



# An Evaluation of 10 Lensing Models of the Frontier Fields Cluster MACS J0416.1-2403

J. D. Remolina González , K. Sharon , and G. Mahler

Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA; [jremolin@umich.edu](mailto:jremolin@umich.edu)

Received 2018 April 24; revised 2018 June 22; accepted 2018 June 25; published 2018 August 9

## Abstract

Galaxy clusters can act as gravitational lenses and magnify the universe behind them, allowing us to see deep into the early universe. The *Hubble Space Telescope* Frontier Fields program uses six galaxy clusters imaged by *Hubble* to discover and study galaxies at  $z \sim 5\text{--}10$ . Seven independent teams developed lens models and derived magnifications for each galaxy cluster, based on positional and redshift constraints from the best available data at the time. In this work we evaluate 10 models for MACS J0416.1-2403 that were made public in 2015 by contrasting them with new spectroscopic redshifts that were measured in 2016. We developed an independent comparison method that uses the source plane root-mean-square as a metric of lensing model performance. Our analysis quantifies the ability of models to predict unknown multiple images. We examine the source plane scatter of multiply imaged systems and explore the dependence of the scatter on the location and the redshift of the background sources. The analysis we present evaluates the performance of the different algorithms in the specific case of the MACS J0416.1-2403 models.

**Key words:** galaxies: clusters: individual (MACS J0416.1-2403) – gravitational lensing: strong

## 1. Introduction

Strong gravitational lensing in massive galaxy clusters provides an opportunity to study both the physical properties of clusters and the highly magnified background universe behind them. Strong lensing analysis can directly inform estimates of the total (baryonic and dark matter) projected mass distributions of galaxy cluster cores (e.g., Richard et al. 2010a; Sharon et al. 2014, 2015; Zitrin & Broadhurst 2016; Monna et al. 2017; Mahler et al. 2018); the high magnifications associated with strong lensing clusters facilitate both observations of the high-redshift ( $z \sim 6\text{--}11$ ) magnified background universe (e.g., Kneib et al. 2004; Richard et al. 2008; Coe et al. 2013, 2015; Bouwens et al. 2014; Bradley et al. 2014; Atek et al. 2015; Salmon et al. 2017) and zoomed-in studies of highly magnified galaxies around  $z \sim 2$  (e.g., Jones et al. 2010; Bayliss et al. 2011; Livermore et al. 2015; Johnson et al. 2017a, 2017b). Strong lensing clusters have also been used to constrain cosmological parameters (e.g., Jullo et al. 2010; Cao et al. 2015; Magaña et al. 2015; Acebron et al. 2017). The *Hubble Space Telescope* (*HST*) has been transformative for strong lensing research, enabling most of the studies listed above.

The *Hubble* Frontier Fields (HFF; Lotz et al. 2017) targets six massive galaxy clusters that are used as cosmic telescopes to observe the background universe ( $z \sim 5\text{--}10$ ). Each of the galaxy clusters and parallel fields was observed with seven optical and near-infrared bands using the Wide Field Camera 3 (WFC3) and the Advanced Camera for Surveys (ACS) over 140 orbits each, for a total of 840 *HST* orbits. Follow-up observations from a variety of observatories both on the ground and in space have provided complementary data, including spectroscopic redshifts ( $z_{\text{spec}}$ ), which allowed for precise lens models and the magnification measurements needed to study the background universe (e.g., Oesch et al. 2015; Alavi et al. 2016; Laporte et al. 2016; McLeod et al. 2016; Bouwens et al. 2017; Livermore et al. 2017; Yue et al. 2017; Atek et al. 2018; Kawamata et al. 2018).<sup>1</sup>

The Space Telescope Science Institute (STScI) contracted several lens modeling teams to compute lens models for each of

the HFF clusters and make them available to the public. As part of this effort, the different teams collaboratively evaluated the robustness of the multiply imaged systems that were found in these data. Other than agreeing on a uniform set of constraints, each team worked independently from the other teams. In that, the HFF initiative presents us with a unique opportunity to compare different algorithms that were used to model the same cosmic lenses, the various assumptions used by the lensing teams, and to learn about systematics in lens modeling and the strengths of different algorithms. Comparison projects using simulated galaxy clusters (Acebron et al. 2017; Meneghetti et al. 2017) and HFF clusters (Prieve et al. 2017) investigate the performance of the algorithms employed by the lensing modelers. Other projects contrasted lens model predictions with observations from strongly lensed supernovae (Rodney et al. 2015; Kelly et al. 2016; Rodney et al. 2016; Treu et al. 2016).

Since the beginning of the HFF program, several versions of the lens models have been computed by each team. With each new version the quantity of constraints has increased due to the deep *HST* observations, allowing for identification of faint multiply imaged systems of background sources. In addition, the quality of the constraints has improved with spectroscopic follow-up providing confirmation of the arcs, and importantly, spectroscopic redshifts (e.g., Jauzac et al. 2014; Richard et al. 2014; Treu et al. 2015; Caminha et al. 2017; Karman et al. 2017; Mahler et al. 2018).

The goal of this paper is to examine the predictive power of strong lensing models. We evaluate the 10 public HFF version 3 lens models for MACS J0416.1-2403, taking advantage of newly measured spectroscopic redshifts (Caminha et al. 2017) that were not available at the time the lens models were computed. We investigate how the different algorithms perform in regions of the field of view where there were no constraints when the lens models were computed. We also check for any dependence of the root-mean-square (rms) scatter of the systems on the source spectroscopic redshifts. With the new data we can explore the predictive power of the lens models through the source plane scatter and image plane projections. This is important for future predictions made using these

<sup>1</sup> See <http://www.stsci.edu/hst/campaigns/frontier-fields/Publications> for a more comprehensive list of references and publications.

models, including the location of new multiple images of known or newly discovered lensed sources.

This paper is organized as follows. The importance of source redshifts in strong gravitational lens models, MACS J0416.1-2403, and the lens models available from HFF are described in Section 2. In Section 3, we describe the lensing equation and the complexities of using the image plane scatter. Next, we discuss the methodology and results from the source plane and image plane projection analysis. Section 4 examines the V4 models for MACS J0416.1-2403, which used the newly measured redshifts as constraints. A summary and conclusions are provided in Section 5. We assume a flat cosmology with  $\Omega_M = 0.3$ ,  $\Omega_\Lambda = 0.7$ , and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2. Background

### 2.1. Strong Lensing Constraints

Strong lensing analysis relies on correct identification of sets of multiply imaged lensed sources (“arcs”). Modelers use several selection criteria to identify multiple images, including morphology, color, redshift, and expected lensing geometry. Deep, high-resolution observations increase the feasibility of finding arc candidates but come with an extensive cost in observational resources. These observations are crucial since both the accuracy and precision of the lens models improve with the use of more lensed galaxies, as has been shown by Richard et al. (2010b), Johnson & Sharon (2016), and Mahler et al. (2018).

The main constraints in strong lensing come from the observed positions of these images. Additionally, the redshift of the lensed source provides the distance (assuming a cosmology), giving us three-dimensional information on the geometry of the system (Equation (1)). Ideally, all of the multiply imaged systems that are used to constrain the model would have spectroscopic redshifts. This is not typically the case, due to the high observational cost of spectroscopic follow-up. As the community devotes more resources to spectroscopic follow-up observations, the data become more complete. In the cases in which spectroscopic redshifts are not available, lens modeling algorithms can treat redshifts as free parameters, increasing the flexibility of the models and thus increasing their uncertainty.

In the absence of spectroscopic redshifts, photometric redshifts can be used in the modeling process. Photometric redshifts are prone to systematic uncertainties, which can be reduced with information from multiple band-passes and deeper data, although catastrophic failures are possible. Nevertheless, photometric measurements can be implemented in the modeling process by using their posterior probability distributions as a prior for the free-redshift parameter of the lensed source.

An investigation of the systematics in strong gravitational lens modeling exploring the importance of using spectroscopic redshifts was done by Johnson & Sharon (2016), who found that the rms scatter in the image plane improves when increasing the number of spectroscopic systems used in computing the lens models. Cerny et al. (2018) investigated methods for treating the redshifts of multiple images in various galaxy clusters and the effects on the magnifications computed from the lens models. The systematic uncertainties due to a small fraction of spectroscopic redshifts, compared to a large fraction, were analyzed in the case of Abell 2744 (Mahler et al. 2018), which was observed as part of the HFF program.

The systematic effect on the retrieval of cosmological parameters from strong lensing models due to the different density profiles used by parametric lens models and the effect of redshifts were explored by Acebron et al. (2017) using simulated galaxy clusters (Meneghetti et al. 2017).

In this work we take advantage of the factor of two increase in the number of lensing constraints with spectroscopically confirmed redshifts to evaluate lens models of the HFF cluster MACS J0416.1-2403.

### 2.2. The Predictive Power of Lensing Models in Clusters of Galaxies

All strong lens models rely on lensing observables as constraints. A variety of predictions can be made from the computed models, including the locations of new multiple images of known or newly discovered lensed sources. The use of lens models to find background-source projections is standard practice by lens modelers as well as the rest of the community, particularly for the purpose of discovering new images of lensed background sources, which in turn can help provide new constraints for the lens models.

With a well-constrained lens model, the angular separation between images of the same source can be used to estimate “geometric” redshifts in the absence of spectroscopic redshift (e.g., Zitrin et al. 2014; Chan et al. 2017). In addition, model-predicted redshifts can be compared to independent photometric redshift estimates and help discriminate between degenerate photometric redshifts (e.g., Lam et al. 2014; Diego et al. 2015a; Cerny et al. 2018).

Time delay can also be predicted by lens models. Since the travel time of light depends on its path between the source and the observer, different images of the same source are not contemporaneous. The arrival time difference, or time delay, between multiple images of the same source depends on the lensing potential and can be directly derived from a lens model. If the source is transient or variable in nature, the time delay can be observationally measured and compared to lens model predictions, e.g., supernovae (Kelly et al. 2014; Oguri 2015; Sharon & Johnson 2015; Diego et al. 2016a; Jauzac et al. 2016; Kelly et al. 2016; Treu et al. 2016; Rodney et al. 2018) and quasars lensed by clusters (Inada et al. 2006; Fohlmeister et al. 2007, 2013; Dahle et al. 2013; Sharon et al. 2017).

Finally, the lensing magnification is seldom a direct observable. In most cases, one can predict and measure the relative lensing magnification between images of the same source. The only known exception thus far is observed standard candles such as Type Ia supernovae, whose peak magnitudes can be known through calibration and thus their absolute magnification can be compared to model predictions (Rodney et al. 2015).

In this paper, we focus on the ability of lens models to properly predict the source locations of lensed images. For assessment of the predicting power of magnification or time delay, the reader is referred to the publications listed above.

### 2.3. MACS J0416.1-2403

MACS J0416.1-2403 is a massive cluster at a redshift  $z = 0.397$  identified in the MASSive Cluster Survey (MACS Ebeling et al. 2007, 2014). It was suggested that it is a merging cluster (Mann & Ebeling 2012) due to the double-peak profile of its bolometric X-ray luminosity. The Cluster Lensing and

**Table 1**Number of Arcs with  $z_{\text{spec}}$  for MACS J0416.1-2403

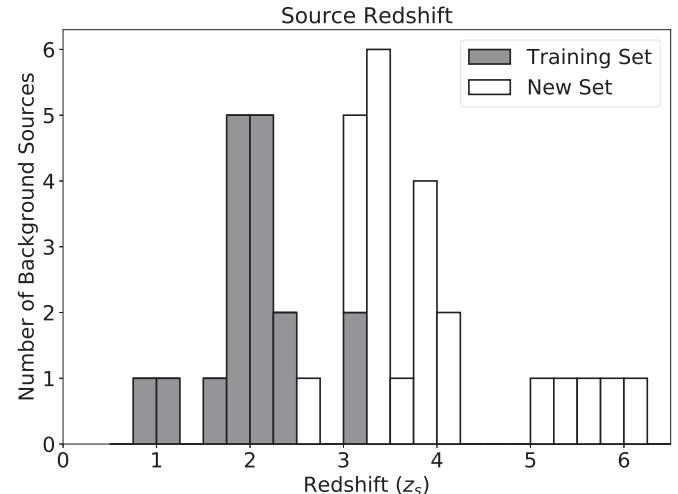
Sets	$N_{\text{sys}}$	$N_{\text{im}}$
Training	17	51
New	22	57
Total	39	108

**Note.** Number of multiple images ( $N_{\text{im}}$ ) and background systems ( $N_{\text{sys}}$ ) with spectroscopic redshifts ( $z_{\text{spec}}$ ) used in our analysis.

Supernova survey with *Hubble* (CLASH) has studied this cluster to constrain its mass distribution (Postman et al. 2012). Using the CLASH *HST* images, Zitrin et al. (2013) identified 70 multiple images of 23 background sources, which were used in the strong lensing analysis. This cluster was previously observed by *Spitzer* IRAC (PI R. Bouwens), *Chandra Space Observatory*, Jansky Very Large Array, Giant Meterwave Radio Telescope (Ogurac et al. 2015), and Very Large Telescope spectroscopy (e.g., Grillo et al. 2015; Balestra et al. 2016). As part of the HFF program, MACS J0416.1-2403 was observed again by *HST* in early 2014 and late 2014. Some of the most recent work exploring MACS J0416.1-2403 includes joint X-ray and strong lensing analysis to investigate the dark matter distribution in the core of the galaxy cluster (Bonamigo et al. 2017). Line-of-sight structure is relevant when performing strong lensing analysis because the lensing signal is sensitive to the total projected mass distribution (Bayliss et al. 2014; D’Aloisio et al. 2014). A study of line-of-sight structure effects on the strong lens modeling of MACS J0416.1-2403 was presented by Chirivì et al. (2018). Natarajan et al. (2017) studied the dark matter substructure of MACS J0416.1-2403 by comparing lensing -inferred substructure to that of simulations.

With the deep *HST* imaging, hundreds of multiple images have been identified and spectroscopic redshifts have been measured in follow-up observations. In 2015, seven independent lensing teams computed lens models for MACS J0416.1-2403 with the positional and redshift constraints available at that time. This included 17 sources with confirmed  $z_{\text{spec}}$  (Jauzac et al. 2014; Richard et al. 2014; Treu et al. 2015). In this analysis, we refer to the set of lensed images with spectroscopic redshifts available when the lens models were computed as the “training set.” In 2016, a new follow-up survey by Caminha et al. (2017) increased the number of spectroscopic redshifts to 39 sources and updated some of the previously found redshifts. The lensed images (arcs) with new spectroscopic redshifts are referred to as “new set” in our analysis. Table 1 has a summary of the number of background sources and the multiple images with spectroscopic redshifts for MACS J0416.1-2403. The distribution of the source spectroscopic redshifts ( $z_s$ ) is shown in Figure 1. We note that most of the new arcs have a higher redshift compared to the training set. Out of the spectroscopic redshifts that were updated, only system 26 had a drastic change in  $z_{\text{spec}}$  from  $z_{26} = 2.1851$  to  $z_{26} = 3.2355$  (see Caminha et al. 2017). In Figure 2, we overplot the positions of the arcs on a WFC/IR (F160W) image of MACS J0416.1-2403. The colors represent the source spectroscopic redshift, the symbols indicate if the lensed system is part of the training (plus) or new (diamond) set, and the box identifies the location of system 26.

The independent lensing teams have ranked the lensed systems based on the confidence associated with the identification of the multiple images, the availability of the spectroscopic redshift for



**Figure 1.** Redshift distribution of the background sources that are lensed by the galaxy cluster MACS J0416.1-2403. The gray area represents the training set (data were available to the lensing groups when computing the models in 2015 from Jauzac et al. 2014; Richard et al. 2014; Treu et al. 2015) and the white is the new set (data were available from the recent follow-up survey by Caminha et al. 2017 that did not have any previous archival spectroscopic redshifts). From the distribution, it is clear that the new arcs tend to have a higher redshift than the arcs in the training set. This color convention is used throughout the paper.

the source, and lensing geometry. The arcs that had the highest reliability and measured spectroscopic redshifts were labeled “GOLD.” Secure arcs with no available  $z_{\text{spec}}$  were labeled “SILVER.” Lastly, arcs labeled “BRONZE” were the multiply imaged systems where there was neither unanimous agreement on the reliability nor spectroscopic redshift confirmation. The lensing teams used these vetted selections to compute the lens models. We refer to these classifications in this paper.

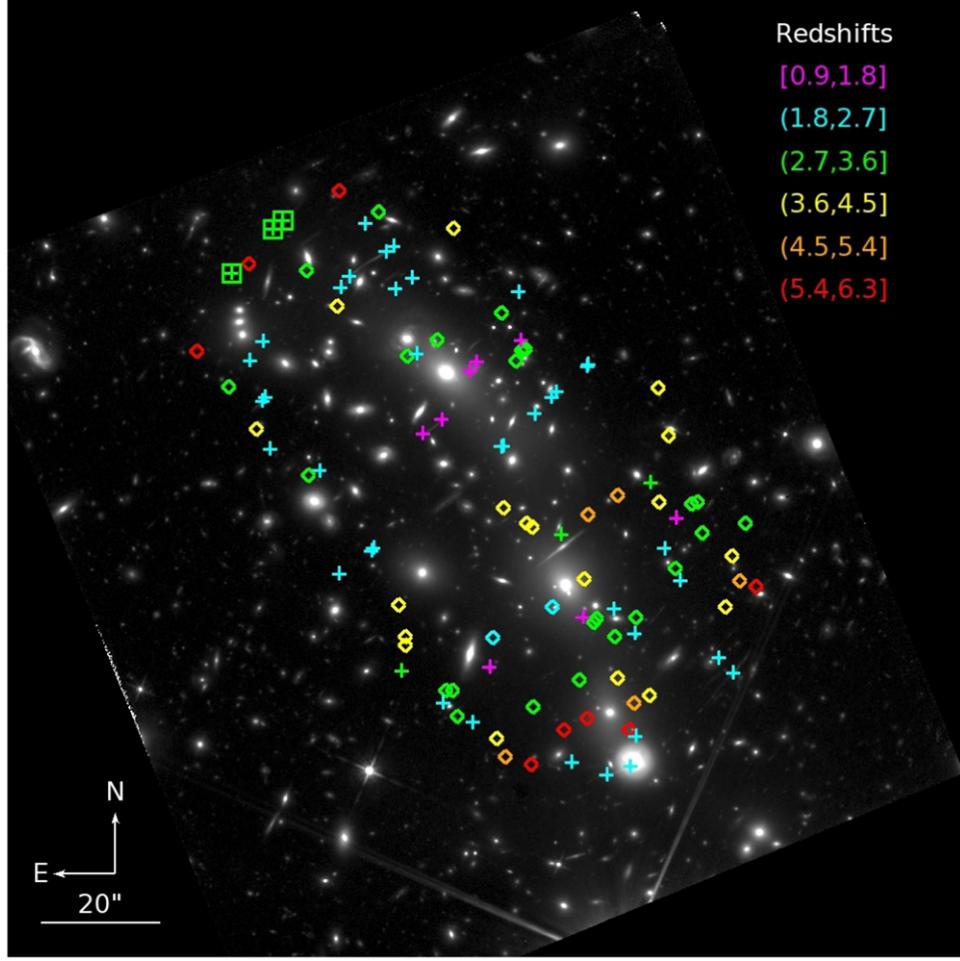
#### 2.4. Lensing Models

The lensing groups contracted by STScI computed independent lens models using a variety of algorithms, which can generally be separated into two categories: “parametric” and “non-parametric” (also called “free-form”). Parametric algorithms use a parameterized density distribution for dark matter halos and cluster-member galaxies to determine the lensing potential and mass distribution of the cluster. Non-parametric or free-form algorithms make little to no assumptions regarding the functional form of the density profile of the galaxy cluster. Hybrid models are some combination of the parametric and non-parametric algorithms. In addition to the different forms of modeling the dark matter halos, an assumption that light traces mass is used in some of the algorithms.

The following is a brief overview of the models analyzed in this work. A summary of the lens model names, algorithms, and constraints can be found in Table 2. A more detailed description of the lens models can be found on the *HST* Frontier Fields Lensing Models webpage.<sup>2</sup>

CATS (C-V3, C-V3.1) collaboration (Jauzac et al. 2012, 2014; Richard et al. 2014) and Sharon/Johnson (JS-V3) team (hereafter JS; Johnson et al. 2014) used the publicly available parametric lensing model software LENSTOOL (Jullo et al. 2007; Jullo & Kneib 2009). JS used only strong lensing,

<sup>2</sup> See <http://stsci.edu/hst/campaigns/frontier-fields/Lensing-Models> for the publicly available lens models.



**Figure 2.** *HST* optical F160W image of MACS J0416.1-2403, where the colored marks represent the positions of the multiple images, and the color indicates the background-source-measured spectroscopic redshift (Jauzac et al. 2014; Richard et al. 2014; Treu et al. 2015; Caminha et al. 2017). The plus symbols represent the training set, and the diamonds indicate the new lensed systems. The boxes indicate lensed system 26, which had a spectroscopic redshift update from  $z_{26} = 2.1851$  to  $z_{26} = 3.2355$ .

**Table 2**  
HFF V3 Lens Models for MACS J0416.1-2403

Lens Model	Algorithm Name	Algorithm	Constraints Used	Model Reference
B-V3	SWUnited	Non-Parametric	“GOLD”	Hoag et al. (2016)
C-V3	LENSTOOL	Parametric	“GOLD,” “SILVER”	Jauzac et al. (2014), Richard et al. (2014)
C-V3.1	LENSTOOL	Parametric	“GOLD,” “SILVER,” “BRONZE”	Jauzac et al. (2014), Richard et al. (2014)
D-V3	WSLAP+	Hybrid	“GOLD,” “SILVER”	Diego et al. (2015b, 2016b)
G-V3	GLAFIC	Parametric	“GOLD,” “SILVER”	Kawamata et al. (2016)
JS-V3	LENSTOOL	Parametric	“GOLD”	Johnson et al. (2014)
W-V3	GRALE	Non-Parametric	“GOLD,” “SILVER”	Mohammed et al. (2014)
W-V3.1	GRALE	Non-Parametric	“GOLD,” “SILVER,” “BRONZE”	Mohammed et al. (2014)
Z-LTM	LTM	Parametric	“GOLD,” “SILVER”	Broadhurst et al. (2005), Zitrin et al. (2013)
Z-NFW	PIEMD+eNFW	Parametric	“GOLD,” “SILVER”	Zitrin et al. (2009, 2013)

**Note.** This is a summary of the 10 public version 3 models of MACS J0416.1-2403, including the name, type of algorithm, and constraints used to compute the model. For definitions of the constraints, see Section 2.3.

while CATS used a combination of strong and weak lensing for the reconstruction of the mass distribution of the galaxy cluster. The total mass distribution of the cluster is parameterized by a combination of cluster-scale halos and a contribution from individual cluster-member galaxies, using scaling relations to observed galaxy properties. LENSTOOL uses a Bayesian Monte

Carlo Markov Chain algorithm to explore the parameter space, minimizing a  $\chi^2$  to identify the best set of parameters. There are two models computed by CATS. Version 3 (C-V3) used only the “GOLD” and “SILVER” arcs, while version 3.1(C-V3.1) used “GOLD,” “SILVER,” and “BRONZE.” The model by JS (JS-V3) only included the “GOLD” arcs.

Glaflac (G-V3) (Oguri 2010; Ishigaki et al. 2015; Kawamata et al. 2016) used a parametric algorithm in which a large smooth elliptical Navarro–Frenk–White (NFW; Navarro et al. 1997) profile represents the cluster mass distribution, while the individual galaxies are modeled using Pseudo-Jaffe ellipsoids (Keeton 2001). To complement these two components, external and internal perturbations of the lensing potential were introduced and represented by a multipole Taylor expansion. The “GOLD” and “SILVER” arcs were used to constrain the lens model.

Zitrin (Z-NFW, Z-LTM) models (Zitrin et al. 2009, 2013) used two different methods of parameterization. The first is a NFW dark matter halo density to describe the cluster mass distribution. The second method, introduced by Broadhurst et al. (2005), is called Light-Traces-Mass (LTM). LTM uses a parametric profile for the cluster-member galaxies scaled to their luminosity, then the mass map is smoothed using a Gaussian kernel to represent the dark matter component of the cluster. Both Z-LTM and Z-NFW use the “GOLD” and “SILVER” arcs.

Bradač (B-V3) (Hoag et al. 2016) used Strong and Weak Lensing United (SWUnited Bradač et al. 2006, 2009). This non-parametric model combines both strong and weak lensing analysis to constrain the gravitational potential through iterative minimization of  $\chi^2$  of a non-regular adaptive grid. The “GOLD” arcs were used to compute the model.

Williams (W-V3, W-V3.1) models (Liesenborgs et al. 2006; Mohammed et al. 2014) used GRALE, a non-parametric method that uses a genetic algorithm to iteratively refine the mass distribution on a grid. The models only use the lensing observables as constraints, with no other assumption on the mass density profile or cluster galaxies. This method is ideal for testing the assumption that light traces the mass of the cluster. The W-V3 model used the “GOLD” and “SILVER” arcs, while the model W-V3.1 used “GOLD,” “SILVER,” and “BRONZE” arcs.

Diego (D-V3)<sup>3</sup> (Diego et al. 2015b, 2016b) used WSLAP+, which is a free-form algorithm with the addition of a parameterized distribution for the mass contribution from the galaxy cluster members. The large-scale dark matter distribution of the cluster is obtained using the non-parametric algorithm, using a superposition of two-dimensional Gaussians in an adaptive grid. The compact mass contribution from the cluster members is tied to a light-to-mass ratio. The “GOLD” and “SILVER” arcs have been used as inputs to compute the lens model.

### 3. Strong Lens Models Evaluation

The lensing equation provides a transformation between each image location in the image plane to that of its source in the source plane:

$$\begin{aligned}\beta_s &= \theta_i - \alpha(\theta_i), \\ \alpha(\theta_i) &= \frac{D_{ls}(z_l, z_s)}{D_s(z_s)} \hat{\alpha}(\theta_i),\end{aligned}\quad (1)$$

where  $\beta_s$  is the source plane position,  $\theta_i$  is the location in the image plane of image  $i$  of source  $s$ ,  $\hat{\alpha}(\theta_i)$  is the deflection angle,  $D_{ls}(z_l, z_s)$  is the angular diameter distance between the lens and the source, and  $D_s(z_s)$  is the angular diameter distance between the observer and the source. Lens modeling algorithms

<sup>3</sup> The publicly released version 3 lens model of Diego was not suitable for this study due to a scaling problem. We obtained the correct version of the files from J. M. Diego (2018, private communication).

find the distribution of projected mass density that minimizes the scatter between the observed and predicted images of all background sources. Ideally, the strong lensing analysis would be done in the image plane where the observables are found, typically minimizing the equation (Kneib & Natarajan 2011)

$$\chi_s^2 = \sum_{i=1}^{n_s} \frac{[\theta_{\text{obs}}^i - \theta_{\text{model}}^i]^2}{\delta_{si}^2}, \quad (2)$$

where  $\theta_{\text{obs}}^i$  and  $\theta_{\text{model}}^i$  are the observed and the predicted positions by the lens model of image  $i$  of system  $s$ , respectively, and  $\delta_{si}^2$  is the error on the position of image  $i$  in the lensed system  $s$ .

When assessing the image plane scatter of the different lens models for the analysis presented in this paper, we found that some models fail to predict all the multiple images for some of the systems (see Section 3.3). In those cases, the scatter of the system would be driven by the largest angular separation between the observed images in the image plane, preventing a quantitatively meaningful assessment of the models.

We overcome this issue by resorting to measuring the scatter in the source plane. Since the lens equation does not need to be inverted, there is always a prediction for  $\beta_s$ . The source plane and image plane areas are related to each other through the lensing magnification, which can be different for each of the multiple images  $\theta_i$ . We therefore average over the magnification-weighted contribution to the scatter of each image, as detailed below.

#### 3.1. Source Plane Scatter per System

To quantify the performance of the lensing models we use the magnification-weighted rms scatter of the projected source positions in the source plane, with the goal of evaluating the predictive power of the models, and exploring its dependence on redshift and position in the image plane.

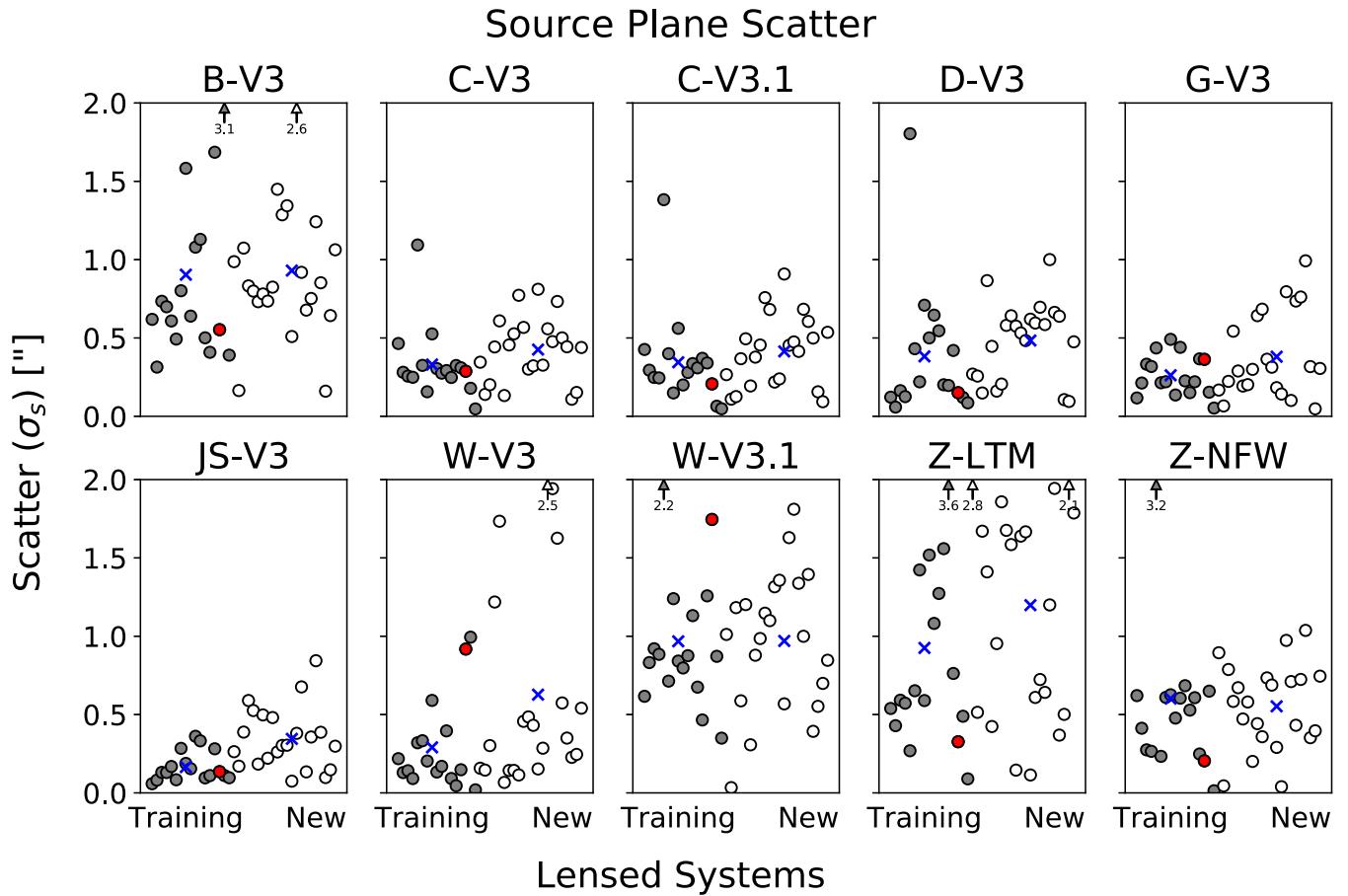
The source plane location  $\beta_i$  of each image  $i$  of source  $s$  is calculated from the lensing equation (Equation (1)). We use the deflection matrices ( $\hat{\alpha}$ ) from the publicly released lens models, observed positions of the arcs ( $\theta_i$ ), and the source spectroscopic redshift data  $z_s$  (see Table 3). The deflection maps are interpolated using a bivariate spline. The interpolation is critical to increasing the precision when determining the location in the source plane of the multiple images.

Since the true position of the background source is non-observable, we compute the barycenter ( $\langle \beta_s \rangle$ ) of the projected source positions (Gade 2010), weighted by the magnification of each image:

$$\langle \beta_s \rangle = \frac{\sum_{i=1}^{n_s} \mu_i \beta_i}{\sum_{i=1}^{n_s} \mu_i}, \quad (3)$$

where  $i = 1, 2, 3, \dots, n_s$  is the index of each image of lensed system  $s$ ,  $n_s$  is the total number of multiple images in the system, and  $\mu_i$  is the magnification of image  $i$ . Lastly, the distances between the barycenter and source plane locations are measured using the Vincenty formulae (Vincenty 1975) to measure an rms scatter in the source plane. The weighted source plane rms scatter ( $\sigma_s$ ) for each of the systems is computed as follows:

$$\sigma_s = \left( \frac{1}{n_s^2} \sum_{j=1}^{n_s} \frac{1}{\mu_j^2} \sum_{i=1}^{n_s} (\mu_i ||\beta_i - \langle \beta_s \rangle||)^2 \right)^{1/2}, \quad (4)$$



**Figure 3.** The system rms scatter for each lensed source ( $\sigma_s$ , Equation (4)) is plotted against the arc systems. The filled circles represent the systems that are part of the training set, and the open circles represent systems that are part of the new set. The arrows at the top of the graph indicate data points that have a scatter  $> 2''.0$ . The blue “x” represents the average scatter of the training systems and new systems, respectively, for each model ( $\bar{\sigma}$ , Equation (5)). The red dot marks system 26, whose redshift has been updated.

where the summation takes into account the different magnifications of the images.

When calculating the average scatter of several systems,  $\bar{\sigma}$ , we use the formula

$$\bar{\sigma} = \frac{1}{N} \sum_s^N (\sigma_s), \quad (5)$$

where  $s = 1, 2, 3, \dots, N$  is the index of the system,  $N$  is the total number of systems, and  $\sigma_s$  is the source plane scatter of system  $s$  (Equation (4)).

We note that even though all of the modelers used the same list of multiple images (as described in Section 2.3), the positions are slightly altered; for example, they may use a unique morphological feature or a bright clump in the arc as the positional constraint. The coordinates used in this paper are tabulated in Table 3. We tested whether the exact choice of coordinate within the arc affects the results of our analysis. We computed the percent difference in the source plane scatter between two sets of positional coordinates from two different lensing teams.<sup>4</sup> We find that the source plane rms scatter changes by less than 5% between the coordinate sets. This test provides confidence that the results described in the following sections are robust to slight differences in the coordinates used by the different teams.

### 3.1.1. General Results

In Figure 3, we plot the rms scatter ( $\sigma_s$ ) for each system for all the lens models, and in Figure 4 we show the distributions (represented by  $\bar{\sigma}$ , first, and third quartiles) of the system rms scatter of the training set, new set, and combined set, for a more direct comparison between the models.

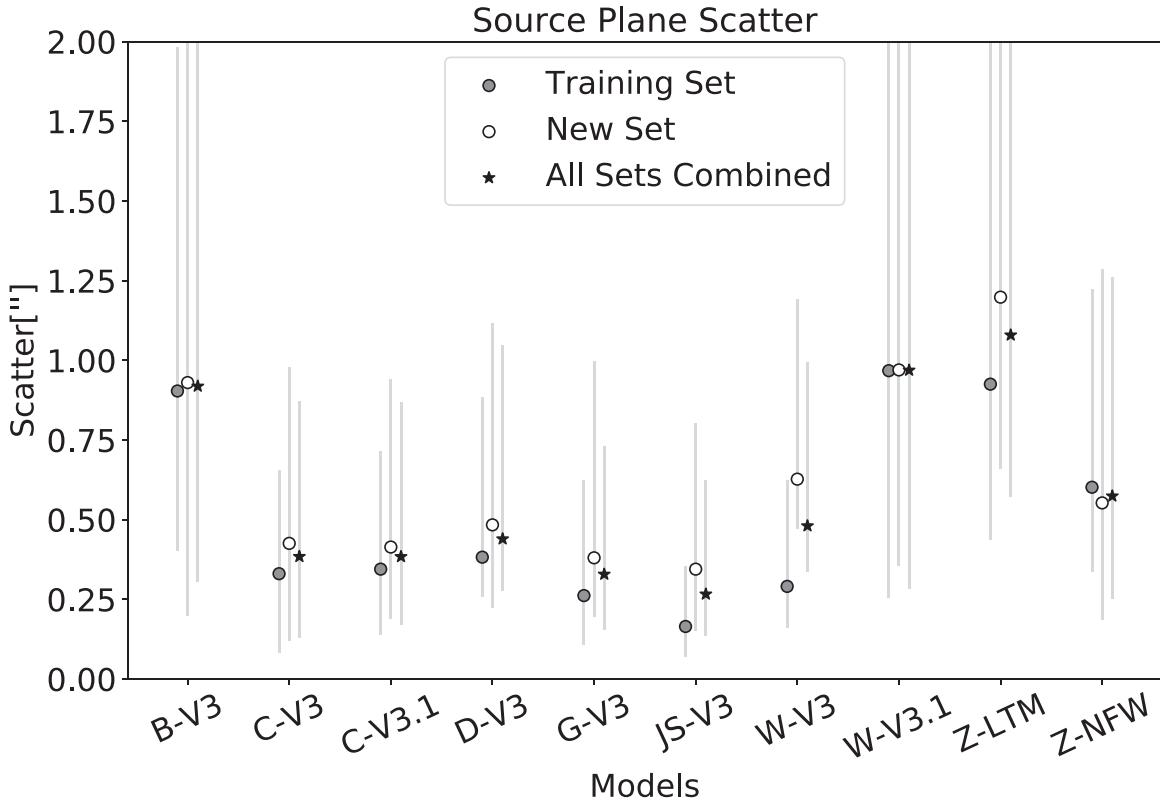
We find that models based on parametric or hybrid algorithms result in low rms scatters (C-V3, C-V3.1, D-V3, G-V3, JS-V3, and Z-NFW), with combined average rms scatters of  $\bar{\sigma} < 0''.5$ ; The only parametric model that has a combined scatter of roughly  $1''.0$  is Z-LTM. On the other hand, most of the non-parametric algorithms result in higher rms scatter (B-V3, W-V3.1, Z-LTM), with the exception of W-V3.

Comparing the scatter of the training set to that of the new set, we find that with the exception of a few outliers, the training set systems tend to have smaller source plane scatter compared to the new systems. This is not surprising, since the training set was used to compute the models, meaning that the lens models are tuned to reproduce these multiple images.

We investigate whether the source plane rms scatter ( $\sigma_s$ ) increases or decreases with the system redshift ( $z_s$ ) in Figure 5; we find no significant trend with redshift. The apparent larger source scatter at  $z > 2.5$  is due to the fact that most of the sources in the new set are at a higher redshift.

Two teams computed additional models (V3.1) that include the “BRONZE” arcs, which allows us to assess their influence.

<sup>4</sup> The coordinates of the constraints are from the JS and Diego teams (2018, private communication).



**Figure 4.** The average model source plane scatter for the training, new, and combined sets for each of the models is plotted. The first and third quartiles are represented by the error bars. Most of the parametric models (C-V3, C-V3.1, G-V3, JS-V3, and Z-NFW) have similar rms scatter, with a combined rms scatter  $\sigma < 0''.5$ . The only parametric model with a high rms scatter is Z-LTM. The hybrid model, D-V3, has a scatter that is similar to that of the parametric models. The non-parametric models, B-V3 and W-V3.1, have a higher rms combined scatter compared to that of the parametric and hybrid models, while W-V3 has a low rms combined scatter.

When comparing C-V3 and C-V3.1, which use a parametric algorithm, we find that these two models have a similar rms. This is not the case in models W-V3 and W-V3.1, which use a non-parametric method. Most of the systems in W-V3 have low scatter, except for a few outliers. In W-V3.1, the rms scatter of all the systems is generally higher compared to that of W-V3. The difference between the V3 and the V3.1 models from both teams is the number and quality of constraints used when they were computed: V3 included “GOLD” and “SILVER” multiple images, while V3.1 used the “GOLD,” “SILVER,” and “BRONZE” arcs. This difference suggests that the GRALE algorithm produces better results when a higher fraction of the lensing constraints are considered “high quality.” Measured by the rms diagnostic, this suggests that, at least for this particular cluster, adding low-confidence arcs that do not have spectroscopic redshifts degrades the lens model despite increasing the quantity of constraints.

### 3.1.2. Arc System 26, Source Plane Scatter

The latest spectroscopic follow-up by Caminha et al. (2017) confirmed most of the previously known spectroscopic redshifts. One exception is the redshift of system 26, which was updated from  $z_{26} = 2.1851$  (Treu et al. 2015) to  $z_{26} = 3.2355$  (Caminha et al. 2017). This change in redshift is of interest because it can be used to test how models perform when an erroneous spectroscopic redshift constraint is used to compute them.

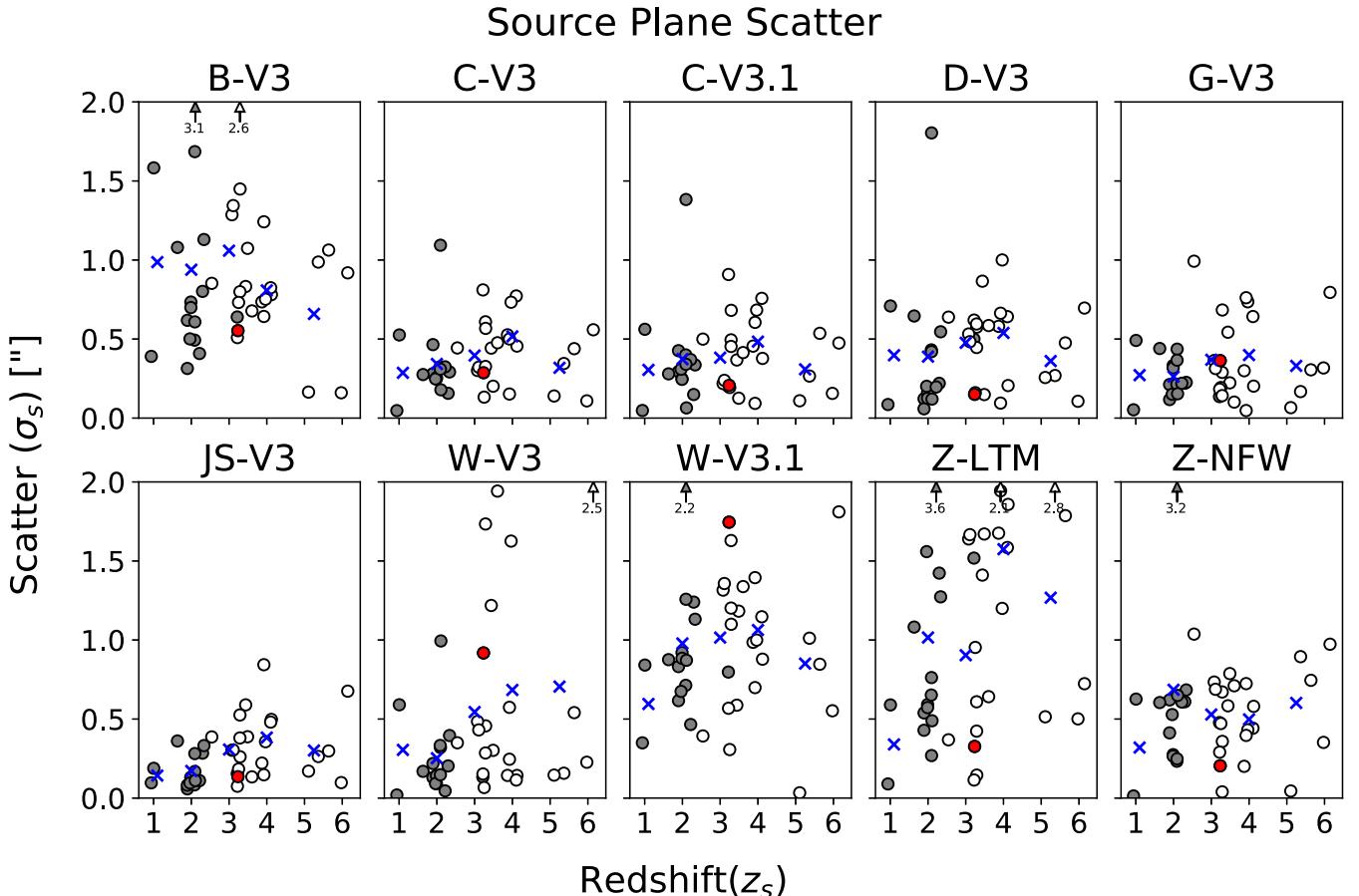
In Figure 6, we plot the rms scatter of system 26 ( $\sigma_{26}$ ) for both of the  $z_s$  values for each model. The difference in source plane system scatter is only due to the change in the

spectroscopic redshift. In most of the models, the rms scatter of system 26 is about  $0''.25$  lower with the new  $z_{\text{spec}}$ , meaning that system 26 is better reproduced by the lens models when the corrected spectroscopic redshift of  $z_{26} = 3.2355$  is used. The only model in which the rms scatter increased was W-V3.1, with a change of  $\sim 1''.0$ ; however, this model had generally higher source plane scatter (Figure 3), therefore this deviation may not be informative. A visual inspection of the positions of the images of system 26 relative to the magnification maps for the two redshifts indicates that in both cases the lensing configuration of this system is not reproduced well by this model. These changes show that in a case like the HFF clusters, which have dozens of systems with spectroscopic redshifts, the effects of one erroneous redshift can be averaged out when computing the model average rms scatter.

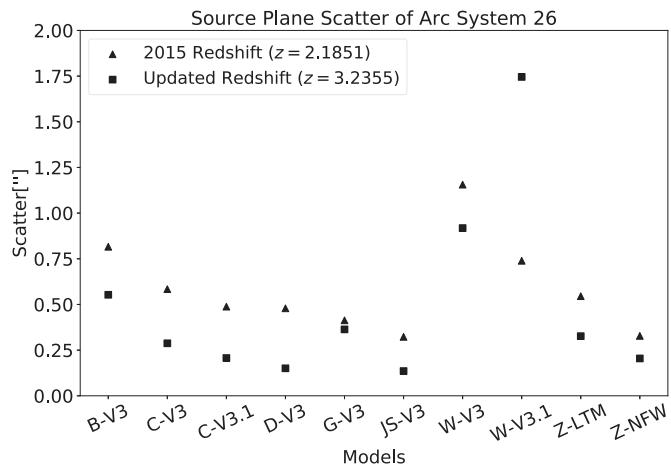
#### 3.1.3. Source Plane Scatter Dependence on Image Location

Exploring the rms scatter of systems with respect to their location in the image plane provides us with information related to regions in the image plane where a model performs well and regions where a model does not. In Figures 7–9, we overplot the positions of the multiple images on a WFC/IR (F160W) image of MACS J0416.1-2403. The source plane system rms scatter is color-coded as shown in the legend.

For some of the models (C-V3, C-V3.1, D-V3, G-V3, and JS-V3) the source plane rms scatter is uniformly low across the image plane, while others show large variation from region to region in the image plane. In addition, the similarity between C-V3 and C-V3.1 implies that when these parametric models already have a large number of constraints with spectroscopic



**Figure 5.** The lensed system scatter ( $\sigma_s$ , Equation (4)) is plotted against the redshift of the system ( $z_s$ ). The red circle shows the lensed system 26, plotted at its updated (correct) redshift. The blue “x” represents the average scatter for lensed systems in the following redshift bins: [0.9, 1.5], (1.5, 2.5], (2.5, 3.5], (3.5, 4.5], and (4.5, 6.2]. We find no significant dependence on redshift within the training set (filled circles) and new sets (open circles). The apparent larger source scatter at higher redshifts is due to the fact that most of the sources in the new set are at higher redshifts, as shown in Figure 1 and described in Section 2.3.



**Figure 6.** The source plane scatter of the lensed system 26 ( $\sigma_{26}$ ) is plotted for the different models. The redshift of this particular system was updated with the spectroscopic data from  $z_{26} = 2.1851$  (Treu et al. 2015) to  $z_{26} = 3.2355$  (Caminha et al. 2017). For most of the lens models the rms scatter is lower with the updated spectroscopic redshift, except for W-V3.1, in which the scatter increased by  $\sim 1.^{\prime\prime}0$ .

redshifts, they do not benefit from adding low-score constraints for which no spectroscopic data are available.

The distribution of the system scatter in B-V3 is roughly uniform throughout the image plane with a slightly higher rms scatter in the southwest region of the field of view.

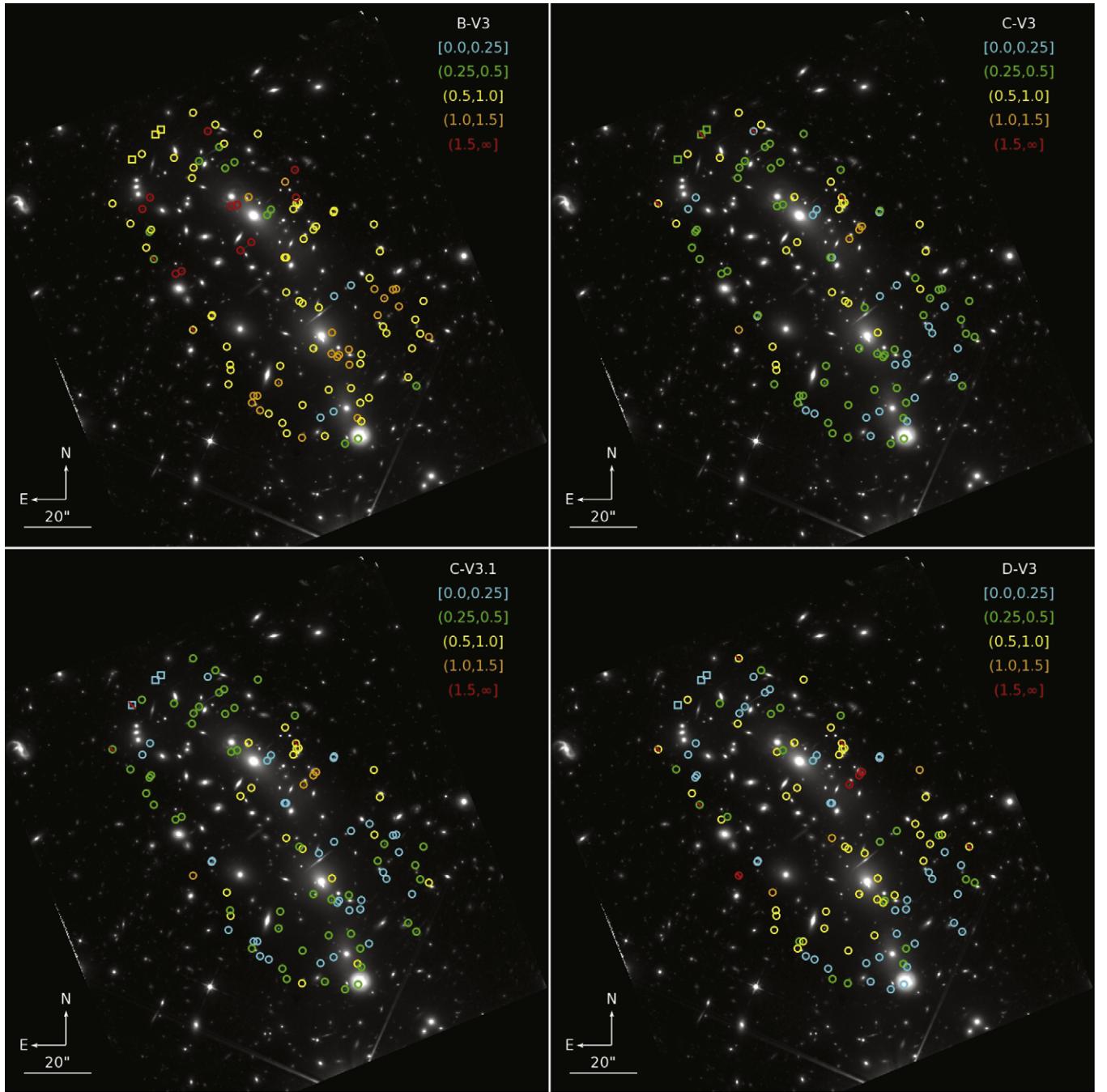
Both Zitrin models, Z-LTM and Z-NFW, have similar behavior across the field of view, with lower rms values in the northeast and relatively high source plane rms scatter in the southwest. However, Z-LTM has higher rms scatter overall, with rms higher than  $1.^{\prime\prime}5$  in the southwest. Most of the systems in Z-NFW have an rms scatter lower than  $1.^{\prime\prime}0$  (Figure 9).

Last, we observe drastic differences between W-V3 and W-V3.1. In W-V3, most of the high-scatter systems are found in the northeast region of the image plane. In contrast, W-V3.1 has a more uniform rms scatter through the image plane but is generally higher everywhere compared to W-V3. From this we can conclude that this model is strongly affected by the quality of the constraints —adding the “BRONZE” images (low level of agreement on the robustness of the arc, and no spectroscopic redshifts) increased the rms scatter in general across the entire image plane.

### 3.2. Evaluating the Models’ Predictive Power

To evaluate the average predictive power of the lens models, we use the distribution of the systems source plane rms scatter (Figure 3). Part of the predictive power is the ability of a model to accurately predict the position of new multiple images, with similar accuracy to that of the set of images that were used as constraints. To do so, we compare the rms scatter of the new systems to that of the training set.

In Figure 10, we plot the distributions of the training set and new systems. We have added dashed gray lines to guide the



**Figure 7.** *HST* optical F160W image of MACS J0416.1-2403. The positions of the multiple images are indicated by the circles, while the color represents the source plane scatter of the lensed systems. These models show the rms scatter is uniform across the image plane, while others have regions of relative high and low rms scatter. The square symbols represent the location of lensed system 26, and the squares and circles with a line across the center, e.g.,  $\square$ , represent “failed” arcs (see Section 3.3).

eye; they have a slope of one and increase by  $0.^{\circ}2$ . The models with the best predictive power are those closest to the 1:1 line (lowest gray line) and low average training set rms; these models, C-V3, C-V3.1, D-V3, G-V3, JS-V3, and Z-NFW, maintain similarly low, or only slightly higher average rms scatter when confronted with lensing evidence that was not used to constrain the model. W-V3 has a low training set mean rms, although when compared to that of the new set, we find it is further away from the self-similar line meaning that the model did not perform as well with the new images. In the case where the models have a high training set arithmetic mean and are close to the 1:1 line, they include W-V3.1 and B-V3. These

models perform similarly between the two sets of images with relative high rms scatter compared to the rest of the models. Last, Z-LTM has a high training set rms distribution and is far from the self-similar line, meaning that the model already has a high rms scatter and the scatter increases even more when confronted with the lensing evidence that was not used to constrain the model.

### 3.3. Image Plane Projection Analysis

In this section, we test the ability of the models to recover the observed lensing configuration in the image plane. We do this

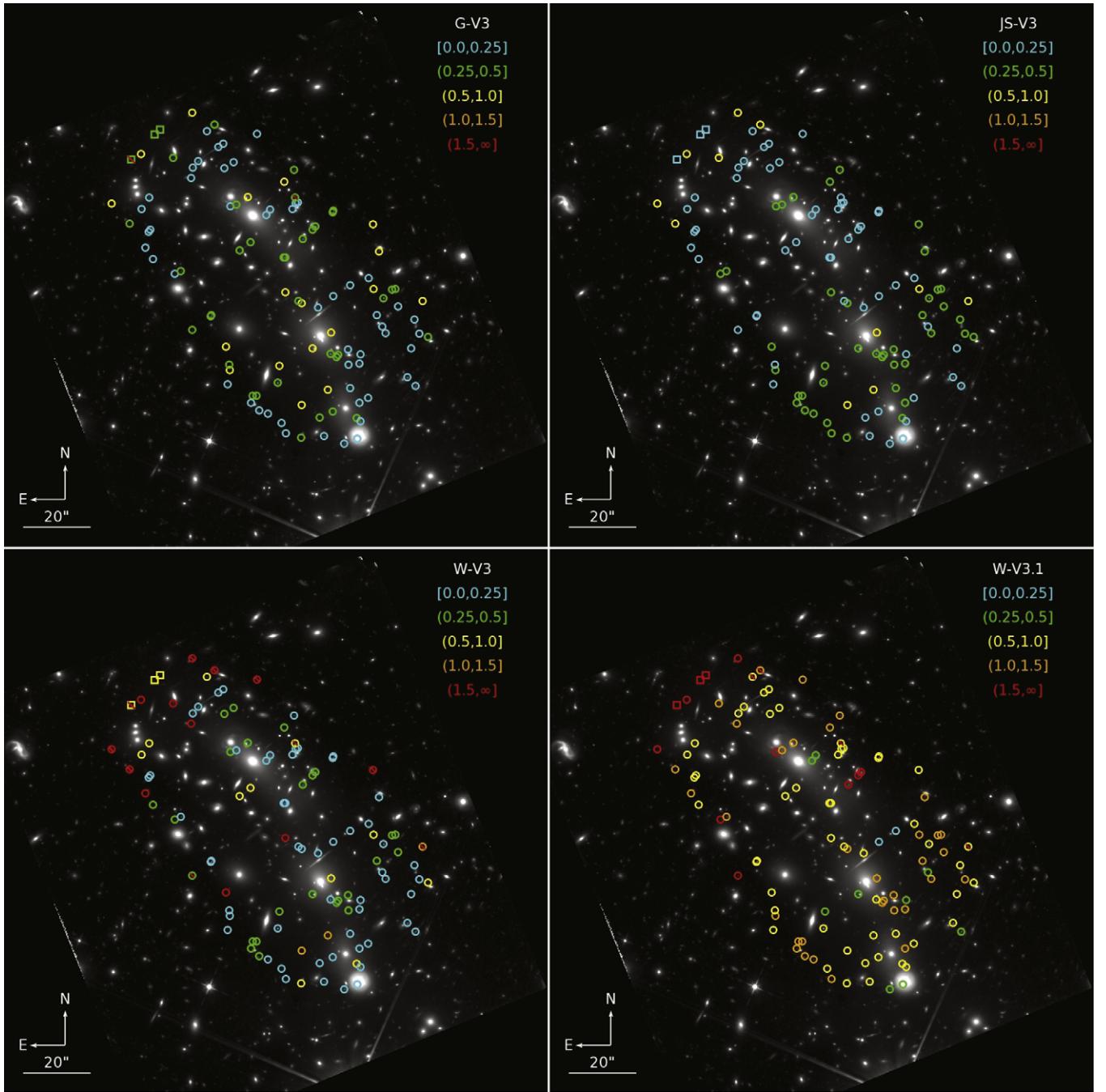


Figure 8. Same as Figure 7.

by computing the predicted source position of each image of each system using Equation (1), then lens it back to all of its predicted positions in the image plane. A correct prediction of the lensing configuration will result in a number of images in the image plane that is at least the same as the number of observed images in each system. It may sometimes predict additional images that may be unobservable, e.g., due to low magnification or obscuration by foreground sources.

To allow for conservative positional uncertainties, we set a circular area with a radius of  $0.^{\prime\prime}1$  around the location of the source plane position. We then ray-trace the area inside the circle to the image plane to predict its multiple images. The predicted images are identified using the Hoshen–Kopelman (Hoshen &

Kopelman 1976) cluster labeling algorithm. The original image should always be reconstructed. If more predictions appear, it means that the model predicts multiple images, corresponding to the circle that we had set in the source plane. In the case where no additional predicted images are found in the image plane, then the model does not predict multiple images from the circle in the source plane. We refer to these multiple images as failed arcs, because they fail to create some or all of the multiple projections.

The process of finding the other image plane projections is then repeated for each one of the multiple images. An arc fails to produce counter images if its source circle falls entirely outside the caustic (the source plane projection of the critical

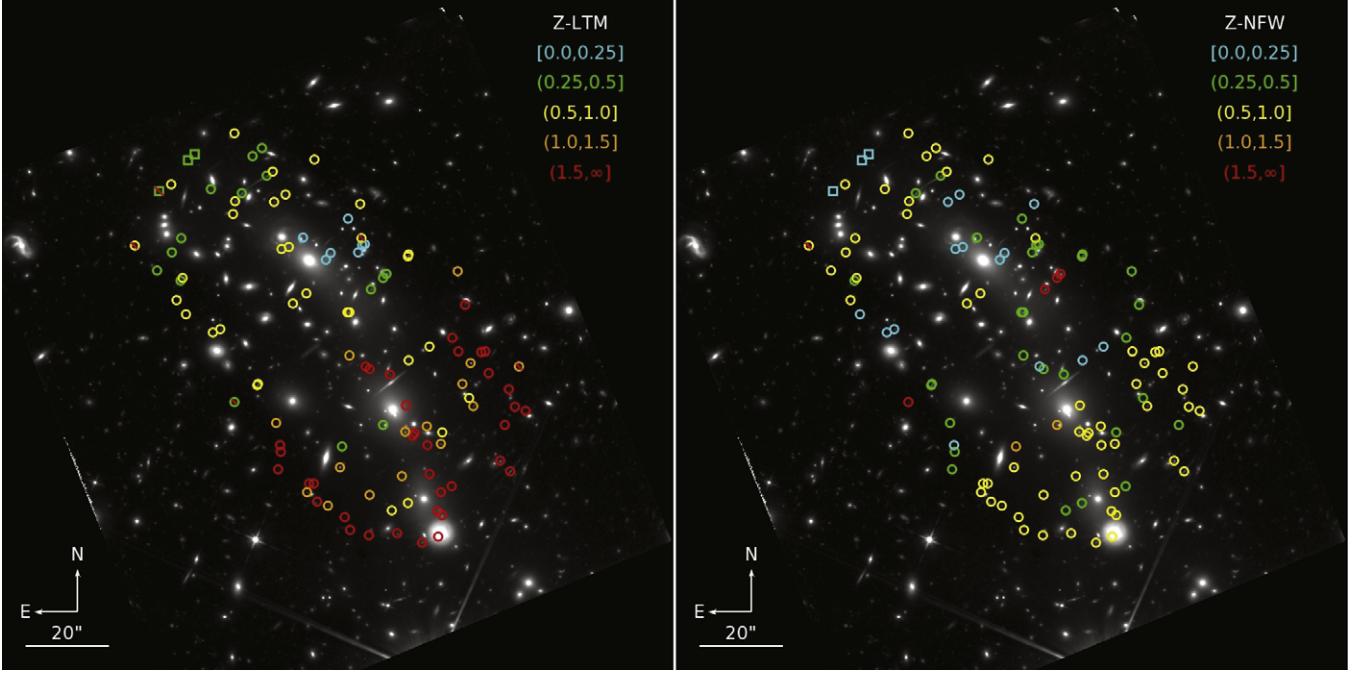
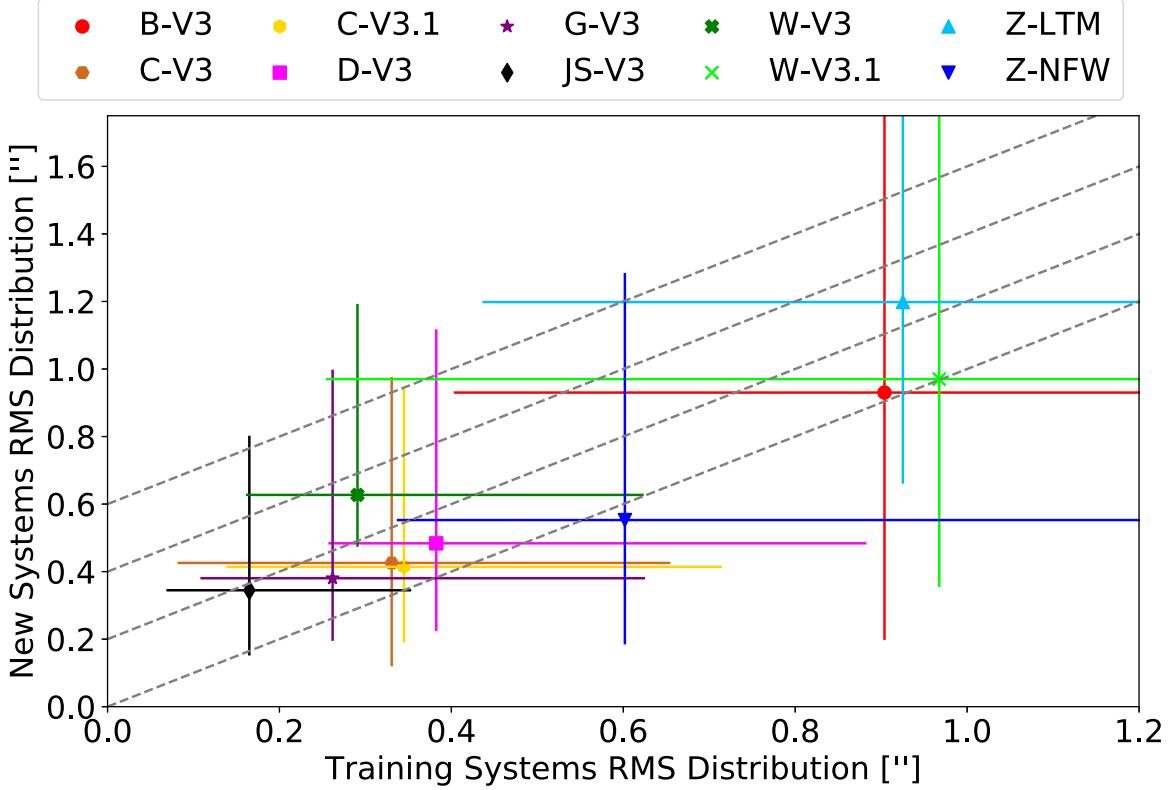


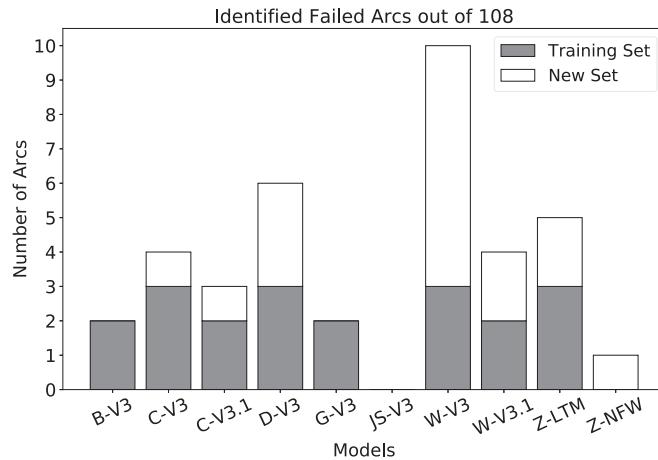
Figure 9. Same as Figure 7.



**Figure 10.** To evaluate the predictive power of the lens models we plot the distribution of the training and new system rms scatter. The error bars represent the first and third quartiles, and the symbols represent the arithmetic mean (Equation (5)) of the distribution. The dashed gray lines are there to guide the eye, they have a slope of one and increase by  $0''.2$ . The models with the best predictive power are those that are closest to the 1:1 line (lowest gray dash line) and have the lowest distribution for the training set.

curve, representing loci of high  $\mu$  in the image plane) for its redshift, in which case the model cannot predict the correct number of multiple projections in the image plane.

As can be seen in Figure 11, we find that the lens models are able to reproduce the majority of the 108 multiple images, with a small fraction of systems that are not fully reproduced. Most



**Figure 11.** Number of the arcs, out 108 total, that were labeled as failed arcs. We compute the source plane position of the observed arc, then lens it forward to the image plane. Failed arcs are defined as those that do not produce additional images other than the original location, i.e., they fail to produce the lensing multiplicity (see Section 3.3). The heights of the bars represent the number of failed arcs by the model and the colors indicate the contribution from the training and new sets.

of the models have less than 5 failed arcs out of the 108 arcs, e.g., overall there is a  $>95\%$  success rate and the models perform well in reproducing the lensing configuration.

The JS-V3 model is able to produce a projection for each one of the 108 arcs, while the W-V3 is the model with the most failed arcs, 10, most of them in the new set, in the northeast region of the cluster—systems that also have high rms. The failed arcs are shown in Figures 7–9 as a square or circle with a line across the center, e.g.,  $\oslash$ .

#### 4. Version 4 Lensing Models

In 2017, the lensing teams repeated the vetting process for all the multiple images in MACS J0416.1-2403, including newly available information on their spectroscopic redshifts from Caminha et al. (2017). Using this new vetted list they computed the next version of lens models (V4).

Five of the teams, CATS (Cats-V4 and Cats-V4.1), Glafic (G-V4), Diego (D-V4 and D-V4.1), Sharon/Johnson (JS-V4), and Williams (W-V4), who had computed version 3 models, have new publicly available V4 models. In addition, some new groups, Caminha (Cam-V4) and Keeton (K-V4), have computed lens models and made them public through the HFF website. At the time of submission of this manuscript, there are no V4 models from the teams of Bradač and Zitrin.

We have described the algorithms used by the CATS, Glafic, Diego, Sharon/Johnson, and Williams teams in Section 2.4. Caminha (Caminha et al. 2017) used LENSTOOL (Jullo et al. 2007; Jullo & Kneib 2009), which is the same algorithm used by CATS and Sharon/Johnson. Keeton uses GRAVELENS (Keeton 2011), which is also a parametric lens modeling algorithm.

We use the exact same constraints as in our analysis of the version 3 models (see Table 3) and the same methodology (see Section 3.1). We compute the source plane system rms scatter ( $\sigma_s$ , Equation (4)) and plot it in Figure 12 for each lensed system. In this case all of the constraints are considered a training set, since they were used as constraints when

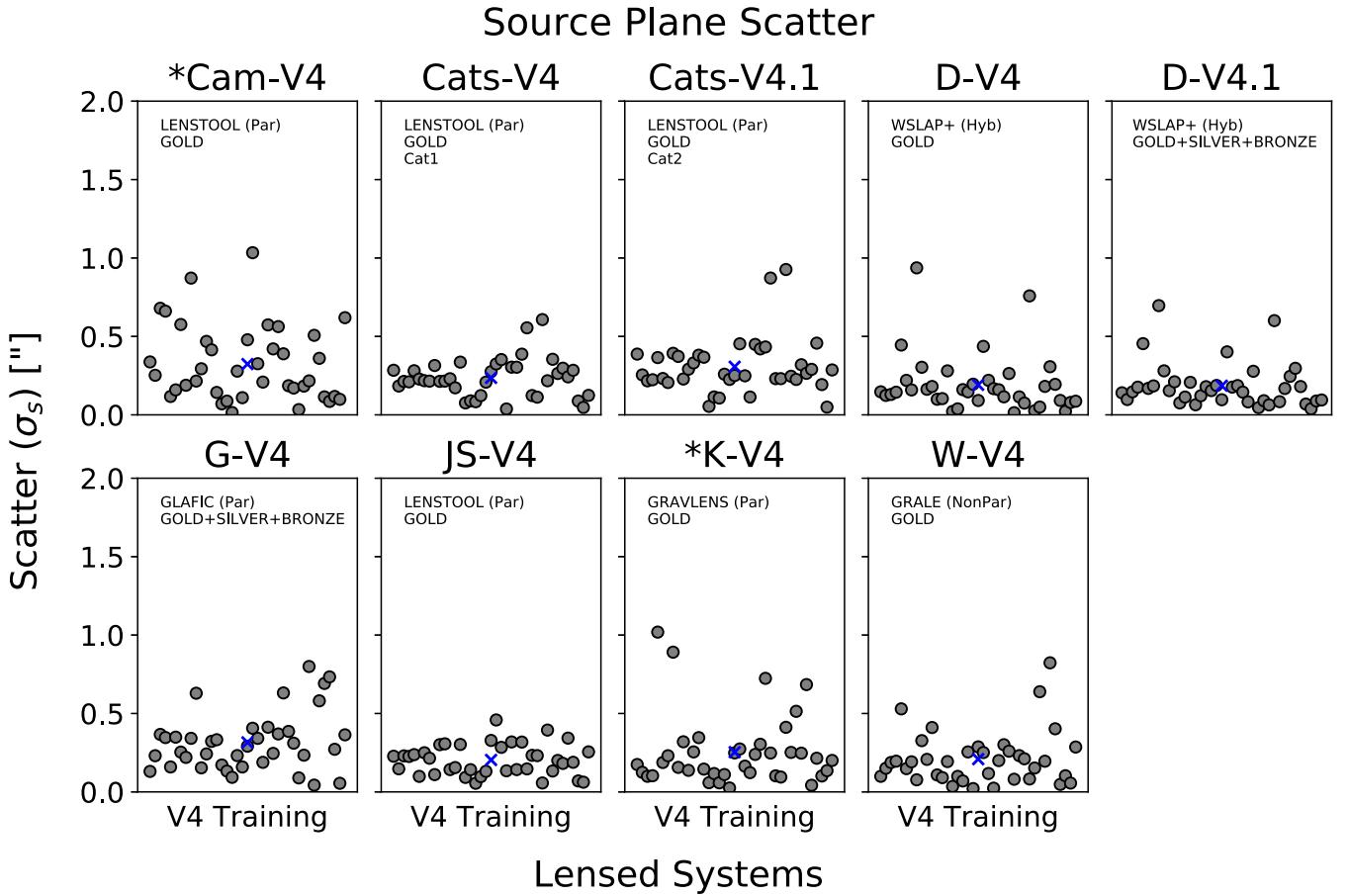
computing the V4 lens models. In the CATS models, the different versions (V4 and V4.1) are not related to the quantity and quality of images, but indicate two different cluster-member galaxy catalogs (Cat1: Grillo et al. 2015 and Cat2: J. Richard et al. 2018, in preparation). In the models by Diego (J. Vega et al. 2018, in preparation) two different versions are related to the quality and quantity of constraints used to compute the models.

In Figure 12, we observe that all of the lens models that are currently available on the Mikulski Archive for Space Telescopes have a low rms scatter and the distribution is similar for all of these models. Similarly, all the models have up to four failed arcs, with most having only 0–2 failed arcs. That the rms scatter is similar and low for all of the lens models and that there are few failed arcs demonstrates how well these lens models are doing in reproducing the lensing configuration.

#### 5. Summary and Conclusions

In this work we evaluated publicly available, version 3 HFF lensing models for MACS J0416.1-2403, using the source plane rms scatter and the image plane projection analysis as the metric. We have developed a method independent of all lensing algorithms used by the lensing teams to compute the source plane rms scatter of the multiple images and predictions in the image plane. We compared the rms scatter of the lens models using the training set—multiple images with spectroscopic redshifts that were known to the modelers at the time the models were computed—and the new arcs, multiple images that did not have spectroscopic redshifts at the time the models were computed. The new redshifts were published later by Caminha et al. (2017). We summarize our findings as follows:

1. When comparing the source plane rms scatter of different models, we found that the new arcs have a higher rms scatter compared to the training set in general for all the lens models (Figure 4). This is not surprising, since the models are tuned to reproduce the training systems.
2. We find that C-V3 and C-V3.1 have a similar source plane scatter. The only difference between these two models is the number and quality of constraints used to compute the model: V3 used “GOLD” and “SILVER” constraints, while V3.1 used “GOLD,” “SILVER,” and “BRONZE.” This indicates that parametric models do not necessarily benefit from adding more constraints with no spectroscopic redshift measurements when there are already hundreds of constraints that do have spectroscopic data, like in the case of the HFF clusters.
3. We observe clear differences between the models W-V3 and W-V3.1. In the case of W-V3, where only “GOLD” and “SILVER” constraints were used, the majority of the rms system scatter is fairly low, with the exception of some outliers. On the other hand, W-V3.1 used all the “GOLD,” “SILVER,” and “BRONZE” arcs. Overall the source plane system rms scatter is higher than that of W-V3. The observed differences between these two models tells us that this algorithm performs better when a higher fraction of the constraints used have high confidence and measured spectroscopic redshifts.
4. Comparing the rms scatter against the source redshift (Figure 5), we found no trend with spectroscopic redshift. The only difference comes from the limited range of redshifts from the training set (Figure 3) and the lower



**Figure 12.** Source plane rms scatter plotted for each of the lensed systems using the version 4 publicly available lens models. All of the constraints are considered training, since they were used as inputs when computing the new lens models. The blue “x” represents the mean of the distribution. In the text of each panel you can find the name of the algorithm used, in parentheses if the algorithm is parametric (Par), non-parametric (NonPar), or hybrid (Hyb). The text of the panels also shows the new constraints used, and additional comments to describe the difference between models from the same lensing group (two sets of galaxy cluster-member catalogs, Cat1: Grillo et al. 2015 and Cat2: J. Richard et al. 2018, in preparation). The “\*\*” in the name is to highlight the new teams that have made V4 lens models that did not compute a V3 lens models for MACS J0416.1-2403.

source plane rms scatter of the training set compared to the new set.

5. We computed the source plane scatter of system 26 (Figure 6) to assess the effects of the drastic change in the spectroscopic redshift of the system (Caminha et al. 2017) from  $z_{26} = 2.1851$  to  $z_{26} = 3.2355$ . We find that in all the models but one, W-V3.1, the rms for this system decreases when calculated with the correct redshift. In HFF, because there are so many constraints, one catastrophic failure is not going to break the model (just because of high scatter for this one system). In cases with a low number of constraints, however, it is significant (Johnson & Sharon 2016).
6. We explore the spatial distribution of the system source plane rms scatter against the location of the multiple images in the image plane with Figures 7–9. The models C-V3, C-V3.1, D-V3, G-V3, and JS-V3 have a uniform performance throughout the image plane. Z-LTM and Z-NFW have relatively high rms scatter in the southwest region of the image plane. Lastly, we observe drastic differences between the W-V3 and W-V3.1 models. Most of the W-V3 multiply imaged systems have a low rms scatter and are located in the middle and southwest regions of the field of view and three high-scatter systems are found in the northeast region of the image plane. On

the other hand, in W-V3.1 the rms scatter is uniformly high across the entire field of view.

7. Most of the parametric (C-V3, C-V3.1, G-V3, JS-V3, and Z-NFW), and hybrid (D-V3) models have a combined average rms scatter below  $0.^{\prime\prime}5$ , while the non-parametric models (B-V3 and W-V3.1) have a higher average rms scatter. The W-V3 non-parametric model has a low average rms scatter. Lastly, the Z-LTM has the highest average rms scatter overall.
8. We have explored the predictive power of the lens models by comparing the distribution of the new and training set system rms scatter (Figure 10). We find that the models that have the best predictive power are C-V3, C-V3.1, D-V3, G-V3, JS-V3, and Z-NFW, which are found in the lower left part of Figure 10 and are closest to the 1:1 line.
9. In the image plane, we find that the models correctly predict projections for most of the arcs, but there are some systems that fail to produce all the lensed images—we named these “failed arcs.” We find that for most of the models, less than 5 out of the 108 arcs failed (Figure 11). The model JS-V3 predicts projections for all of the multiple images, while W-V3 has the most failed arcs at 10, most of them found in the northeast region, near system 26, where the scatter is generally higher. This

- means that the lens models, regardless of the algorithm used, are very successful in reproducing most of the multiply imaged systems.
10. The new version 4 lens models for MACS J0416.1-2403 have been made public by most of the lensing teams, as well as two new ones. We plot the source plane scatter in Figure 12 for the models available at the time of the submission of this paper. We observe a similar low source plane rms scatter distribution between all the lens models, demonstrating the progress in the quality of data, better characterization of MACS J0416.1-2403, and understanding of strong lens modeling systematics.

Recent studies that make extensive use of the HFF lens models note differences between model outputs from the different algorithms (Livermore et al. 2017) particularly at high magnifications (Bouwens et al. 2017), highlighting a need for a direct comparison between lensing algorithms. Mahler et al. (2018) compared the effect of the addition of new spectroscopic redshifts on the uncertainties in the lensing mass reconstruction, magnifications, and redshift predictions. Our findings are consistent with their rms analysis concluding that lensing models benefit from having more spectroscopic measurements to the constraints.

This work complements other comparison projects like the one done by Prieve et al. (2017), where the magnification of the lensing models was used as a metric, and work done by Meneghetti et al. (2017), where simulated clusters were used to evaluate the algorithms. While the rms scatter of the source plane is not the only measure of performance of the models, it provides important insights. The analysis presented here forms a quantitative assessment of different models of real lenses by confronting them with new observational lensing evidence. It

provides the community with a means to understand where the strengths and differences of lens modeling algorithms lie; and it enables confidence that models of strong lensing clusters are becoming extremely powerful and reliable, allowing for their use as cosmic telescopes.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of the Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. Support for MAST for non-HST data are provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts.

The authors thank the anonymous referee for insightful suggestions that improved this manuscript. We thank all of the lensing teams and all who contributed to the data that were used in this paper. In particular we thank J. M. Diego and J. Vega for providing the correct lensing model and the coordinates of the constraints, and L. Williams and D. Coe for providing files that were not yet readily available on MAST. We thank M. Bradač for useful discussion.

## Appendix

We provide in Table 3 the list of multiple images that were used in this work and by the lens modeling teams, along with their coordinates and redshift measurements.

**Table 3**  
List of the Lens Model Constraints Used in This Analysis

ID	R.A. (deg)	Decl. (deg)	$z$	$z_{\text{new}}$	Training or New
1.1	64.04075	-24.061592	1.896	1.896	Training
1.2	64.043479	-24.063542	1.896	1.896	Training
1.3	64.047354	-24.068669	1.896	1.896	Training
2.1	64.041165	-24.061853	1.8925	1.895	Training
2.2	64.043022	-24.063024	1.8925	1.895	Training
2.3	64.047487	-24.068856	1.8925	1.895	Training
3.1	64.03077	-24.067128	1.9885	1.9894	Training
3.2	64.035252	-24.07099	1.9885	1.9894	Training
3.3	64.041802	-24.075731	1.9885	1.9894	Training
4.1	64.03082	-24.067231	1.9886	1.9887	Training
4.2	64.035142	-24.070968	1.9887	1.9887	Training
4.3	64.041891	-24.075856	1.9888	1.9887	Training
5.1	64.032406	-24.068414	2.0918	2.0948	Training
5.2	64.032651	-24.068666	2.0922	2.0948	Training
5.3	64.03352	-24.069451	2.0952	2.0948	Training
5.4	64.0435551	-24.07695762	2.088	2.0948	Training
7.1	64.039798	-24.063088	2.0854	2.0881	Training
7.2	64.040667	-24.063592	2.0854	2.0881	Training
7.3	64.047102	-24.071106	2.0854	2.0881	Training
10.1	64.026053	-24.077256	2.2982	2.2982	Training
10.2	64.0284	-24.079743	2.2982	2.2982	Training
10.3	64.03669519	-24.08390609	2.2982	2.2982	Training
11.1	64.03925	-24.070386	1.0054	1.0054	Training
11.2	64.038296	-24.06973	1.0054	1.0054	Training
11.3	64.034233	-24.066021	1.0054	1.0054	Training
13.1	64.027567	-24.072672	3.2226	3.2175	Training
13.2	64.032161	-24.075103	3.2226	3.2175	Training

**Table 3**  
(Continued)

ID	R.A. (deg)	Decl. (deg)	$z$	$z_{\text{new}}$	Training or New
13.3	64.040353	-24.081486	3.2226	3.2175	Training
14.1	64.026233	-24.074339	1.6324	1.6333	Training
14.2	64.031042	-24.078961	1.6344	1.6333	Training
14.3	64.03583	-24.081322	1.6344	1.6333	Training
15.1	64.02686	-24.075745	2.3355	2.3355	Training
15.2	64.02944699	-24.07858856	2.3355	2.3355	Training
15.3	64.038234	-24.082985	2.3355	2.3355	Training
16.1	64.02406843	-24.08089322	1.9644	1.9644	Training
16.2	64.028334	-24.084547	1.9644	1.9644	Training
16.3	64.03161189	-24.085762	1.9644	1.9644	Training
17.1	64.029818	-24.086364	2.2181	2.2182	Training
17.2	64.028608	-24.085981	2.2181	2.2182	Training
17.3	64.02334	-24.081581	2.2181	2.2182	Training
23.1	64.044566	-24.072098	2.0943	2.0932	Training
23.2	64.039578	-24.066631	2.0943	2.0932	Training
23.3	64.034342	-24.063742	2.091	2.0932	Training
26.1	64.046452	-24.0604	2.1851	3.2355	Training
26.2	64.046963	-24.060796	2.1851	3.2355	Training
26.3	64.049083	-24.062863	2.1851	3.2355	Training
27.1	64.04813491	-24.06695355	2.1067	2.1067	Training
27.2	64.047469	-24.066045	2.1067	2.1067	Training
27.3	64.04221281	-24.06054175	2.1067	2.1067	Training
28.1	64.036506	-24.067024	0.9377	0.9397	Training
28.2	64.036838	-24.067457	0.9377	0.9397	Training
33.1	64.02840815	-24.08299301	...	5.365	New
33.2	64.03505002	-24.08549952	...	5.365	New
33.3	64.02298453	-24.07726749	...	5.365	New
34.1	64.029254	-24.073289	...	5.106	New
34.2	64.030798	-24.07418	...	5.106	New
35.1	64.037492	-24.083636	...	3.4909	New
35.2	64.029418	-24.079861	...	3.4909	New
35.3	64.024937	-24.075016	...	3.4909	New
38.1	64.033625	-24.083178	...	3.4406	New
38.2	64.031255	-24.081905	...	3.4406	New
38.3	64.022701	-24.074589	...	3.4406	New
44.1	64.045259	-24.062757	...	3.2885	New
44.2	64.041543	-24.059997	...	3.2885	New
44.3	64.049237	-24.068168	...	3.2885	New
47.1	64.026328	-24.076694	...	3.2526	New
47.2	64.028329	-24.078999	...	3.2526	New
48.1	64.035489	-24.084668	...	4.1218	New
48.2	64.029244	-24.081802	...	4.1218	New
48.3	64.023416	-24.076122	...	4.1218	New
49.1	64.033944	-24.074569	...	3.871	New
49.2	64.040175	-24.079864	...	3.871	New
51.1	64.04013543	-24.08029748	...	4.1032	New
51.2	64.03366471	-24.0747629	...	4.1032	New
51.3	64.02663527	-24.07047437	...	4.1032	New
55.1	64.035233	-24.064726	...	3.2922	New
55.3	64.038514	-24.065965	...	3.2922	New
58.1	64.025187	-24.073582	...	3.0773	New
58.2	64.03773	-24.08239	...	3.0773	New
58.3	64.030481	-24.07922	...	3.0773	New
67.1	64.038075	-24.082404	...	3.1103	New
67.2	64.025451	-24.073651	...	3.1103	New
67.3	64.030363	-24.079019	...	3.1103	New
22_D15.1	64.03448151	-24.06695655	...	3.2215	New
22_D15.2	64.034181	-24.066489	...	3.2215	New
22_D15.3	64.034006	-24.066447	...	3.2215	New
32_D15.1	64.045119	-24.072336	...	3.2882	New
32_D15.2	64.040081	-24.06673	...	3.2882	New
2_C16.a	64.050865	-24.066538	...	6.1452	New
2_C16.b	64.048179	-24.062406	...	6.1452	New
2_C16.c	64.043572	-24.059004	...	6.1452	New

**Table 3**  
(Continued)

ID	R.A. (deg)	Decl. (deg)	$z$	$z_{\text{new}}$	Training or New
6_C16.a	64.047781	-24.070169	...	3.6065	New
6_C16.b	64.043657	-24.064401	...	3.6065	New
6_C16.c	64.037679	-24.060756	...	3.6065	New
17_C16.a	64.040489	-24.07838	...	3.9663	New
17_C16.b	64.035107	-24.073864	...	3.9663	New
17_C16.c	64.027171	-24.068224	...	3.9663	New
22_C16.b	64.030997	-24.077173	...	3.923	New
22_C16.c	64.027127	-24.073572	...	3.923	New
23_C16.a	64.035668	-24.07992	...	2.5425	New
23_C16.b	64.032638	-24.078508	...	2.5425	New
33_C16.a	64.032017	-24.08423	...	5.9729	New
33_C16.b	64.030821	-24.083697	...	5.9729	New
34_C16.b	64.027632	-24.082609	...	3.9228	New
34_C16.c	64.023731	-24.078477	...	3.9228	New
35_C16.a	64.033681	-24.085855	...	5.639	New
35_C16.b	64.028654	-24.08424	...	5.639	New
35_C16.c	64.022187	-24.077559	...	5.639	New

**Note.** Each image is identified by its ID, right ascension (R.A.), and declination (Decl.) in degrees (J2000). The next columns are the spectroscopic redshifts ( $z$ ) that were available in 2015 (Jauzac et al. 2014; Richard et al. 2014; Treu et al. 2015), and the newly available spectroscopic redshifts ( $z_{\text{new}}$ ) from Caminha et al. (2017). The last column indicates whether the system is considered part of the training set or the new set.

## ORCID iDs

- J. D. Remolina González <https://orcid.org/0000-0002-7868-9827>  
K. Sharon <https://orcid.org/0000-0002-7559-0864>  
G. Mahler <https://orcid.org/0000-0003-3266-2001>

## References

- Acebron, A., Jullo, E., Limousin, M., et al. 2017, *MNRAS*, **470**, 1809  
Alavi, A., Siana, B., Richard, J., et al. 2016, *ApJ*, **832**, 56  
Atek, H., Richard, J., Jauzac, M., et al. 2015, *ApJ*, **814**, 69  
Atek, H., Richard, J., Kneib, J.-P., & Schaefer, D. 2018, *MNRAS*, **479**, 5184  
Balestra, I., Mercurio, A., Sartoris, B., et al. 2016, *ApJS*, **224**, 33  
Bayliss, M. B., Gladders, M. D., Oguri, M., et al. 2011, *ApJL*, **727**, L26  
Bayliss, M. B., Johnson, T., Gladders, M. D., Sharon, K., & Oguri, M. 2014, *ApJ*, **783**, 41  
Bonamigo, M., Grillo, C., Ettori, S., et al. 2017, *ApJ*, **842**, 132  
Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014, *ApJ*, **795**, 126  
Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, *ApJ*, **843**, 129  
Bradač, M., Clowe, D., Gonzalez, A. H., et al. 2006, *ApJ*, **652**, 937  
Bradač, M., Treu, T., Applegate, D., et al. 2009, *ApJ*, **706**, 1201  
Bradley, L. D., Zitrin, A., Coe, D., et al. 2014, *ApJ*, **792**, 76  
Broadhurst, T., Benéz, N., Coe, D., et al. 2005, *ApJ*, **621**, 53  
Caminha, G. B., Grillo, C., Rosati, P., et al. 2017, *A&A*, **600**, A90  
Cao, S., Biesiada, M., Gavazzi, R., Piórkowska, A., & Zhu, Z.-H. 2015, *ApJ*, **806**, 185  
Cerny, C., Sharon, K., Andrade-Santos, F., et al. 2018, *ApJ*, **859**, 159  
Chan, B. M. Y., Broadhurst, T., Lim, J., et al. 2017, *ApJ*, **835**, 44  
Chirivì, G., Suyu, S. H., Grillo, C., et al. 2018, *A&A*, **614**, 8  
Coe, D., Bradley, L., & Zitrin, A. 2015, *ApJ*, **800**, 84  
Coe, D., Zitrin, A., Carrasco, M., et al. 2013, *ApJ*, **762**, 32  
Dahle, H., Gladders, M. D., Sharon, K., et al. 2013, *ApJ*, **773**, 146  
D’Aloisio, A., Natarajan, P., & Shapiro, P. R. 2014, *MNRAS*, **445**, 3581  
Diego, J. M., Broadhurst, T., Benitez, N., et al. 2015a, *MNRAS*, **446**, 683  
Diego, J. M., Broadhurst, T., Chen, C., et al. 2016a, *MNRAS*, **456**, 356  
Diego, J. M., Broadhurst, T., Molnar, S. M., Lam, D., & Lim, J. 2015b, *MNRAS*, **447**, 3130  
Diego, J. M., Broadhurst, T., Wong, J., et al. 2016b, *MNRAS*, **459**, 3447  
Ebeling, H., Barrett, E., Donovan, D., et al. 2007, *ApJL*, **661**, L33  
Ebeling, H., Ma, C.-J., & Barrett, E. 2014, *ApJS*, **211**, 21  
Fohlmeister, J., Kochanek, C. S., Falco, E. E., et al. 2007, *ApJ*, **662**, 62  
Fohlmeister, J., Kochanek, C. S., Falco, E. E., et al. 2013, *ApJ*, **764**, 186  
Gade, K. 2010, *JNav*, **63**, 395  
Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, *ApJ*, **800**, 38  
Hoag, A., Huang, K.-H., Treu, T., et al. 2016, *ApJ*, **831**, 182  
Hoshen, J., & Kopelman, R. 1976, *PhRvB*, **14**, 3438  
Inada, N., Oguri, M., Morokuma, T., et al. 2006, *ApJL*, **653**, L97  
Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2015, *ApJ*, **799**, 12  
Jauzac, M., Clément, B., Limousin, M., et al. 2014, *MNRAS*, **443**, 1549  
Jauzac, M., Jullo, E., Kneib, J.-P., et al. 2012, *MNRAS*, **426**, 3369  
Jauzac, M., Richard, J., Limousin, M., et al. 2016, *MNRAS*, **457**, 2029  
Johnson, T. L., Rigby, J. R., Sharon, K., et al. 2017b, *ApJL*, **843**, L21  
Johnson, T. L., & Sharon, K. 2016, *ApJ*, **832**, 82  
Johnson, T. L., Sharon, K., Bayliss, M. B., et al. 2014, *ApJ*, **797**, 48  
Johnson, T. L., Sharon, K., Gladders, M. D., et al. 2017a, *ApJ*, **843**, 78  
Jones, T. A., Swinbank, A. M., Ellis, R. S., Richard, J., & Stark, D. P. 2010, *MNRAS*, **404**, 1247  
Jullo, E., & Kneib, J.-P. 2009, *MNRAS*, **395**, 1319  
Jullo, E., Kneib, J.-P., Limousin, M., et al. 2007, *NJPh*, **9**, 447  
Jullo, E., Natarajan, P., Kneib, J.-P., et al. 2010, *Sci*, **329**, 924  
Karman, W., Caputi, K. I., Caminha, G. B., et al. 2017, *A&A*, **599**, A28  
Kawamata, R., Ishigaki, M., Shimasaku, K., et al. 2018, *ApJ*, **855**, 4  
Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016, *ApJ*, **819**, 114  
Keeton, C. R. 2001, arXiv:[astro-ph/0102341](https://arxiv.org/abs/astro-ph/0102341)  
Keeton, C. R. 2011, GRAVELENS: Computational Methods for Gravitational Lensing, Astrophysics Source Code Library, ascl:[1102.003](https://ascl.net/1102.003)  
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2014, ATel, **6729**  
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2016, *ApJL*, **819**, L8  
Kneib, J.-P., Ellis, R. S., Santos, M. R., & Richard, J. 2004, *ApJ*, **607**, 697  
Kneib, J.-P., & Natarajan, P. 2011, *A&ARv*, **19**, 47  
Lam, D., Broadhurst, T., Diego, J. M., et al. 2014, *ApJ*, **797**, 98  
Laporte, N., Infante, L., Troncoso Iribarren, P., et al. 2016, *ApJ*, **820**, 98  
Liesenborgs, J., De Rijcke, S., & Dejonghe, H. 2006, *MNRAS*, **367**, 1209  
Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, *ApJ*, **835**, 113  
Livermore, R. C., Jones, T. A., Richard, J., et al. 2015, *MNRAS*, **450**, 1812  
Lotz, J. M., Koeckemoer, A., Coe, D., et al. 2017, *ApJ*, **837**, 97  
Magaña, J., Motta, V., Cárdenas, V. H., Verdugo, T., & Jullo, E. 2015, *ApJ*, **813**, 69  
Mahler, G., Richard, J., Clément, B., 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., Natarajan, P., Coe, D., et al. 2017, *MNRAS*, **472**, 3177  
Mohammed, I., Liesenborgs, J., Saha, P., & Williams, L. L. R. 2014, *MNRAS*, **439**, 2651  
Monna, A., Seitz, S., Balestra, I., et al. 2017, *MNRAS*, **466**, 4094  
Natarajan, P., Chadayammuri, U., Jauzac, M., et al. 2017, *MNRAS*, **468**, 1962

- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493  
 Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2015, *ApJ*, **808**, 104  
 Ogrean, G. A., van Weeren, R. J., Jones, C., et al. 2015, *ApJ*, **812**, 153  
 Oguri, M. 2010, *PASJ*, **62**, 1017  
 Oguri, M. 2015, *MNRAS*, **449**, L86  
 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  
 Richard, J., Kneib, J.-P., Limousin, M., Edge, A., & Jullo, E. 2010a, *MNRAS*, **402**, L44  
 Richard, J., Smith, G. P., Kneib, J.-P., et al. 2010b, *MNRAS*, **404**, 325  
 Richard, J., Stark, D. P., Ellis, R. S., et al. 2008, *ApJ*, **685**, 705  
 Rodney, S. A., Balestra, I., Bradac, M., et al. 2018, *NatAs*, **2**, 324  
 Rodney, S. A., Patel, B., Scolnic, D., et al. 2015, *ApJ*, **811**, 70  
 Rodney, S. A., Strolger, L.-G., Kelly, P. L., et al. 2016, *ApJ*, **820**, 50  
 Salmon, B., Coe, D., Bradley, L., et al. 2017, arXiv:1710.08930  
 Sharon, K., Bayliss, M. B., Dahle, H., et al. 2017, *ApJ*, **835**, 5  
 Sharon, K., Gladders, M. D., Marrone, D. P., et al. 2015, *ApJ*, **814**, 21  
 Sharon, K., Gladders, M. D., Rigby, J. R., et al. 2014, *ApJ*, **795**, 50  
 Sharon, K., & Johnson, T. L. 2015, *ApJL*, **800**, L26  
 Treu, T., Brammer, G., Diego, J. M., et al. 2016, *ApJ*, **817**, 60  
 Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, *ApJ*, **812**, 114  
 Vincenty, T. 1975, *Survey Review*, **23**, 88  
 Yue, B., Castellano, M., Ferrara, A., et al. 2017, arXiv:1711.05130  
 Zitrin, A., & Broadhurst, T. 2016, *ApJ*, **833**, 25  
 Zitrin, A., Broadhurst, T., Umetsu, K., et al. 2009, *MNRAS*, **396**, 1985  
 Zitrin, A., Meneghetti, M., Umetsu, K., et al. 2013, *ApJL*, **762**, L30  
 Zitrin, A., Zheng, W., Broadhurst, T., et al. 2014, *ApJL*, **793**, L12