbol{\alpha}_{\rm halo}$ and $\boldsymbol{\alpha}_{\rm mem}$, then gives

$$\boldsymbol{y} = \boldsymbol{x}_i - [\boldsymbol{\alpha}_{\rm halo}(\boldsymbol{x}_i) + \boldsymbol{\alpha}_{\rm mem}(\boldsymbol{x}_i) + \boldsymbol{\alpha}_{\rm los}(\boldsymbol{x}_i) - \boldsymbol{\Gamma}_{{\rm clus},i}\boldsymbol{\alpha}_{\rm fg}(\boldsymbol{x}_i)]. \tag{A3}$$

The likelihood function for data given the model can then be written as (see Keeton 2010)

$$\mathscr{L} = |\boldsymbol{C}_{\rm obs}|^{-1/2} \exp\left(-\frac{1}{2}\sum_{\{s,i\}}[\boldsymbol{x}_{\rm i} - (\boldsymbol{\alpha}_{\rm halo} + \boldsymbol{\alpha}_{\rm mem} + \boldsymbol{\alpha}_{\rm los} - \boldsymbol{\Gamma}_{\rm clus,i}\boldsymbol{\alpha}_{\rm fg}) - \boldsymbol{y}_{\rm s}]^{\rm T}\boldsymbol{\mu}_{\rm i}^{\rm T}\boldsymbol{C}_{\rm obs}^{-1}\boldsymbol{\mu}_{\rm i}^{\rm T}[\boldsymbol{x}_{\rm i} - (\boldsymbol{\alpha}_{\rm halo} + \boldsymbol{\alpha}_{\rm mem} + \boldsymbol{\alpha}_{\rm los} - \boldsymbol{\Gamma}_{\rm clus,i}\boldsymbol{\alpha}_{\rm fg}) - \boldsymbol{y}_{\rm s}]\right), \tag{A4}$$

where we include the measurement uncertainty for each image i in $\mathbf{C}_{\rm obs}$. We note the sum uses $\{s, i\}$ to denote that we keep track of which images i go with which source s. The magnification tensor $\boldsymbol{\mu}_i$ is used to convert between source and lens planes for a given image i. We also keep track of determinant factors that are important and thus include $|\mathbf{C}_{\rm obs}|^{-1/2}$ for the normalization, but omit factors of 2π for simplicity.

Thus, we see that our likelihood function is Gaussian in our previously discussed random variables, i.e. $\boldsymbol{\alpha}_{\rm mem}$, $\boldsymbol{\alpha}_{\rm los}$, and $\boldsymbol{\alpha}_{\rm fg}$, as well as in the source positions. We can collect these into a block vector as

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{\alpha}_{\rm mem} \\ \boldsymbol{\alpha}_{\rm los} \\ \boldsymbol{\alpha}_{\rm fg} \\ \boldsymbol{y} \end{bmatrix}. \tag{A5}$$

We also define the helper matrix $\boldsymbol{A}$ such that

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} & -\boldsymbol{\Gamma}_{\rm clus} & \boldsymbol{U} \end{bmatrix}, \tag{A6}$$

where $\boldsymbol{I}$ is an identity matrix with dimensions $2N_{\rm img} \times 2N_{\rm img}$ and $N_{\rm img}$ is the number of images. The matrix $\boldsymbol{U}$ is used to connect which images go with which sources. For example, if the constraints consisted of two sources, the first with two images and the second with three, $\boldsymbol{U}$ would take the form

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{I}_2 & \boldsymbol{0} \\ \boldsymbol{I}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_2 \\ \boldsymbol{0} & \boldsymbol{I}_2 \\ \boldsymbol{0} & \boldsymbol{I}_2 \end{bmatrix}. \tag{A7}$$

Here we use $\boldsymbol{I}_2$ to denote the 2×2 identity matrix. With these definitions, we can rewrite our likelihood in a more compact form:

$$\mathscr{L} = |\boldsymbol{C}_{obs}|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\alpha}_{\rm halo}-\boldsymbol{A}\boldsymbol{v})^T \boldsymbol{\mu}^T \boldsymbol{C}_{\rm obs}^{-1} \boldsymbol{\mu}(\boldsymbol{x}-\boldsymbol{\alpha}_{\rm halo}-\boldsymbol{A}\boldsymbol{v})}. \tag{A8}$$

Next, we must consider the pertinent priors, specifically for $\boldsymbol{\alpha}_{\rm mem}$, $\boldsymbol{\alpha}_{\rm los}$, $\boldsymbol{\alpha}_{\rm fg}$, and $\boldsymbol{y}$. We assume that the deflections are Gaussian, although we note that this would not necessarily hold if an image lies very close to a member galaxy and is thus more heavily influenced by it. We show examples of deflection distributions for a set of images in Section 3.

If the cluster member deflections are indeed Gaussian, then they can be described by a mean vector $\bar{\boldsymbol{a}}_{\rm mem}$ and covariance matrix $\boldsymbol{C}_{\rm mem}$. Similarly, $\boldsymbol{\alpha}_{\rm los}$ and $\boldsymbol{\alpha}_{\rm fg}$ can also be described in this manner, although we must allow for cross terms in the covariance matrices. For the source positions $\boldsymbol{y}$, we assume Gaussian priors but with a standard deviation σ_y large enough such that the priors are effectively uninformative. Thus, the priors can be written as

$$\mathscr{P} = |\boldsymbol{C}_{\rm pri}|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{v}-\bar{\boldsymbol{a}})^T \boldsymbol{C}_{\rm pri}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{a}})}, \tag{A9}$$

where

$$\bar{\boldsymbol{a}} = \begin{bmatrix} \bar{\boldsymbol{a}}_{\rm mem} \\ \bar{\boldsymbol{a}}_{\rm los} \\ \bar{\boldsymbol{a}}_{\rm fg} \\ \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{C}_{\rm pri} = \begin{bmatrix} \boldsymbol{C}_{\rm mem} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{\rm los} & \boldsymbol{C}_{\rm los,fg} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_{\rm los,fg}^T & \boldsymbol{C}_{\rm fg} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \sigma_y^2 \boldsymbol{I} \end{bmatrix}, \tag{A10}$$

and we have once again dropped any factors of 2π.

The important takeaway is that the full posterior $P \propto \mathscr{L}\mathscr{P}$ is still Gaussian in our nuisance parameters $\boldsymbol{v}$:

$$P \propto |\boldsymbol{C}_{\rm obs}|^{-1/2} |\boldsymbol{C}_{\rm pri}|^{-1/2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\alpha}_{\rm halo}-\boldsymbol{A}\boldsymbol{v})^T \boldsymbol{\mu}^T \boldsymbol{C}_{\rm obs}^{-1} \boldsymbol{\mu}(\boldsymbol{x}-\boldsymbol{\alpha}_{\rm halo}-\boldsymbol{A}\boldsymbol{v}) - \frac{1}{2}(\boldsymbol{v}-\bar{\boldsymbol{a}})^T \boldsymbol{C}_{\rm pri}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{a}})}. \tag{A11}$$

This is useful because we can then marginalize over the nuisance parameters using standard Gaussian integration. In order to do so, we define

$$\begin{aligned} \boldsymbol{d} &= \boldsymbol{x} - \boldsymbol{\alpha}_{\rm halo} - \boldsymbol{A}\bar{\boldsymbol{a}} \\ &= \boldsymbol{x} - \boldsymbol{\alpha}_{\rm halo} - \bar{\boldsymbol{a}}_{\rm mem} - \bar{\boldsymbol{a}}_{\rm los} + \boldsymbol{\Gamma}_{\rm halo} \bar{\boldsymbol{a}}_{\rm fg}, \end{aligned} \tag{A12}$$

and

$$\begin{aligned} \boldsymbol{C}_{\rm marg} &= \boldsymbol{\mu}^{-1} \boldsymbol{C}_{\rm obs} \boldsymbol{\mu}^{-T} + \boldsymbol{A} \boldsymbol{C}_{\rm pri} \boldsymbol{A}^T \\ &= \boldsymbol{\mu}^{-1} \boldsymbol{C}_{\rm obs} \boldsymbol{\mu}^{-T} + \boldsymbol{C}_{\rm mem} + \boldsymbol{C}_{\rm los} - \boldsymbol{\Gamma}_{\rm clus} \boldsymbol{C}_{\rm los,fg}^T - \boldsymbol{C}_{\rm los,fg} \boldsymbol{\Gamma}_{\rm clus}^T + \boldsymbol{\Gamma}_{\rm clus} \boldsymbol{C}_{\rm fg} \boldsymbol{\Gamma}_{\rm clus}^T + \sigma_y^2 \boldsymbol{U}\boldsymbol{U}^T. \end{aligned} \tag{A13}$$

Then our marginalized posterior distribution $P_{\rm marg} = \int \mathscr{L}\mathscr{P} {\rm d}\boldsymbol{v}$ is given as

$$P_{\rm marg}(\boldsymbol{q}) \propto |\boldsymbol{\mu}|^{-1} |\boldsymbol{C}_{\rm marg}|^{-1/2} e^{-\frac{1}{2}\boldsymbol{d}^T \boldsymbol{C}_{\rm marg}^{-1} \boldsymbol{d}}. \tag{A14}$$

We note that formally we allow $\sigma_y \to \infty$ so that the priors on the source positions are uniform over an infinite domain. This may seem to pose a problem in calculating $\boldsymbol{C}_{\rm marg}$, specifically in the last term of equation (A13). However, the inverse as needed for equation (A14) is still well defined. We can use the Woodbury matrix identity to write

$$\boldsymbol{C}_{\rm marg}^{-1}\Big|_{\sigma_y \to \infty} = \boldsymbol{S}_{\rm eff} - \boldsymbol{S}_{\rm eff} \boldsymbol{U} (\boldsymbol{U}^T \boldsymbol{S}_{\rm eff} \boldsymbol{U})^{-1} \boldsymbol{U}^T \boldsymbol{S}_{\rm eff}, \tag{A15}$$

where

$$\boldsymbol{S}_{\rm eff}^{-1} = \boldsymbol{\mu}^{-1} \boldsymbol{C}_{\rm obs} \boldsymbol{\mu}^{-T} + \boldsymbol{C}_{\rm mem} + \boldsymbol{C}_{\rm los} - \boldsymbol{\Gamma}_{\rm clus} \boldsymbol{C}_{\rm los,fg}^T - \boldsymbol{C}_{\rm los,fg} \boldsymbol{\Gamma}_{\rm clus}^T + \boldsymbol{\Gamma}_{\rm clus} \boldsymbol{C}_{\rm fg} \boldsymbol{\Gamma}_{\rm clus}^T. \tag{A16}$$

The determinant $|\boldsymbol{C}_{\rm marg}|$ has asymptotic behaviour when $\sigma \to \infty$:

$$\ln|\boldsymbol{C}_{\rm marg}| \approx \ln|\boldsymbol{U}^T\boldsymbol{S}_{\rm eff}\boldsymbol{U}| - \ln|\boldsymbol{S}_{\rm eff}| + 4N_{\rm src}\ln\sigma_y, \tag{A17}$$

where the last term is such because the priors have σ_y^2 terms for each spatial dimension. Since we either normalize the posterior or consider relative posteriors (i.e. when comparing models), we can omit the last term.

APPENDIX B: SIMPLIFYING THE DISTANCE METRIC

Formally, the distance metric Δ^2 defined in equation (7) involves big double sums, but with some algebra it can be written in terms of moments of the distributions. To see this, let's write out the sums in D_X (omitting limits to simplify the notation):

$$\begin{aligned}
D_X^2 &= \frac{1}{n_X(n_X-1)}\sum_{i,j}\left(\boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_i + \boldsymbol{x}_j^T\boldsymbol{C}^{-1}\boldsymbol{x}_j - \boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_j - \boldsymbol{x}_j^T\boldsymbol{C}^{-1}\boldsymbol{x}_i\right)\\
&= \frac{2}{n_X-1}\sum_i \boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_i - \frac{2}{n_X(n_X-1)}\sum_{i,j}\boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_j\\
&= \frac{2}{n_X-1}\sum_i \boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_i - \frac{2}{n_X(n_X-1)}\left(\sum_i \boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_i + \sum_{i\neq j}\boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_j\right)\\
&= \frac{2}{n_X}\sum_i \boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_i - \frac{2}{n_X(n_X-1)}\sum_{i\neq j}\boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_j\\
&= 2(\langle\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\rangle - \langle\boldsymbol{x}\rangle^T\boldsymbol{C}^{-1}\langle\boldsymbol{x}\rangle)
\end{aligned} \tag{B1}$$

In the first line, index manipulation shows that the first and second terms are identical, as are the third and fourth terms, leading to the second line. Then we separate the double sum into terms with $i=j$ and terms with $i\neq j$. In the third line, we can combine the sums in the first and second terms. To further process the last line, let's write out the first term in component notation, where indices a and b now run over the components:

$$\begin{aligned}
\left\langle\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right\rangle &= \left\langle\sum_{a,b}(\boldsymbol{C}^{-1})_{ab}x_a x_b\right\rangle\\
&= \sum_{a,b}(\boldsymbol{C}^{-1})_{ab}\,\langle x_a x_b\rangle\\
&= \sum_{a,b}(\boldsymbol{C}^{-1})_{ab}\,[\mu_a\mu_b + (\boldsymbol{C}_X)_{ab}]\\
&= \boldsymbol{m}_X^T\boldsymbol{C}^{-1}\boldsymbol{m}_X + {\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X)
\end{aligned} \tag{B2}$$

where $\boldsymbol{m}_X$ and $\mathbf{C}_X$ are, respectively, the mean vector and covariance matrix for distribution X. Combining equations (B1) and (B2) yields

$$D_X^2 = 2\,{\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X) \tag{B3}$$

We can similarly find $\mathrm{D}_Y^2 = 2\,{\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_Y)$. These expressions quantify an overall 'size' for each distribution, according to the metric $\mathbf{C}$. If we use Euclidean distances ($\boldsymbol{C}=\boldsymbol{I}$) then D_X^2 is twice the sum of the variances across all dimensions of X (and likewise for Y).

Now consider D_{XY}:

$$\begin{aligned}
D_{XY}^2 &= \frac{1}{n_X n_Y}\sum_{i=1}^{n_X}\sum_{j=1}^{n_Y}\left(\boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{x}_i + \boldsymbol{y}_j^T\boldsymbol{C}^{-1}\boldsymbol{y}_j - \boldsymbol{x}_i^T\boldsymbol{C}^{-1}\boldsymbol{y}_j - \boldsymbol{y}_j^T\boldsymbol{C}^{-1}\boldsymbol{x}_i\right)\\
&= \left\langle\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right\rangle + \left\langle\boldsymbol{y}^T\boldsymbol{C}^{-1}\boldsymbol{y}\right\rangle - \boldsymbol{m}_X^T\boldsymbol{C}^{-1}\boldsymbol{m}_Y - \boldsymbol{m}_Y^T\boldsymbol{C}^{-1}\boldsymbol{m}_X,
\end{aligned} \tag{B4}$$

where we use the fact that the two samples are independent in the cross terms. Now, we substitute for the first two terms using equation (B2) and the corresponding expression for Y, and collect terms to obtain

$$D_{XY}^2 = (\boldsymbol{m}_X-\boldsymbol{m}_Y)^T\boldsymbol{C}^{-1}(\boldsymbol{m}_X-\boldsymbol{m}_Y) + {\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X) + {\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_Y). \tag{B5}$$

Combining the pieces, we can write our test statistic as

$$\Delta^2 = D_{XY}^2 - D_X D_Y = (\boldsymbol{m}_X-\boldsymbol{m}_Y)^T\boldsymbol{C}^{-1}(\boldsymbol{m}_X-\boldsymbol{m}_Y) + {\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X) + {\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_Y) - 2\sqrt{{\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X)\,{\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_Y)}. \tag{B6}$$

We can combine the last three terms to obtain our final expression:

$$\Delta^2 = (\boldsymbol{m}_X-\boldsymbol{m}_Y)^T\boldsymbol{C}^{-1}(\boldsymbol{m}_X-\boldsymbol{m}_Y) + \left[\sqrt{{\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_X)} - \sqrt{{\rm tr}(\boldsymbol{C}^{-1}\boldsymbol{C}_Y)}\right]^2. \tag{B7}$$

The first term is identical to Hotelling's t^2 statistic, and it measures a shift in the means. The second term measures a difference in the sizes/shapes of the distributions.

Note: Baringhaus & Franz (2004) introduced a different multivariate two-sample test. They considered average rather than rms distances

$$\tilde{D}_X = \frac{1}{n_X(n_X-1)} \sum_{i=1}^{n_X} \sum_{j=1}^{n_X} ||\boldsymbol{x}_i - \boldsymbol{x}_j|| \qquad \text{(and likewise for } Y\text{)} \tag{B8}$$

$$\tilde{D}_{XY} = \frac{1}{n_X n_Y} \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} ||\boldsymbol{x}_i - \boldsymbol{y}_j|| \tag{B9}$$

and defined their test statistic as

$$T = \tilde{D}_{XY} - \frac{1}{2}(\tilde{D}_X + \tilde{D}_Y) \tag{B10}$$

Their definition used the Euclidean distance, but it can be generalized to use a Mahalanobis distance. As far as we can tell, the T statistic cannot be simplified to avoid the double sums. We favour the Δ^2 statistic because it gives a nice generalization of Hotelling's t^2 statistic that clearly accounts for differences in both the locations and shapes of the distributions, while only requiring first and second moments (which can be computed more rapidly than double sums).

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