Draft version October 21, 2015
Preprint typeset using LaTeX style emulateapj v. 5/2/11

'REFSDAL' MEETS POPPER: COMPARING PREDICTIONS OF THE RE-APPEARANCE OF THE MULTIPLY IMAGED SUPERNOVA BEHIND MACSJ1149.5+2223

T. Treu[1,2], G. Brammer[3], J. M. Diego[4], C. Grillo[5], P. L. Kelly[6], M. Oguri[7,8,9], S. A. Rodney[10,11,12], P. Rosati[13], K. Sharon[14], A. Zitrin[15,12], I. Balestra[16], M. Bradač[17], T. Broadhurst[18,19], G. B. Caminha[13], M. Ishigaki[20,8], R. Kawamata[21], T. L. Johnson[14], A. Halkola, A. Hoag[17], W. Karman[22], A. Mercurio[23], K. B. Schmidt[24], L.-G. Strolger[3,25], and S. H. Suyu[26]

Draft version October 21, 2015

ABSTRACT

Supernova 'Refsdal', multiply imaged by cluster MACS1149.5+2223, represents a rare opportunity to make a true blind test of model predictions in extragalactic astronomy, on a time scale that is short compared to a human lifetime. In order to take advantage of this event, we produced seven gravitational lens models with five independent methods, based on Hubble Space Telescope (HST) Hubble Frontier Field images, along with extensive spectroscopic follow-up from *HST* and from the Very Large Telescope. We compare the model predictions and show that they agree reasonably well with the measured time delays and magnification ratios between the known images, even though these quantities were not used as input. This agreement is encouraging, considering that the models only provide statistical uncertainties, and do not include additional sources of uncertainties such as structure along the line of sight, cosmology, and the mass sheet degeneracy. We then present the model predictions for the other appearances of SN 'Refsdal'. A future image will reach its peak in the first half of 2016, while another image appeared between 1994 and 2004. The past image would have been too faint to be detected in archival images. The future image should be approximately one third as bright as the brightest known images and thus detectable in *HST* images, as soon as the cluster can be targeted again (beginning 2015 October 30). We will find out soon whether our predictions are correct.

Subject headings: gravitational lensing: strong

tt@astro.ucla.edu
[1] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095
[2] Packard Fellow
[3] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
[4] IFCA, Instituto de Física de Cantabria (UC-CSIC), Av. de Los Castros s/n, 39005 Santander, Spain
[5] Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark
[6] Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA
[7] Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8583, Japan
[8] Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[9] Research Center for the Early Universe, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[10] Department of Physics and Astronomy, University of South Carolina, 712 Main St., Columbia, SC 29208, USA
[11] Department of Physics and Astronomy, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA
[12] Hubble Fellow
[13] Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, via Saragat 1, I-44122, Ferrara, Italy
[14] Department of Astronomy, University of Michigan, 1085 S. University Avenue, Ann Arbor, MI 48109, USA
[15] California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125
[16] University Observatory Munich, Scheinerstrasse 1, D-81679 Munich, Germany
[17] University of California Davis, 1 Shields Avenue, Davis, CA 95616
[18] Fisika Teorikoa, Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea UPV/EHU
[19] IKERBASQUE, Basque Foundation for Science, Alameda Urquijo, 36-5 48008 Bilbao, Spain
[20] Institute for Cosmic Ray Research, The University of Tokyo, Kashiwa, Chiba 277-8582, Japan
[21] Department of Astronomy, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[22] Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, the Netherlands
[23] INAF, Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127 Bologna, Italy
[24] Department of Physics, University of California, Santa Barbara, CA 93106-9530, USA
[25] Department of Physics, Western Kentucky University, Bowling Green, KY 42101, USA
[26] Institute of Astronomy and Astrophysics, Academia Sinica, P.O. Box 23-141, Taipei 10617, Taiwan

1. INTRODUCTION

In 1964 Sjur Resfdal speculated that a supernova multiply imaged by a foreground massive galaxy could be used to measure distances and, therefore, the Hubble Constant (Refsdal 1964). The basic physics behind this phenomenon is very simple. According to Fermat's principle, in gravitational optics as in standard optics, multiple images form at the extrema of the excess arrival time (Schneider 1985; Blandford & Narayan 1986). The excess arrival time is the result of the competition between the geometric time delay and the Shapiro (1964) delay. The arrival time thus depends on the apparent position of the image on the sky as well as the gravitational potential. Since the arrival time is measured in seconds, while all the other lensing observables are measured in angles on the sky, their relationship depends on the angular diameter distance D. In the simplest case of single plane lensing the time delay between two images is proportional to the so-called time-delay distance, i.e. $D_d D_s (1+z_d)/D_{ds}$, where d and s represent the deflector and the source, respectively, and the so-called Fermat potential (see, e.g., Meylan et al. 2006; Treu 2010; Suyu et al. 2010).

Over the past decades many authors have highlighted the importance and applications of identifying such events (e.g., Kolatt & Bartelmann 1998; Holz 2001; Goobar et al. 2002; Bolton & Burles 2003; Oguri & Kawano 2003), computed rates and proposed search strategies (Linder et al. 1988; Sullivan et al. 2000; Oguri et al. 2003; Oguri & Marshall 2010), and identified highly magnified supernova (Quimby et al. 2014). Finally, 50 years after the initial proposal by Refsdal, the first multiply imaged supernova was discovered in November 2014 (Kelly et al. 2015) in Hubble Space Telescope (*HST*) images of the cluster MACSJ1149.5+2223 (Ebeling et al. 2007; Smith et al. 2009; Zitrin & Broadhurst 2009), taken as part of the Grism Lens Amplified Survey from Space (GLASS; GO-13459, PI Treu; Schmidt et al. 2014; Treu et al. 2015), and aptly nicknamed 'Refsdal'. SN 'Refsdal' was identified in difference imaging as four point-sources that were not present in earlier images taken as part of the CLASH survey (Postman et al. 2012). Luckily, the event was discovered just before the beginning of an intensive imaging campaign as part of the Hubble Frontier Field (HFF) initiative (Lotz et al. 2015, in preparation; Coe et al. 2015). Additional epochs were obtained as part of the FrontierSN program (GO-13790. PI: Rodney), and a director discretionary time program (GO/DD-14041. PI: Kelly). The beautiful images that have emerged (Figure 1) are an apt celebration of the international year of light and the one hundredth anniversary of the theory of general relativity (e.g., Treu & Ellis 2015).

The gravitational lensing configuration of the 'Refsdal' event is very remarkable. The supernova exploded in one arm of an almost face-on spiral galaxy that is multiply imaged and highly magnified by the cluster gravitational potential. Furthermore, the spiral arm hosting 'Refsdal' happens to be sufficiently close to a cluster member galaxy that four additional multiple images are formed with average separation of order arcseconds, i.e., typical of galaxy-scale strong lensing. This set of four images close together in an "Einstein cross" configuration is where 'Refsdal' has been detected so far (labeled S1-S4 in Figure 1). As we discuss below, the cluster-scale images are more separated in terms of their arrival time, with time delays that can be much longer than the duration of the event, and therefore it is consistent with the lensing interpretation that they have not been seen yet.

The original suggestion by Refsdal (1964) was to use such events to measure distances and therefore cosmological parameters, starting from the Hubble constant. While distances with interesting accuracy and precision have been obtained from gravitational time delays in galaxy scale systems lensing quasars (e.g., Suyu et al. 2014), it is premature to attempt this in the case of 'Refsdal'. The time delay is not yet known with precision comparable to that attained for lensed quasars (e.g., Tewes et al. 2013b), and the mass distribution of the cluster MACSJ1149.5+2223 is inherently much more complex than that of a single elliptical galaxy.

However, 'Refsdal' gives us a unique opportunity to test the current mass models of MACSJ1149.5+2223, by conducting a textbook-like falsifiable experiment (Popper 1992). All the models that have been published after the discovery of 'Refsdal' (Kelly et al. 2015; Oguri 2015; Sharon & Johnson 2015; Diego et al. 2015; Jauzac et al. 2015) predict that an additional image will form some time in the near future (near image 1.2 of the host galaxy, shown in Figure 1). It could appear as early as October 2015 or in a few years. The field of MACSJ1149.5+2223 is currently unobservable with *HST*, but observations will resume at the end of October 2015 as part of an approved cycle 23 program (GO-14199. PI: Kelly). We thus have the opportunity to carry out a true blind test of the models, if we act fast enough. This test is similar in spirit to the test of magnification models using supernova 'Tomas', a Type-Ia SN magnified by Abell 2744 (Rodney et al. 2015). The uniqueness of our test lies in the fact that it is based on the prediction of an event that has not happened yet and it is thus intrinsically blind and immune from experimenter bias.

The quality and quantity of data available to lens modelers have improved significantly since the discovery of 'Refsdal' and the publication of the first modeling papers. As part of the HFF and follow-up programs there are now significantly deeper multiband *HST* images. Spectroscopy for hundreds of sources in the field (Figure 1) is now available from *HST* grism data obtained as part of GLASS and 'Refsdal' follow-up (aimed primarily at typing the supernova; Kelly et al. 2015, in preparation), as well as from Multi Unit Spectroscopic Explorer (MUSE) Very Large Telescope (VLT) Director's Discretionary Time follow-up (PI: Grillo).

The timing is thus perfect to ask the question: "Given state of the art data and models, how accurately can we predict the arrival time and magnification of the next appearance of a multiply-imaged supernova?" Answering this question will give us an absolute measurement of the quality of present-day models, although one should keep in mind that this is a very specific test. The arrival time and especially the magnification of a point source depend strongly on the details of the gravitational potential in the vicinity of the images. Additional uncertainties on the time delay and magnification arise from the inhomogenous distribution of mass along the line of sight (Suyu et al. 2010; Collett et al. 2013; Greene et al. 2013), the mass-sheet degeneracy and its generalizations

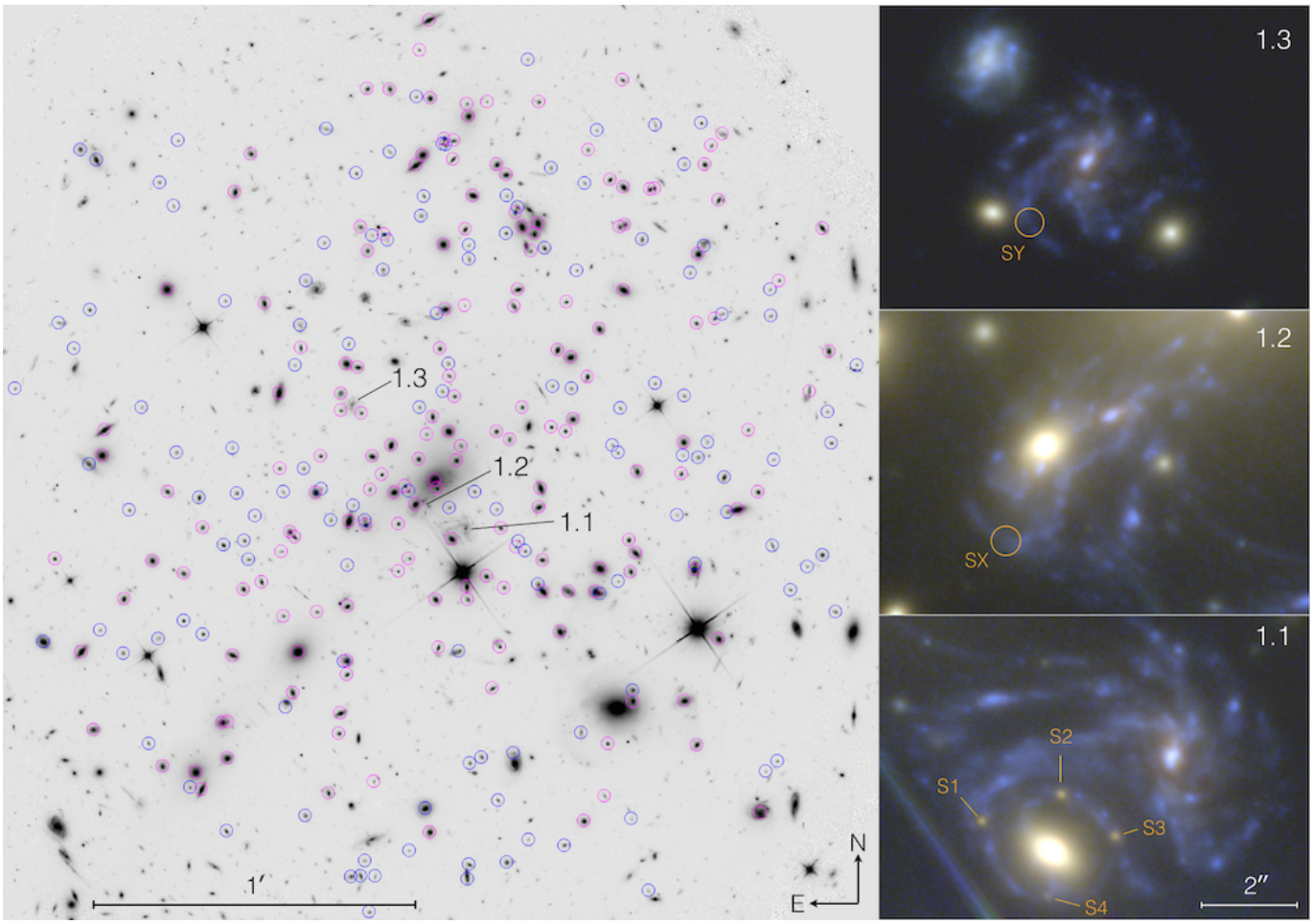

Figure 1. Multiple images of the SN 'Refsdal' host galaxy behind MACS1149. The left panel shows a wide view of the cluster, encompassing the entire footprint of the WFC3-IR camera. Spectroscopically confirmed cluster member galaxies are highlighted in magenta circles. Cyan circles indicate those associated with the cluster based on their photometric properties. The three panels on the right show in more detail the multiple images of the SN 'Refsdal' host galaxy (labeled 1.1 1.2 and 1.3). The positions of the known images of 'Refsdal' are labeled as S1-S4, while the model-predicted locations of the future and past appearance are labeled as SX and SY, respectively.

(Falco et al. 1985; Schneider & Sluse 2013, 2014; Suyu et al. 2014; Xu et al. 2015), and the residual uncertainties in cosmological parameters, especially the Hubble Constant (Riess et al. 2011; Freedman et al. 2012). Average or global quantities of more general interest, such as the total volume behind the cluster, or the average magnification, are much less sensitive to the details of the potential around a specific point.

In order to answer this question in the very short amount of time available, the 'Refsdal' follow-up team worked hard to reduce and analyze the follow-up data. By May 2015 it was clear that the quality of the follow-up data would be sufficient to make substantial improvements to their lens models. Therefore the follow-up team contacted the three other groups who had by then published predictions for 'Refsdal', and offered them the new datasets to update their models, as part of a concerted comparison effort. Thus, the five groups worked together to incorporate the new information into lensing analysis, first by identifying and rigorously vetting new sets of multiple images, and then to update their models in time to make a timely prediction. A synopsis and comparison between the results and predictions of the various models is presented in this paper. Companion papers by the individual groups will describe the follow-up campaigns as well as the details of each modeling effort.

This paper is organized as follows. In Section 2, we briefly summarize the datasets and measurements that are used in this comparison effort. In Section 3, we review the constraints used by the modeling teams. Section 4 gives a concise description of each of the five lens modeling techniques adopted. Section 5 presents the main results of this paper, i.e. a comparison of the predictions of the different models. Section 6 discusses the results, and Section 7 concludes with a summary. To ensure uniformity with the modeling effort for the Hubble Frontier Fields clusters, we adopt a concordance cosmology with $h = 0.7$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. All magnitudes are given in the AB system.

2. SUMMARY OF DATASETS AND MEASUREMENTS

We briefly summarize the datasets and measurements used in this paper. An overview of the field of view and pointing of the instruments used in this paper is shown in Figure 2.

2.1. *HST imaging*

Different versions of the images were used at different stages of the process. However, the final identification of multiple images and their positions were based on the HFF data release v1.0, and their world coordinate system. The reader is referred to the HFF data release webpages[27] for more information on this data.

2.1.1. *The light curves of SN 'Refsdal'*

Two teams measured the light curves of 'Refsdal' independently and derived initial measurements of the time delays and magnification ratios. The difference between the two measurements provides an estimate of the systematic uncertainties associated with the measurement, even though both measurements ignore effects like microlensing fluctuations (Dobler & Keeton 2006), and therefore this should be considered as a lower limit to the total uncertainty. A third effort (Rodney et al. 2015, in preparation) is under way to determine the time delays using methods developed for lensed quasars (Tewes et al. 2013a) that do not use a template for the light curve. Preliminary results from this third method indicate time delays and magnifications consistent with those presented here, albeit with larger uncertainties, as expected for the more flexibile procedure. More details about the supernova light curve, final measurements of time delays and magnification ratios and their uncertainties will be presented separately by each team in forthcoming publications.

The measurement of the first team (Kelly et al.), is based on the wide-band *F160W* (approximately rest-frame *R* band) WFC3-IR light curves from imaging taken between 2014 November 11 and 2015 July 21. The *F160W* photometry of S1–S4 was fit with the *R*-band light curves of SN 1987A (Hamuy & Suntzeff 1990) and three additional events showing similar luminosity evolution: NOOS-005 (*I* band)[28], SN 2006V (*r* band; Taddia et al. 2012), and SN 2009E (*R* band; Pastorello et al. 2012). The team first performed a spline interpolation of the comparison light curves, and then iteratively searched for the light-curve normalization and date of maximum that provide the best fit to the photometry of each SN image. The uncertainties in Table 1 are the standard deviation among the time delays and magnifications for the four light-curve templates. The peak brightness of image S1 occured approximately on 2015 April 26 (±20 days in the observer frame). We note that the light curve is very extended in time, and the time of the peak brightness is more uncertain than the relative time delay.

The second team (Strolger et al.) proceeded as follows. The multiple exposures on the target field of MACSJ1149+2223 were combined in visits, 2 to 4 exposure combinations by passband, each typically about 250, 1200, and 5000 seconds in total exposure time. Each visit-based filter combination was corrected to a rectified astrometric grid using DrizzlePac routines. Photometric measures were made with aperture photometry using the weighted average of circular apertures of $r = 2$, 3, and 4 pixels, corrected to infinite aperture magnitudes using encircled energy tables from Sirianni et al. (2005, ACS) and the WFC3 instrument handbook.

[27] http://www.stsci.edu/hst/campaigns/frontier-fields/
[28] http://ogle.astrouw.edu.pl/ogle3/ews/NOOS/2003/noos.html

Table 1
Measured time delays and magnification ratios

Image pair	Δt (K) (days)	μ ratio (K)	Δt (S) (days)	μ ratio (S)
S2 S1	-2.1±1.0	1.09±0.01	-9.0	1.06
S3 S1	5.6±2.2	1.04±0.02	-11	0.87
S4 S1	22±11	0.35±0.01	15.6	0.36

Note. — Observed delays and relative magnifications between the images S1–S4 of SN 'Refsdal'. For the values in column 2 and 3 Kelly et al. have fit the WFC3-IR *F160W* photometry of the images using the light curves of four separate 87A-like SN. The uncertainties listed are the standard deviation among the estimates made using the four light-curve templates, and may significantly underestimate the actual uncertainty. The values listed in column 4 and 5 are obtained independently by Strolger et al., and have similar uncertainties.

The light curve and spectral models in SNANA (Kessler et al. 2009) were used to construct multi-passband template light curves for five SN types (Ia, IIP, IIL/n, and Ib/c), corrected to appear as they would at the redshift of the event and through the observed passbands. An artificial SN 1987A-like model was then added, based on optical observations of SN 1987A (Hamuy & Suntzeff 1990), de-reddened by an $E(B-V) = 0.16$ (Fitzpatrick & Walborn 1990) and $R_V = 4.5$ (De Marchi & Panagia 2014) appropriate for the region of the Large Magellanic Cloud where SN 1987A appeared. The goodness of fit was evaluated through a least-squares fit (as χ^2_ν) to the multi-passband data for each image independently, with magnification and date of maximum light as free parameters. The slow rise of SN 'Refsdal' to maximum light (∼200 days observed, ∼80 days in the rest-frame *R*-passband) was seen early on to be generally inconsistent with the rise times for the common SN types, but broadly consistent with the rise of SN 1987A-like events, taking time delation into account. The best fit to all four images was found to be the SN 1987A-like template, with $\chi^2_\nu < 28$ for all images. The low quality of fit can be attributed principally to the difference in color between SN 1987A and the much bluer SN 'Refsdal', as well as the relative lack of flexibility in using a single template to represent a heterogeneous SN class. These fits were improved ($\chi^2_\nu < 10$) by adding non-positive extinction correction, with $A_V = -2.1$ (assuming $R_V = 4.05$).

2.2. *Spectroscopy*

2.2.1. *HST spectroscopy*

The *HST* grism spectroscopy is comprised of two datasets. The GLASS data consist of 10 orbits of exposures taken through the G102 grism and 4 orbits of exposures taken through the G141 grism, spanning the wavelength range $0.81 - 1.69\mu m$. The GLASS data were taken at two approximately orthogonal position angles to mitigate contamination by nearby sources (the first one in 2014 February 23-25, the second PA in 2014 November 3-11). The 'Refsdal' follow-up effort was focused on the G141 grism, reaching a depth of 30 orbits. The pointing and position angle of the follow-up grism data were

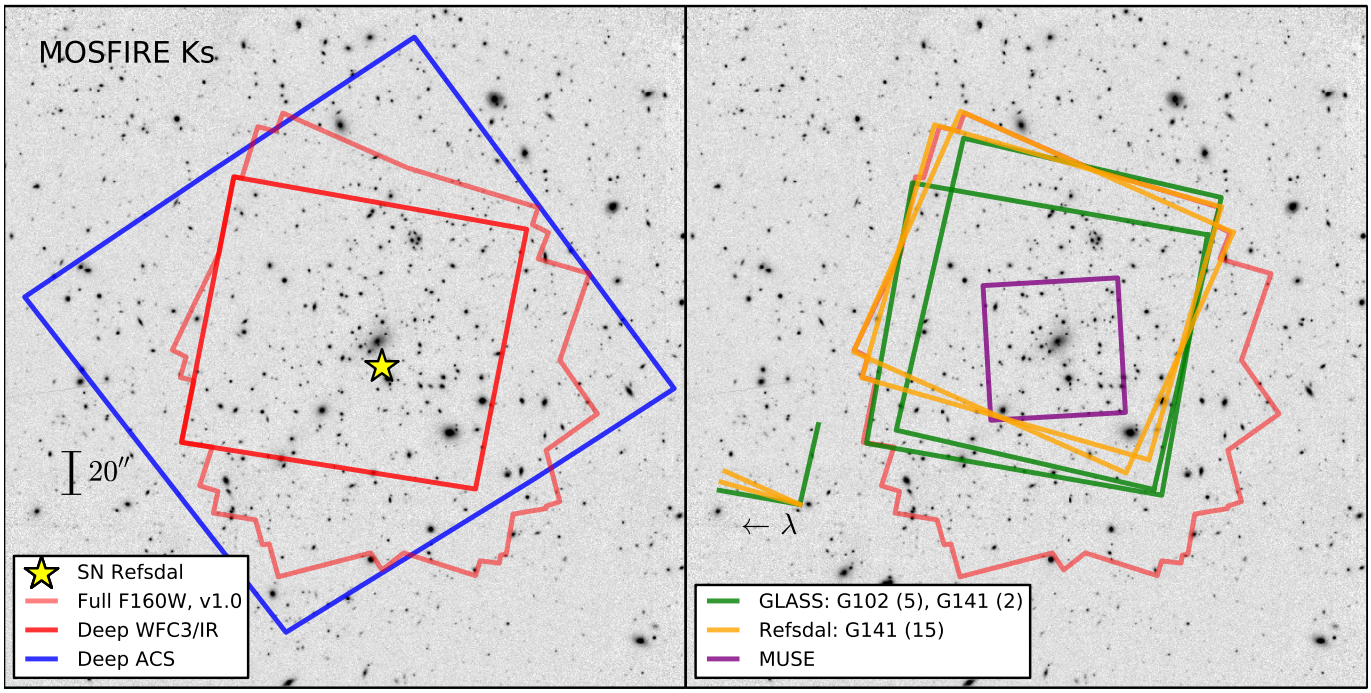

Figure 2. Observational layout of the MUSE and *HST* spectroscopy in the context of existing imaging data for MACSJ1149.5+2223. The "Full F160W" polygon is the full footprint of the F160W v1.0 FF release image. The numbers in parentheses in the spectroscopy panel at the right are the number of orbits per grism in each of two orients. The background image has been taken with the MOSFIRE instrument on the W.M.Keck-I Telescope (Brammer et al. 2015, in preparation).

chosen to optimize the spectroscopy of the supernova itself, and are therefore different from the ones adopted by GLASS. The 'Refsdal' follow-up spectra were taken between 2014 December 23 and 2015 January 4. Only a brief description of the data is given here. For more details the reader is referred to Schmidt et al. (2014) and Treu et al. (2015) for GLASS, and Brammer et al. (2015, in preparation) and Kelly et al. (2015, in preparation) for the deeper follow-up data.

The observing strategies and data reduction schemes were very similar for the two datasets, building on previous work by the 3D-HST survey (Brammer et al. 2012). At least 4 sub-exposures were taken during each visit with semi-integer pixel offsets. This enables rejection of defects and cosmic rays as well as recovery of some of the resolution lost to undersampling of the PSF through interlacing. The data were reduced with an updated version of the 3D-HST reduction pipeline[29] described by Brammer et al. (2012) and Momcheva et al. (2015). The pipeline takes care of alignment, defect removal, background removal, image combination, and modeling of contamination by nearby sources. One and two dimensional spectra are extracted for each source.

The spectra were inspected independently by two of us (T.T. and G.B.) using custom tools and the interfaces GiG and GiGz (available at `https://github.com/kasperschmidt/GLASSinspectionGUIs`) developed as part of the GLASS project. Information obtained from the multiband photometry, continuum, and emission line was combined to derive a redshift and quality flag. The few discrepancies between redshifts and quality flags were resolved by mutual agreement. In the end, we determined redshifts for 389 sources, with quality 3 or 4 (probable or secure, respectively, as defined by Treu et al. 2015).

[29] http://code.google.com/p/threedhst/

Table 2
Redshift catalog

ID*	RA (J2000)	DEC (J2000)	z	quality	source	Notes
1	177.397188	22.393744	0.0000	4	2	···
2	177.404017	22.403067	0.5660	4	2	···
3	177.394525	22.400653	0.5410	4	2	···
4	177.399663	22.399597	0.5360	4	2	···
5	177.404054	22.392108	0.0000	4	2	···
6	177.398554	22.389792	0.5360	4	2	···
7	177.393010	22.396799	2.9490	4	2	···
8	177.394400	22.400761	2.9490	4	2	···
9	177.404192	22.406125	2.9490	4	2	···
10	177.392904	22.404014	0.5140	4	2	···

Note. — First entries of the redshift catalog. The full catalog is given in its entirety in the electronic edition. The column "quality" contains the quality flag (3=secure, 4=probable). The column "source" gives the original source of the redshift (1=*HST*, Brammer et al. 2015, in prep; 2=MUSE, Grillo et al. 2015, in prep; 3=both). The column "note" lists special comments about the object, e.g. if the object is part of a known multiply image system.

2.2.2. *VLT spectroscopy*

Integral field spectroscopy was obtained with the MUSE instrument on the VLT between 2015 February 14 and 2015 April 12, as part of a Director Discretionary Time program to followup 'Refsdal' (PI: Grillo). The main goal of the program was to facilitate the computation of an accurate model to forecast the next appearance of the lensed SN. MUSE covers the wavelength range 480-

930 nm, with an average spectral resolution of $R \sim 3000$, over a $1 \times 1\,\mathrm{arcmin}^2$ field of view, with a pixel scale of $0\farcs2$ px^{-1}. Details of the data acquisition and processing are given in a separate paper (Grillo et al. 2015, in preparation). Only a brief summary of relevant information is given here to guide the reader.

Twelve exposures were collected in dark time under clear conditions and with an average seeing of approximately $1\farcs0$. Bias subtraction, flatfielding, wavelength and flux calibrations were obtained with the MUSE Data Reduction Software version 1.0, as described by Karman et al. (2015a,b). The different exposures were combined into a single datacube, with a spectral sampling of 1.25 Å px^{-1}, and a resulting total integration time of 4.8 hours. 1D spectra within circular apertures of $0\farcs6$ radius were extracted for all the objects visible in the coadded image along the spectral direction. We also searched in the datacube for faint emission line galaxies that were not detected in the stacked image. Redshifts were first measured independently by two of the co-authors (W.K. and I.B.) and later reconciled in the very few cases with inconsistent estimates. The analysis yielded secure redshift values for 111 objects, of which 15 are multiple images of 6 different background sources.

Table 3
Multiply imaged systems

ID	R.A. (J2000)	Decl. (J2000)	Z09	S09	R14, J14	D15	Spec z	ref	Spec z	source	Notes	Avg. Score	Category
1.1	177.39700	22.396000	1.2	A1.1	1.1	1.1	1.4906	S09	1.488	3	···	1.0	Gold
1.2	177.39942	22.397439	1.3	A1.2	1.2	1.2	1.4906	S09	1.488	3	···	1.0	Gold
1.3	177.40342	22.402439	1.1	A1.3	1.3	1.3	1.4906	S09	1.488	3	···	1.0	Gold
1.5	177.39986	22.397133	1.4	···	···	1.5	···	···	···	···	1	2.0	···
2.1	177.40242	22.389750	3.3	A2.1	2.1	2.3	1.894	S09	1.891	3	···	1.0	Gold
2.2	177.40604	22.392478	3.2	A2.2	2.2	2.2	1.894	S09	1.891	3	···	1.0	Gold
2.3	177.40658	22.392886	3.1	A2.3	2.3	2.1	1.894	S09	1.891	3	···	1.0	Gold
3.1	177.39075	22.399847	2.1	A3.1	3.1	3.1	2.497	S09	3.129	3	2	1.0	Gold
3.2	177.39271	22.403081	2.2	A3.2	3.2	3.2	2.497	S09	3.129	3	2	1.0	Gold
3.3	177.40129	22.407189	2.3	A3.3	3.3	3.3	···	···	3.129	3	2	1.1	Gold
4.1	177.39300	22.396825	4.1	···	4.1	4.1	···	···	2.949	2	···	1.0	Gold
4.2	177.39438	22.400736	4.2	···	4.2	4.2	···	···	2.949	2	···	1.0	Gold
4.3	177.40417	22.406128	4.3	···	4.3	4.3	···	···	2.949	2	···	1.0	Gold
5.1	177.39975	22.393061	5.1	···	5.1	5.1	···	···	2.80	1	···	1.0	Gold
5.2	177.40108	22.393825	5.2	···	5.2	5.2	···	···	···	···	···	1.0	Gold
5.3	177.40792	22.403553	5.3	···	5.3	···	···	···	···	···	···	1.7	Silver
6.1	177.39971	22.392544	6.1	···	6.1	6.1	···	···	···	···	···	1.1	Gold
6.2	177.40183	22.393858	6.2	···	6.2	6.2	···	···	···	···	···	1.1	Gold
6.3	177.40804	22.402506	5.4/6.3	···	6.3	···	···	···	···	···	···	1.7	Silver
7.1	177.39896	22.391339	7.1	···	7.1	···	···	···	···	···	···	1.1	Gold
7.2	177.40342	22.394269	7.2	···	7.2	···	···	···	···	···	···	1.1	Gold
7.3	177.40758	22.401242	···	···	7.3	···	···	···	···	···	···	1.2	Gold
8.1	177.39850	22.394350	8.1	···	8.1	8.1	···	···	···	···	···	1.2	Gold
8.2	177.39979	22.395044	8.2	···	8.2	8.2	···	···	···	···	···	1.2	Gold
8.4	177.40709	22.404722	···	···	···	···	···	···	···	···	3	1.2	Gold
···	177.40704	22.405553	···	···	···	···	···	···	2.78	1	3	3.0	Rejected
···	177.40517	22.401563	8.3	···	···	···	···	···	···	···	3	3.0	Rejected
9.1	177.40517	22.426233	···	A6.1	9.1	···	···	···	···	···	···	1.8	···
9.2	177.40388	22.427231	···	A6.2	9.2	···	···	···	···	···	···	1.8	···
9.3	177.40325	22.427228	···	A6.3	9.3	···	···	···	···	···	···	1.8	···
9.4	177.40364	22.426422	···	A6.4?	···	···	···	···	···	···	···	1.8	···
10.1	177.40450	22.425514	···	A7.1	10.1	···	···	···	···	···	···	1.8	···
10.2	177.40362	22.425636	···	A7.2	10.2	···	···	···	···	···	···	1.8	···
10.3	177.40221	22.426625	···	A7.3	10.3	···	···	···	···	···	···	1.8	···
12.1	177.39857	22.389356	···	···	···	···	···	···	1.020	3	4	2.6	Rejected
12.2	177.40375	22.392345	···	···	···	···	···	···	0.929	2	4	2.9	Rejected
12.3	177.40822	22.398801	···	···	···	···	···	···	1.118	3	4	2.6	Rejected
13.1	177.40371	22.397786	···	···	13.1	···	···	···	1.23	1	···	1.0	Gold
13.2	177.40283	22.396656	···	···	13.2	···	···	···	1.25	1	···	1.0	Gold
13.3	177.40004	22.393858	···	···	13.3	···	···	···	1.23	1	···	1.3	Gold
14.1	177.39167	22.403489	···	···	14.1	···	···	···	3.703	2	···	1.3	Gold
14.2	177.39083	22.402647	···	···	14.2	···	···	···	3.703	2	···	1.3	Gold
110.1	177.40014	22.390162	···	···	···	···	···	···	3.214	2	···	1.0	Gold
110.2	177.40402	22.392894	···	···	···	···	···	···	3.214	2	···	1.0	Gold
110.3	177.40907	22.400242	···	···	···	···	···	···	···	···	···	2.0	···
21.1	177.40451	22.386704	···	···	···	···	···	···	···	···	···	1.8	Silver
21.2	177.40800	22.389057	···	···	···	···	···	···	···	···	···	1.6	Silver
21.3	177.40907	22.390407	···	···	···	···	···	···	···	···	···	1.6	Silver
22.1	177.40370	22.386838	···	···	···	···	···	···	···	···	···	1.8	···
22.2	177.40791	22.389232	···	···	···	···	···	···	···	···	···	1.8	···
22.3	177.40902	22.391053	···	···	···	···	···	···	···	···	···	1.8	···
23.1	177.39302	22.411428	···	A5	···	···	···	···	···	···	···	1.8	···
23.2	177.39308	22.411455	···	A5	···	···	···	···	···	···	···	1.8	···
23.3	177.39315	22.411473	···	A5	···	···	···	···	···	···	···	1.8	···
24.1	177.39285	22.412872	···	···	···	···	···	···	···	···	···	1.7	···
24.2	177.39353	22.413071	···	···	···	···	···	···	···	···	···	1.7	···
24.3	177.39504	22.412697	···	···	···	···	···	···	···	···	···	1.8	···
25.1	177.40428	22.398782	···	···	···	···	···	···	···	···	···	2.0	···

Table 3 — *Continued*

ID	R.A. (J2000)	Decl. (J2000)	Z09	S09	R14, J14	D15	Spec z	ref	Spec z	source	Notes	Avg. Score	Category
25.2	177.40411	22.398599	···	···	···	···	···	···	···	···	···	2.0	···
25.3	177.39489	22.391796	···	···	···	···	···	···	···	···	···	2.3	···
26.1	177.41035	22.388749	9.1	···	···	···	···	···	···	···	···	1.8	Silver
26.2	177.40922	22.387697	9.2	···	···	···	···	···	···	···	···	1.8	Silver
26.3	177.40623	22.385369	···	···	···	···	···	···	···	···	···	1.8	Silver
27.1	177.40971	22.387665	···	···	···	···	···	···	···	···	···	1.8	Silver
27.2	177.40988	22.387835	···	···	···	···	···	···	···	···	···	1.8	Silver
27.3	177.40615	22.385142	···	···	···	···	···	···	···	···	···	2.5	···
28.1	177.39531	22.391809	···	···	···	···	···	···	···	···	···	2.0	···
28.2	177.40215	22.396750	···	···	···	···	···	···	···	···	···	2.2	···
28.3	177.40562	22.402434	···	···	···	···	···	···	···	···	···	2.0	···
200.1	177.40875	22.394467	···	···	···	···	···	···	2.32	1	···	2.6	···
200.2	177.40512	22.391261	···	···	···	···	···	···	···	···	···	2.6	···
200.3	177.40256	22.389233	···	···	···	···	···	···	···	···	···	2.8	···
201.1	177.40048	22.395444	···	···	···	···	···	···	···	···	5	1.6	···
201.2	177.40683	22.404517	···	···	···	···	···	···	···	···	5	1.6	···
202.1	177.40765	22.396789	···	···	···	···	···	···	···	···	···	2.0	···
202.2	177.40224	22.391489	···	···	···	···	···	···	···	···	···	2.0	···
202.3	177.40353	22.392586	···	···	···	···	···	···	···	···	···	2.0	···
203.1	177.40995	22.387244	···	···	···	···	···	···	···	···	···	1.8	Silver
203.2	177.40657	22.384511	···	···	···	···	···	···	···	···	···	2.0	Silver
203.3	177.41123	22.388461	···	···	···	···	···	···	···	···	···	1.8	Silver
204.1	177.40961	22.386661	···	···	···	···	···	···	···	···	···	1.8	Silver
204.2	177.40668	22.384322	···	···	···	···	···	···	···	···	···	1.8	Silver
204.3	177.41208	22.389056	···	···	···	···	···	···	···	···	···	1.8	Silver
205.1	177.40520	22.386042	···	···	···	···	···	···	···	···	···	2.0	···
205.2	177.40821	22.388119	···	···	···	···	···	···	···	···	···	2.0	···
205.3	177.41038	22.390625	···	···	···	···	···	···	···	···	···	2.0	···
206.1	177.40764	22.385647	···	···	···	···	···	···	···	···	···	2.2	···
206.2	177.40863	22.386453	···	···	···	···	···	···	···	···	···	2.2	···
206.3	177.41133	22.388997	···	···	···	···	···	···	···	···	···	2.2	···
207.1	177.40442	22.397303	···	···	···	···	···	···	···	···	···	2.2	···
207.2	177.40397	22.396039	···	···	···	···	···	···	···	···	···	2.2	···
208.1	177.40453	22.395761	···	···	···	···	···	···	···	···	···	2.0	···
208.2	177.40494	22.396397	···	···	···	···	···	···	···	···	···	2.0	···
209.1	177.38994	22.412694	···	···	···	···	···	···	···	···	···	3.0	···
209.2	177.39055	22.413408	···	···	···	···	···	···	···	···	···	3.0	···
210.1	177.39690	22.398061	···	···	···	···	···	···	0.702	2	···	3.0	···
210.2	177.39505	22.397497	···	···	···	···	···	···	0.702	2	···	3.0	···

Note. — Coordinates and ID notations of multiply-imaged families of lensed galaxies. The labels in previous publications are indicated for Zitrin et al. (2009; Z09), Smith et al. (2009; S09), Richard et al. (2014; R14), Johnson et al. (2014; J14) and Diego et al. (2015; D15). New identifications were made by Sharon, Oguri, and Hoag. Each modeling team used a modified version or subset of the list above, with coordinates of each knots varying slightly between modelers. The source of the new spectroscopic redshift is as in Table 2 (1=HST, Brammer et al. 2015, in prep; 2=MUSE, Grillo et al. 2015, in prep; 3=both). The average score among the team is recorded, with "1" denotes secore identification, "2" is a possible identification, and higher score are considered unreliable by the teams.

[1] See Table 4 for information on all the knots in source 1.

[2] We revise the redshift of source 3 with the new and reliable measurement from MUSE (see § 2.2).

[3] We revise the identification of a counter image of 8.1 and 8.2, and determine that it is at a different position compared to previous publications. To limit confusion we label the newly identified counter image 8.4.

[4] The identification of source 12 was ruled out in HFF work prior to the 2014 publications; we further reject this set with spectroscopy.

[5] This image is identified as part of the same source as source 8; the third image is buried in the light of a nearby star.

2.2.3. *Combined redshift catalog*

Redshifts for 70 objects were measured independently using both MUSE and GLASS data. We find that the redshifts of all objects in common agree within the uncertainties, attesting to the excellent quality of the data. The final redshift catalog, comprising of 429 entries, is given in electronic format in in Table 2, and is available through the GLASS public website at URL `https://archive.stsci.edu/prepds/glass/`. We note that owing to the high resolution of the MUSE data we improved the precision of the redshift of the Refsdal host galaxy to $z = 1.488$ (c.f. 1.491 previously reported by Smith et al. 2009). Also, we revise the redshift of the multiply imaged source 3 with the new and reliable measurement $z = 3.129$ based on unequivocal multiple line identifications ([OII] in the grism data, plus Lymanα in the MUSE data).

3. SUMMARY OF LENS MODELING CONSTRAINTS

3.1. *Multiple images*

The strong lensing models that are considered in this paper use as constraints sets of multiply-imaged lensed galaxies, as well as knots in the host galaxy of SN 'Refsdal'. The five teams independently evaluated known sets of multiple images (Zitrin & Broadhurst 2009; Smith et al. 2009; Johnson et al. 2014; Sharon & Johnson 2015; Diego et al. 2015), and suggested new identifications of image across the entire field of view, based on the new HFF data. In evaluating the image identifications, the teams relied on their preliminary lens models and the newly measured spectroscopic redshifts (Section 2.2). Each team voted on known and new system on a scale of 1–4, where 1 denotes secure identification, 2 is a possible identification, and higher values are considered unreliable. Images that had large variance in their scores were discussed and re-evaluated, and the final score was then recorded. The list of multiple images considered in this work is given in Table 3. For each system we give coordinates, average score, and redshift if available. We also indicate the labels given to known images that were previously identified in the literature, previously published redshifts, and references to these publications.

We define three samples of image sets, "gold", "silver", and "all", based on the voting process. Following the approach of Wang et al. (2015), we conservatively include in our "gold" sample only the systems that every team was confident about. The "silver" sample includes images that were considered secure by most teams, or are outside the MUSE field of view. The "all" sample includes all the images that were not rejected as false identification, based on imaging and/or spectroscopy. In order to facilitate the comparison, most teams produced baseline models based on the "gold" sample of images, and some of the teams produced additional models based on larger sets of images. However, owing to differences in investigator's opinions and specifics of each code, small differences between the constraints adopted by each team persist. They are described below for each of the teams. The reader is referred to the publications of each individual team for more details.

We also evaluated the identification of knots in the grand design spiral galaxy hosting 'Refsdal'. Table 4 and Figure 3 list the emission knots and features in the host galaxy of SN 'Refsdal' that were considered in this work. Not all knots were used in all models, and again, there are slight differences between the teams as the implementation of these constraints vary among lensing algorithms. Nevertheless, the overall mapping of morphological features between the images of this galaxy was in agreement between the modeling teams.

3.2. *Time delays*

The time delay and magnification ratios between the known images were not yet measured at the time when the models were being finalized. Therefore they were not used as input and they can be considered as a valuable test of the lens model.

3.3. *Cluster members*

Cluster member galaxies were selected based on their redshifts in the combined redshift catalog and their photometry, as follows. In order to account for the cluster velocity dispersion, as well as the uncertainty on the grism-based redshifts, we define cluster membership loosely as galaxies with spectroscopic redshift in the range $0.520 < z < 0.570$, i.e. within a few thousand kilometers per second of the fiducial cluster redshift ($z = 0.542$). This is sufficiently precise for the purpose of building lens models, even though not all the cluster members are necessarily physically bound to the cluster, from a dynamical point of view. Naturally, these cluster members still contribute to the deflection field as the dynamically bound cluster members. The spectroscopic cluster-member catalog comprises 170 galaxies.

To obtain a more complete member catalog, the spectroscopically-confirmed members were supplemented by photometrically selected galaxies. This list includes galaxies down to the magnitude limit (F814W$\sim$25) of spectroscopically confirmed members. It is constituted mostly of galaxies belonging to the last two-magnitude bins of the luminosity distribution, for which the spectroscopic sample is significantly incomplete. The missing galaxies from the spectroscopic catalog are the brightest ones that fall outside the MUSE field of view or the ones that are contaminated in the *HST* grism data. The photometric analysis is restricted to the WFC3-IR area, in order to exploit the full multi-band photometric catalog from CLASH. The method is briefly described by Grillo et al. (2015), and it uses a Bayesian technique to compute the probability for a galaxy to be a member from the distribution in color space of all spectroscopic galaxies (from 13 bands, i.e. not including the 3 in the UV). For the photometric selection, we started from spectroscopically confirmed members, with redshift within $0.520 < z < 0.570$, and provided a catalog with only the objects with measured F160W magnitudes. The total catalog of cluster members comprises 170 galaxies with spectroscopically determined membership, and 136 galaxies with photometrically determined membership.

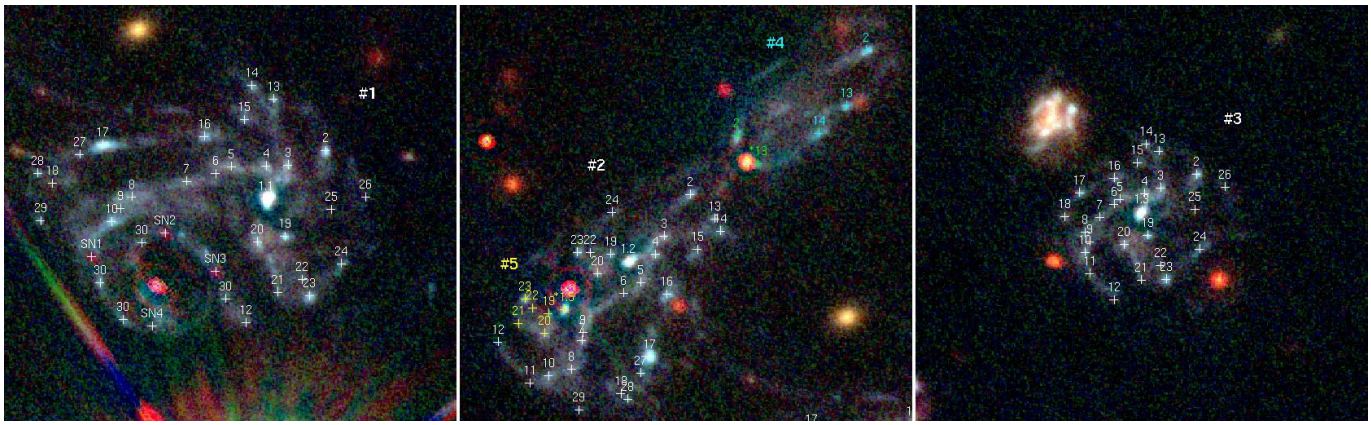

Figure 3. Knots and morphological features in the host galaxy of SN 'Refsdal' at $z = 1.488$. The color composite on which the regions are overplotted is generated by scaling and subtracting the F814W image from the F435W, F606W, and F105W images, in order to suppress the light from the foreground cluster galaxies. The left panel shows image 1.1, and the right panel shows image 1.3. In the middle panel, the complex lensing potential in the central region is responsible for one full image, 1.2, and additional partial images of the galaxy, 1.4, and 1.5 (see also Smith et al. 2009, Zitrin et al. 2009, and Sharon & Johnson 2015). To guide the eye, we label knots that belong to 1.4 and 1.5 in cyan and yellow, respectively. A possible sixth image of a small region of the galaxy is labeled in green. The two features marked with an asterisks in this panel, *1.5 and *13, are the only controversial identifications. We could not rule out the identification of *1.5 (knot 1.1.5 in Table 4) as counterpart of the bulge of the galaxy, however, it is likely only partly imaged. Image *13 (1.13.6 in Table 4) is suggested by some of the models, but hard to confirm, and is thus not used as constraints in the "gold" lens models considered here. We note that the exact coordinates of each feature may vary slightly between modelers, and we refer the reader to detailed publications (in preparation) by each modeling team for exact positions and features used.

Table 4
Knots in the host galaxy of 'Refsdal'

ID	R.A. (J2000)	Decl. (J2000)	ID Smith et al. (2009)	ID Sharon et al. (2015)	ID Diego et al. (2015)	Notes
1.1.1	177.39702	22.396003	2	1.1	1.1.1	1
1.1.2	177.39942	22.397439	2	1.2	1.2.1	1
1.1.3	177.40341	22.402444	2	1.3	1.3.1	1
1.*1.5	177.39986	22.397133	···	···	1.5.1	1,2
1.2.1	177.39661	22.396308	19	23.1	1.1.8	···
1.2.2	177.39899	22.397867	19	23.2	1.2.8	···
1.2.3	177.40303	22.402681	19	23.3	1.3.8	···
1.2.4	177.39777	22.398789	19	23.4	1.4.8a	···
1.2.6	177.39867	22.398242	···	···	1.4.8b	···
1.3.1	177.39687	22.396219	16	31.1	1.1.15	···
1.3.2	177.39917	22.397600	16	31.2	1.2.15	···
1.3.3	177.40328	22.402594	16	31.3	1.3.15	···
1.4.1	177.39702	22.396214	11	32.1	···	···
1.4.2	177.39923	22.397483	11	32.2	···	···
1.4.3	177.40339	22.402558	11	32.3	···	···
1.5.1	177.39726	22.396208	18	33.1	···	···
1.5.2	177.39933	22.397303	18	33.2	···	···
1.5.3	177.40356	22.402522	18	33.3	···	···
1.6.1	177.39737	22.396164	···	···	1.1.13	···
1.6.2	177.39945	22.397236	···	···	1.2.13	···
1.6.3	177.40360	22.402489	···	···	1.3.13	···
1.7.1	177.39757	22.396114	···	40.1	···	···
1.7.2	177.39974	22.396933	···	40.2	···	···
1.7.3	177.40370	22.402406	···	40.3	···	···
1.8.1	177.39795	22.396014	···	···	···	···
1.8.2	177.39981	22.396750	···	···	···	···
1.8.3	177.40380	22.402311	···	···	···	···
1.9.1	177.39803	22.395939	···	···	1.1.9	···
1.9.2	177.39973	22.396983	···	···	1.2.9	···
1.9.3	177.40377	22.402250	···	···	1.3.9	···
1.10.1	177.39809	22.395856	···	···	···	···
1.10.2	177.39997	22.396708	···	36.2	···	···
1.10.3	177.40380	22.402183	···	36.3	···	···
1.11.2	177.40010	22.396661	···	···	1.2.3	···
1.11.3	177.40377	22.402047	···	···	1.3.3	···
1.12.1	177.39716	22.395211	···	···	1.1.14	···
1.12.2	177.40032	22.396925	···	···	1.2.14	···
1.12.3	177.40360	22.401878	···	···	1.3.14	···
1.13.1	177.39697	22.396639	7	24.1	1.1.19	···
1.13.2	177.39882	22.397711	7	24.2	1.2.19	···
1.13.3	177.40329	22.402828	7	24.3	1.3.19	···
1.13.4	177.39791	22.398433	7	24.4	1.4.19	···
1.*13.6	177.39852	22.398061	···	···	···	3
1.14.1	177.39712	22.396725	6	25.1	1.1.7	···
1.14.2	177.39878	22.397633	6	25.2	1.2.7	···

Table 4 — *Continued*

ID	R.A. (J2000)	Decl. (J2000)	ID Smith et al. (2009)	ID Sharon et al. (2015)	ID Diego et al. (2015)	Notes
1.14.3	177.40338	22.402872	6	25.3	1.3.7	···
1.14.4	177.39810	22.398256	···	25.4	1.4.7	···
1.15.1	177.39717	22.396506	···	41.1	1.1.20	···
1.15.2	177.39894	22.397514	···	41.2	1.2.20	···
1.15.3	177.40344	22.402753	···	41.3	1.3.20	···
1.16.1	177.39745	22.396400	4	26.1	1.1.6	···
1.16.2	177.39915	22.397228	4	26.2	1.2.6	···
1.16.3	177.40360	22.402656	4	26.3	1.3.6	···
1.17.1	177.39815	22.396347	3	11.1	1.1.5	···
1.17.2	177.39927	22.396831	3	11.2	1.2.5	···
1.17.3	177.40384	22.402564	3	11.3	1.3.5	···
1.18.1	177.39850	22.396100	···	···	1.1.11	···
1.18.2	177.39947	22.396592	···	···	1.2.11	···
1.18.3	177.40394	22.402408	···	···	1.3.11	···
1.19.1	177.39689	22.395761	···	21.1	1.1.17	···
1.19.2	177.39954	22.397486	···	21.2	1.2.17	···
1.19.3	177.40337	22.402292	···	21.3	1.3.17	···
1.19.5	177.39997	22.397106	···	21.4	1.5.17	···
1.20.1	177.39708	22.395728	···	27.1	1.1.16	···
1.20.2	177.39963	22.397361	···	···	1.2.16	···
1.20.3	177.40353	22.402233	···	27.3	1.3.16	···
1.20.5	177.40000	22.396981	···	27.2	1.5.16	···
1.21.1	177.39694	22.395406	···	···	1.1.18	···
1.21.3	177.40341	22.402006	···	···	1.3.18	···
1.21.5	177.40018	22.397042	···	···	1.5.18	···
1.22.1	177.39677	22.395487	···	···	···	···
1.22.2	177.39968	22.397495	···	···	···	···
1.22.3	177.40328	22.402098	···	···	···	···
1.22.5	177.40008	22.397139	···	···	···	···
1.23.1	177.39672	22.395381	15	22.1	1.1.2	···
1.23.2	177.39977	22.397497	15	22.2	1.2.2	···
1.23.3	177.40324	22.402011	15	22.3	1.3.2	···
1.23.5	177.40013	22.397200	···	22.2	1.5.2	···
1.24.1	177.39650	22.395589	···	28.1	1.1.4	···
1.24.2	177.39953	22.397753	···	28.2	1.2.4	···
1.24.3	177.40301	22.402203	···	28.3	1.3.4	···
1.25.1	177.39657	22.395933	···	···	1.1.21	···
1.25.3	177.40304	22.402456	···	···	1.3.21	···
1.27.1	177.39831	22.396285	···	37.1	···	···
1.27.2	177.39933	22.396725	···	37.2	···	···
1.26.1	177.39633	22.396011	···	···	1.1.12	···
1.26.3	177.40283	22.402600	···	···	1.3.12	···
1.28.1	177.39860	22.396166	···	38.1	···	···
1.28.2	177.39942	22.396559	···	38.2	···	···
1.29.1	177.39858	22.395860	···	39.1	···	···
1.29.2	177.39976	22.396490	···	39.2	···	···
1.30.1	177.39817	22.395465	···	35.1	···	···
1.30.2	177.39801	22.395230	···	35.2	···	···
1.30.3	177.39730	22.395364	···	35.3	···	···
1.30.4	177.39788	22.395721	···	35.4	···	···
SN1	177.39823	22.395631	···	30.1	1.1.3a	···
SN2	177.39772	22.395783	···	30.2	1.1.3b	···
SN3	177.39737	22.395539	···	30.3	1.1.3c	···
SN4	177.39781	22.395189	···	30.4	1.1.3d	···

Note. — Coordinates and ID notations of emission knots in the multiply-imaged host of SN Refsdal, at $z = 1.488$. The labels in previous publications are indicated. New identifications were made by C.G., K.S., and J.D. Each modeling team used a modified version or subset of the list above, with coordinates of each knots varying slightly between modelers. Nevertheless, there is consensus among the modelers on the identification and mapping of the different features between the multiple images of the same source.

[1] Images 1.1, 1.2, 1.3, 1.5 were labeled by Zitrin & Broadhurst (2009) as 1.2, 1.3, 1.1, 1.4, respectively. The labels of other knots were not given in that publication.

[2] This knot was identified as a counter image of the bulge of the galaxy by Zitrin & Broadhurst (2009), but rejected by Smith et al. (2009). As in the paper by Sharon & Johnson (2015), the modelers consensus is that this knot is likely at least a partial image of the bulge.

[3] Image 1.13.6 is predicted by some models as counter image of 1.13, but its identification is not confident enough to be used as constraint.

4. BRIEF DESCRIPTION OF MODELING TECHNIQUES AND THEIR INPUTS

For convenience of the reader, in this section we give a brief description of each of the modeling techniques compared in this work (summarized briefly in Table 4). We note that the five models span a range of very different assumptions. Three of the teams (Grillo et al., Oguri et al., Sharon et al.) used an approach based on modeling the mass distribution with a set of physically motivated components, described each by a small number of parameters, representing the galaxies in the cluster and the overall cluster halo. We refer to these models as "simply-parametrized". One of the approaches (Diego et al.) describes the mass distribution with a larger number of components. The components are not associated with any specific physical object and are used as building blocks, allowing for significant flexibility, balanced by regularization. We refer to this model as "free-form"[30]. The fifth approach (Zitrin et al.), is based on the assumption that light approximately traces mass, and the mass components are built by smoothing and rescaling the observed surface brightness of the cluster members. We refer to this approach as "light-traces-mass". All the models considered here are single-plane lens models. As we will discuss in Section 6, each type of model uses a different approach to account for the effects of structure along the line of sight, and to break the mass sheet degeneracy. All model outputs will be made available through the HFF website after the acceptance of the individual modeling papers.

Table 5
Summary of models

Short name	Team	Type	RMS	Images
Die-a	Diego et al.	Free-form	0.78	gold+sil
Gri-g	Grillo et al.	Simply-param	0.26	gold
Ogu-g	Oguri et al.	Simply-param	0.43	gold
Ogu-a	Oguri et al.	Simply-param	0.31	all
Sha-g	Sharon et al.	Simply-param	0.16	gold
Sha-a	Sharon et al.	Simply-param	0.19	gold+sil
Zit-g	Zitrin et al.	Light-tr-mass	1.3	gold

Note. — For each model we provide a short name as well as basic features and inputs. The column RMS lists the r.m.s. scatter of the observed vs predicted image positions in arcseconds.

We note that members of our team have developed another complementary "free-form" approach, based on modeling the potential in pixels on an adaptive grid (Bradač et al. 2004b, 2009). However, given the pixellated nature of the reconstruction and the need to compute numerical derivatives and interpolate from noisy pixels in order to compute time delays and magnification at the location of 'Refsdal', we did not expect this method to be competitive for this specific application. Therefore in the interest of time we did not construct this model. A pre-HFF model of MACSJ1149.5+2223 using this approach is available through the HFF website and will be updated in the future.

When appropriate, we also describe additional sets of constraints used by each modeler.

4.1. *Diego et al.*

A full description of the modeling technique used by this team (J.D., T.B.) and the various improvements implemented in the code can be found in the literature (Diego et al. 2005, 2007; Sendra et al. 2014; Diego et al. 2015). Here is a brief summary of the basic steps.

4.1.1. *Definition of the mass model*

The algorithm (WSLAP+) relies on a division of the mass distribution in the lens plane into two components. The first is compact and associated with the member galaxies (mostly red ellipticals). The second is diffuse and distributed as a superposition of Gaussians on a regular (or adaptive) grid. In this specific case, a grid of 512×512 pixels $0\farcs1875$ on a side was used. For the compact component, the mass associated with the galaxies is assumed to be proportional to their luminosity. If all the galaxies are assumed to have the same mass-to-light (M/L) ratio, the compact component (galaxies) contributes with just one ($N_g = 1$) extra free-parameter wich corresponds to the correction that needs to be applied to the fiducial M/L ratio. In some particular cases, some galaxies (like the BCG or massive galaxies very close to an arclet) are allowed to have their own M/L ratio adding additional free-parameters to the lens model but typically no more than a few ($N_g \sim$ O(1)). For this component associated with the galaxies, the total mass is assumed to follow either a Navarro et al. (1997, hereafter NFW) profile (with a fixed concentration, and scale radius scaling with the fiducial halo mass) or be proportional to the observed surface brightness. For this work the team adopted $N_g = 2$ or $N_g = 3$. The case $N_g = 2$ considers one central brightest cluster galaxy (BCG) and the elliptical galaxy near image 1.2 to have the same M/L ratio, while the remaining galaxies have a different one. In the case $N_g = 3$, the BCG and the galaxy near image 1.2 have each their own M/L ratio, and the remaining galaxies are assumed to have a third independent value. In all cases, it is important to emphasize that the member galaxy between the 4 observed images of 'Refsdal' was not allowed to have its own independent M/L ratio. This results in a model that is not as accurate on the smallest scales around this galaxy as other models that allow this galaxy to vary.

The diffuse, or 'soft', component, is described by as many free parameters as grid (or cell) points. This number (N_c) varies but is typically between a few hundred to one thousand ($N_c \sim$ O(100)-O(1000)) depending on the resolution and/or use of the adaptive grid. In addition to the free-parameters describing the lens model, the problem includes as unknowns the original positions of the lensed galaxies in the source plane. For the clusters included in the HFF program the number of background sources, N_s, is typically a few tens ($N_s \sim$ O(10)), each contributing with two unknowns (β_x and β_y). All the unknowns are then combined into a single array X with N_x elements ($N_x \sim O(1000)$).

4.1.2. *Definition of the inputs*

[30] These models are sometimes described incorrectly as "non-parametric", even though they typically have more parameters than the so-called parametric models.

The inputs are the pixel position of the strongly lensed galaxies (not just the centroids) for all the multiple images listed in Tables 3 and 4. In the case of elongated arcs near the critical curves with no features, the entire arc is mapped and included as a constraint. If the arclets have individual features, these can be incorporated as semi-independent constraints but with the added condition that they need to form the same source in the source plane. The following inputs are added to the default set of image and knots centers listed in Section 3:

1. Shape of the arclets. This is particularly useful for long elongated arcs (with no counter images) which lie in the regime between the weak and strong lensing. These arcs are still useful constraints that add valuable information beyond the Einstein radius.

2. Shape and morphology of arcs. By including this information one can account (at least partially) for the magnification at a given position.

3. Resolved features in individual systems. This new addition to the code is motivated by the host galaxy of 'Refsdal' where multiple features can be identified in the different counter images. In addition, the counter image in the North, when re-lensed, offers a robust picture of the original source morphology (size, shape, orientation). This information acts as an anchor, constraining the range of possible solutions.

Weak lensing shear measurements can also be used as input to the inference. For the particular case of MACSJ1149.5+2223 the weak lensing measurements are not used, to ensure homogeneity with the other methods.

4.1.3. *Description of the inference process and error estimation*

The array of best fit parameters X, is obtained after solving the system of linear equations

$$\Theta = \Gamma X \quad (1)$$

where the N_o observations (strong lensing, weak lensing, time delays) are included in the array Θ and the matrix Γ is known and has dimension $N_o \times (N_c + N_g + 2N_s)$.

In practice, X is obtained by solving the set of linear equations described in Eq. 1 via a fast bi-conjugate algorithm, or inverted with a singular value decomposition (after setting a threshold for the eigenvalues) or solved with a more robust but slower quadratic algorithm. The quadratic algorithm is the preferred method as it imposes the physical constraint that the solution X must be positive. This eliminates unphysical solutions with negative masses and reduces the space of possible solutions. Like in the case of the bi-conjugate gradient, the quadratic programming algorithm solves the system of linear equations by finding the minimum of the associated quadratic function. Errors in the solution are derived by minimizing the quadratic function multiple times, after varying the initial conditions of the minimization process, and/or varying the grid configuration.

4.2. *Grillo et al.*

The software used by this team (C.G., S.H.S., A.H., P.R., W.K., I.B., A.M., G.B.C.) is Glee (Suyu & Halkola 2010; Suyu et al. 2012). The strong lensing analysis performed here follows very closely the one presented by Grillo et al. (2015) for another HFF target, i.e. MACSJ0416.1−2403. Cosmological applications of Glee can be found in Suyu et al. (2013, 2014) and further details on the strong lensing modeling of MACSJ1149.5+2223 are provided in a dedicated paper (Grillo et al. 2015, in preparation).

4.2.1. *Definition of the mass model*

Different mass models have been explored for this galaxy cluster, but only the best-fitting one is discussed here. The projected dimensionless total surface mass density of each of the 306 cluster members within the WFC3 field of view of the CLASH observations is modeled as a dual pseudoisothermal elliptical mass distribution (dPIE; Elíasdóttir et al. 2007) with vanishing ellipticity and core radius. The galaxy luminosity values in the F160W band are used to assign the relative weights to their total mass profile. The galaxy total mass-to-light ratio is scaled with luminosity as $M_{\rm T}/L \sim L^{0.2}$, thus mimicking the so-called tilt of the Fundamental Plane. The values of axis ratio, position angle, effective velocity dispersion and truncation radius of the two cluster members closest in projection to the central and southern images of the 'Refsdal' host are left free. To complete the total mass distribution of the galaxy cluster, three additional mass components are added to describe the cluster dark matter halo on physical scales larger than those typical of the individual cluster members. These cluster halo components are parametrized as two-dimensional, pseudo-isothermal elliptical mass profiles (PIEMD; Kassiola & Kovner 1993). No external shear or higher order perturbations are included in the model. The number of free parameters associated with the model of the cluster total mass distribution is 28.

4.2.2. *Definition of the inputs*

The positions of the multiple images belonging to the 10 systems of the "gold" sample and to 18 knots of the 'Refsdal' host are the observables over which the values of the model parameters are optimized. The adopted positional uncertainty of each image is $0\farcs065$. The redshift values of the 7 spectroscopically confirmed "gold" systems are fixed, while the remaining 3 systems are included with a uniform prior on the value of D_{ds}/D_s, where D_{ds} and D_s are the deflector-source and observer-source angular diameter distances, respectively. In total, 88 observed image positions are used to reconstruct the cluster total mass distribution.

4.2.3. *Description of the inference process and error estimation*

The best-fitting, minimum-χ^2 model is obtained by minimizing the distance between the observed and model-predicted positions of the multiple images in the lens plane. A minimum χ^2 value of 1441, corresponding to a RMS offset between the image observed and reconstructed positions of $0\farcs26$, is found. To sample the posterior probability distribution function of the model parameters, the image positional uncertainty is increased

until the value of the χ^2 is comparable to the number of the degrees of freedom (89) and standard Markov chain Monte Carlo (MCMC) methods are used. The quantities shown in Figures 9 to 12 are for the model-predicted images of 'Refsdal' and are obtained from 100 different models extracted from an MCMC chain with 10^6 samples and an acceptance rate of approximately 0.13.

4.3. *Oguri et al.*

4.3.1. *Definition of the mass model*

This team (M.O., M.I., R.K.) uses the public software GLAFIC (Oguri 2010). This "simply-parametrized" method assumes that the lens potential consists of a small number of components describing dark halos, cluster member galaxies, and perturbations in the lens potential. The dark halo components are assumed to follow the elliptical NFW mass density profile. In contrast, the elliptical pseudo-Jaffe profile is adopted to describe the mass distribution of cluster member galaxies. In order to reduce the number of free parameters, the velocity dispersion σ and the truncation radius $r_{\rm cut}$ for each galaxy are assumed to scale with the ($F814W$-band) luminosity of the galaxy as $\sigma \propto L^{1/4}$ and $r_{\rm cut} \propto L^{\eta}$ with η being a free parameter. In addition, the second (external shear) and third order perturbations are included in order to account for asymmetry of the overall lens potential. Interested readers are referred to Oguri (2010, 2015), Oguri et al. (2012, 2013), and Ishigaki et al. (2015) for more detailed descriptions and examples of cluster mass modeling with GLAFIC. Additional details are given in a dedicated paper (Kawamata et al. 2015, in preparation).

4.3.2. *Definition of the inputs*

The positions of multiple images and knots listed in Section 3 are used as constraints. Image 1.5 was not used as a constraint. To accurately recover the position of SN 'Refsdal', different positional uncertainties are assumed for different multiple images. Specifically, while the positional uncertainty of $0\farcs4$ in the image plane is assumed for most of multiple images, smaller positional uncertainties of $0\farcs05$ and $0\farcs2$ are assumed for SN 'Refsdal' and knots of the SN host galaxy, respectively (see also Oguri 2015). When spectroscopic redshifts are available, their redshifts are fixed to the spectroscopic redshifts. Otherwise source redshifts are treated as model parameters and are optimized simultaneously with the other model parameters. For a subsample of multiple image systems for which photometric redshift estimates are secure and accurate, a conservative Gaussian prior with the dispersion of $\sigma_z = 0.5$ for the source redshift is added. While GLAFIC allows one to include other types of observational constraints, such as flux ratios, time delays, and weak lensing shear measurements, those constraints are not used in the mass modeling of MACSJ1149.5+2223.

4.3.3. *Description of the inference process and error estimation*

The best-fit model is obtained simply by minimizing χ^2. The so-called source plane χ^2 minimization is used for an efficient model optimization (see Appendix 2 of Oguri 2010). A standard MCMC approach is used to estimate errors on model parameters and their covariance.

The predicted time delays and magnifications are computed at the model-predicted positions. For each mass model (chain) the best-fit source position of the SN is derived. From that, the corresponding SN image positions in the image plane (which can be slightly different from observed SN positions) are obtained for that model, and finally the time delays and magnifications of the images are calculated.

4.4. *Sharon et al.*

The approach of this team (K.S., T.J.) was based on the publicly available software Lenstool (Jullo et al. 2007). Lenstool is a "simply-parametrized" lens modeling code. In practice, the code assumes that the mass distribution of the lens can be described by a combination of mass halos, each of them taking a functional form whose properties are defined by a set of parameters. The method assumes that mass generally follows light, and assigns halos to individual galaxies that are identified as cluster members. Cluster- or group-scale halos represent the cluster mass components that are not directly related to galaxies. The number of cluster or group-scale halos is determined by the modeler. Typically, the positions of the cluster scale halos are not fixed and are left to be determined by the modeling algorithms. A hybrid "simply-parametrized"/"free-form" approach has also been implemented in Lenstool (Jullo & Kneib 2009), where numerous halos are placed on a grid, representing the overall cluster component. This hybrid method is not implemented in this work.

4.4.1. *Definition of the mass model*

All the halos are represented by a PIEMD mass distribution with density profile $\rho(r)$ defined as:

$$\rho(r) = \frac{\rho_0}{(1 + r^2/r_{core}^2)(1 + r^2/r_{cut}^2)}. \quad (2)$$

These halos are isothermal at intermediate radii, i.e., $\rho \propto r^{-2}$ at $r_{core} \lesssim r \lesssim r_{cut}$, and have a flat core internal to r_{core}. The transition between the different slopes is smooth. σ_0 defines the overall normalization as a fiducial velocity dispersion. In Lenstool, each PIEMD halo has seven free parameters: centroid position x,y; ellipticity $e = (a^2 - b^2)/(a^2 + b^2)$ where a and b are the semi major and minor axis, respectively; position angle θ; and r_{core}, r_{cut}, σ_0 as defined above.

The selection of cluster member galaxies is described in Section 3.3. In this model, 286 galaxies were selected from the cluster member catalog, by a combination of their luminosity and projected distance from the cluster center, such that the deflection caused by an omitted galaxy is much smaller than the typical uncertainty due to unseen structure along the line of sight. This selection criterion results in removal of faint galaxies at the outskirts of the cluster, and inclusion of all the galaxies that pass the cluster-member selection in the core.

Cluster member galaxies are modeled as PIEMD as well. Their positional parameters are fixed on their observed properties as measured with SExtractor (Bertin & Arnouts 1996) for x, y, e, and θ. The other parameters, r_{core}, r_{cut}, and σ_0, are linked to their luminosity in the $F814W$ band through scaling relations (e.g., Limousin et al. 2005) assuming a constant mass-to-light ratio for

all galaxies,

$$\sigma_0 = \sigma_0^* \Big(\frac{L}{L^*}\Big)^{1/4} \quad \text{and} \quad r_{cut} = r_{cut}^* \Big(\frac{L}{L^*}\Big)^{1/2}. \quad (3)$$

4.4.2. *Definition of the inputs*

The lensing constraints are the positions of multiple images of each lensed source, plus those of the knots in the host galaxy of 'Refsdal', as listed in Section 3. In cases where the lensed image is extended or has substructure, the exact positions were selected to match similar features within multiple images of the same galaxy with each other thus obtaining more constraints, a better local sampling of the lensing potential, and better handle on the magnification, locally. Where available, spectroscopic redshifts are used as fixed constraints. For sources with no spectroscopic redshift, the redshifts are considered as free parameters with photometric redshifts informing their Bayesian priors. The uncertainties of the photometric redshifts are relaxed in order to allow for outliers. We present two models here: Sha-g uses as constraints the 'gold' sample of multiply-imaged galaxies, and Sha-a uses 'gold', 'silver', and secure arcs outside the MUSE field of view, to allow a better coverage of lensing evidence in the outskirts of the cluster and in particular to constrain the sub halos around MACSJ1149.5+2223.

4.4.3. *Description of the inference process and error estimation*

The parameters of each halo are allowed to vary under Bayesian priors, and the parameter space is explored in a n MCMC process to identify the set of parameters that provide the best fit. The quality of the lens model is measured either in the source plane or in the image plane. The latter requires significantly longer computation time. In source plane minimization, the source positions of all the images of each set are computed, by ray-tracing the image plane positions through the lens model to the source plane. The best-fit model is the one that results in the smallest scatter in the source positions of multiple images of the same source. In image-plane minimization, the model-predicted counter images of each of the multiple images of the same source is computed. This results in a set of predicted images near the observed positions. The best-fit model is the one that minimizes the scatter among these image-plane positions. The MCMC sampling of the parameter space is used to estimate the statistical uncertainties that are inherent to the modeling algorithm. In order to estimate the uncertainties on the magnification and time delay magnification, potential maps are generated from sets of parameters from the MCMC chain that represent 1-σ in the parameter space.

4.5. *Zitrin et al.*

4.5.1. *Definition of the mass model*

The method used by this team (A.Z.) is a Light Traces Mass (LTM) method, so that both the galaxies *and* the dark matter follow the light distribution. The method is described in detail by Zitrin et al. (2009, 2013) and is inspired by the LTM assumptions outlined by Broadhurst et al. (2005). The model consists of two main components. The first component is a mass map of the cluster galaxies, chosen by following the red sequence. Each galaxy is represented with a power-law surface mass density distribution, where the surface density is proportional to its surface brightness. The power-law is a free parameter of the model and is iterated for (all galaxies are forced to have the same exponent). The second component is a smooth dark matter map, obtained by smoothing (with a Spline polynomial or with a Gaussian kernel) the first component, i.e. the superposed red sequence galaxy mass distribution. The smoothing degree is the second free parameter of the model. The two components are then added with a relative weight which whis is a free parameter, along with the overall normalization. A two-component external shear can be then added to add flexibility and generate ellipticity in the magnification map. Lastly, individual galaxies can be assigned with free masses to be optimized by the minimization procedure, to allow more degrees of freedom deviating form the initial imposed LTM. This procedure has been shown to be very effective in locating multiple images in many clusters (e.g., Zitrin et al. 2009, 2012b, 2013, 2015) even without any multiple images initially used as input (Zitrin et al. 2012a). Most of the multiple images in MACSJ1149.5+2223 that were found by Zitrin & Broadhurst (2009) and Zheng et al. (2012), were identified with this method.

4.5.2. *Definition of the inputs*

All sets of multiple images in the gold list were used except system 14. Most knots were used except those in the fifth radial BCG image. All systems listed with spec-z (aside for system 5) were kept fixed at that redshift while all other gold systems were left to be freely optimized with a uniform flat prior. Image position uncertainties were adopted to be $0\farcs5$, aside for the four SN images for which $0\farcs15$ was used.

4.5.3. *Description of the inference process and error estimation*

The best fit solution and errors are obtained via converged MCMC chains.

5. COMPARISON OF LENS MODELS

In this section we carry out a comparison of the 7 models, focusing specifically on the quantities that are relevant for 'Refsdal'. We start in Section 5.1 by presenting the two-dimensional maps of convergence, magnification and time delay, for a deflector at the redshift of the cluster and a source at the redshift of 'Refsdal' ($z = 1.488$; we note that assuming $z = 1.491$, the redshift published by Smith et al. (2009), would not have made any significant difference). Then, in Section 5.2, we compare quantitatively the predicted time delays and magnification ratios of the known images with their mesaured values. Finally, in Section 5.3 we present the forecast for the future (and past) SN image. All the lens models predict the appearance of an image of the SN in the two other images of the host galaxy. In the following sections, we refer to the predicted SN in image 1.2 of the host galaxy as SX, and the one in image 1.3 of the host as SY, following the labeling of previous publications.

5.1. *Convergence, magnification and time delay maps*

Figure 4 shows the convergence (i.e., surface mass density in units of the lensing critical density) maps. There

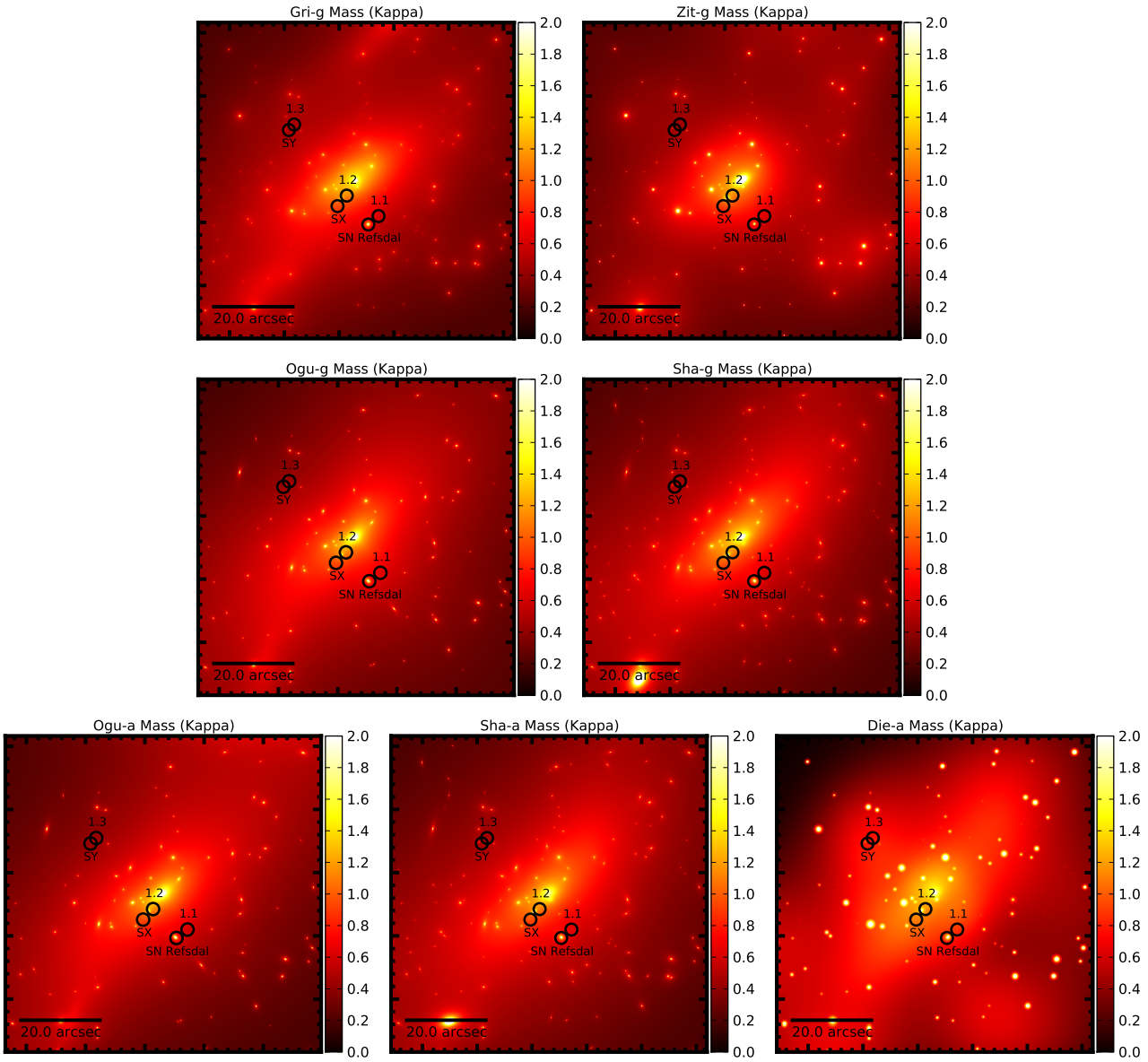

Figure 4. Comparing the mass distributions for the models, labeled as in Table 4. Convergence is computed relative to the critical density with the deflector at the redshift of the cluster and the source at the redshift of the supernova. The cirles identify the positions of the observed and predicted images of Refsdal and those of the multiple images of its host galaxy. The top four panels are models including only the gold sample of images as constraints.

are striking qualitative differences. The Zit-g map is significantly rounder than the others. The Die-a map has significantly more structure, notably two overdensities near SY/1.3 and at the bottom right of the map. These features were to be expected based on the assumptions used by their methods. The Grillo, Oguri, Sharon convergence maps are the most qualitatively similar. This is perhaps unsurprising since the three codes are based on fairly similar assumptions.

Magnification maps are shown in Fig. 5. The regions of extreme magnification are qualitatively similar, even though, similarly to the convergence maps, the Zit-g model is overall rounder, while the Die-a model has more structure.

The time delay surfaces are shown at three zoom levels to highlight different features. Fig. 6 shows the global topology of the time delay surfaces, which is very similar for all models, with minima near 1.1 and SY/1.3 and a saddle point near SX/1.2. As was the case with convergence and magnification, the Zit-g and Die-a time delays surfaces are rounder and have more structure, respectively, than those produced by the other models.

Zooming in the region of SX/1.2 and 1.1 in Fig. 7 reveals more differences. The locations of the minimum near SN 'Refsdal' and of the saddle point near 1.2 are significantly different for the Zit-g model, seemingly as a result of the different contribution of the bright galaxy to the NW of 1.2.

A further zoom in the region of the known images is shown in Figure 8. The time delay surface contour levels are shown in step of 10 days to highlight the behaviour relevant for the cross configuration. Whereas the "simply

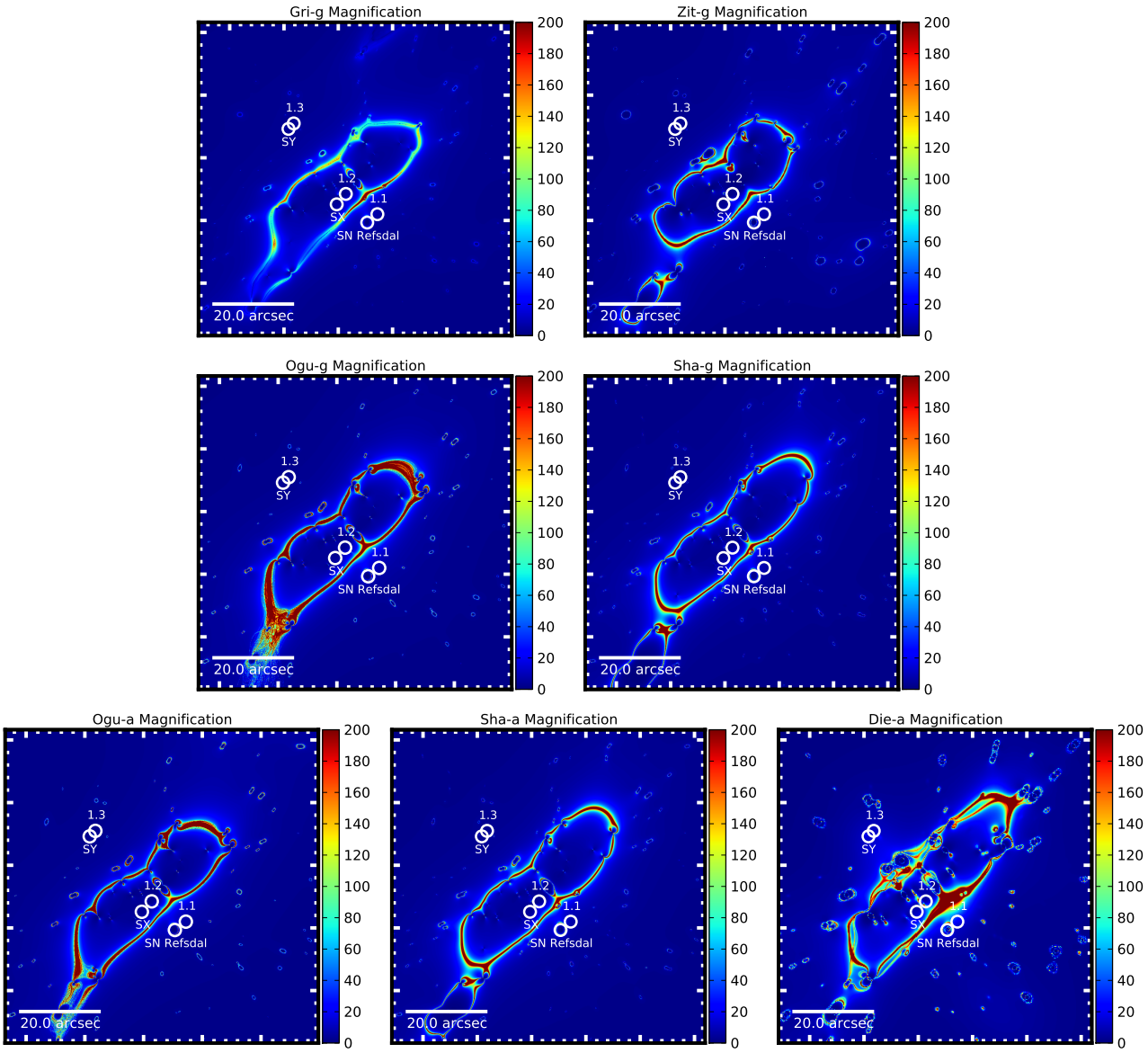

Figure 5. As Figure 4 for magnification.

parametrized" models are topologically very similar to each other, the Die-a and Zit-g models are qualitatively different. The time delay surface is shifted upwards - probably as a result of the nearby perturber highlighted in the previous paragraph. We stress that all the models here are global models, developed to reproduce the cluster potential on larger scale. Hence, local differences should be expected, even though of course they are particularly important in this case.

5.2. *Comparing model predictions with measured time delays and magnification ratios*

Before proceeding with a quantitative comparison, we emphasize once again that the uncertainties discussed in this Section only include statistical uncertainties. Furthermore, in the comparison we neglect for computational reasons the covariance between the uncertainties in time delay and between those in magnification. Systematic uncertainties will be discussed in Section 6.

Figure 9 compares the measured time delays with those predicted by the models for the cross configuration. We stress that the measurements were not used in the construction of the models (or known to the modelers) and therefore they can be considered an independent test of the models. The time delay between S2 and S1 (and to some extent that between S3 and S1) is very short and in fact not all the models agree on the ordering of the two images. The time delay between S4 and S1 is longer and better behaved, with all the models agreeing on the order of the images and with the measured value within the uncertainties. Overall the models are in reasonable agreement with the measurements, even though formally some of them are in statistical tension. This indicates that the uncertainties for some of the parametric models are underestimated.

Interestingly, the models appear to predict rather accurately the observed magnification ratios (Fig. 10), even though these quantitites should be more sensitive to sys-

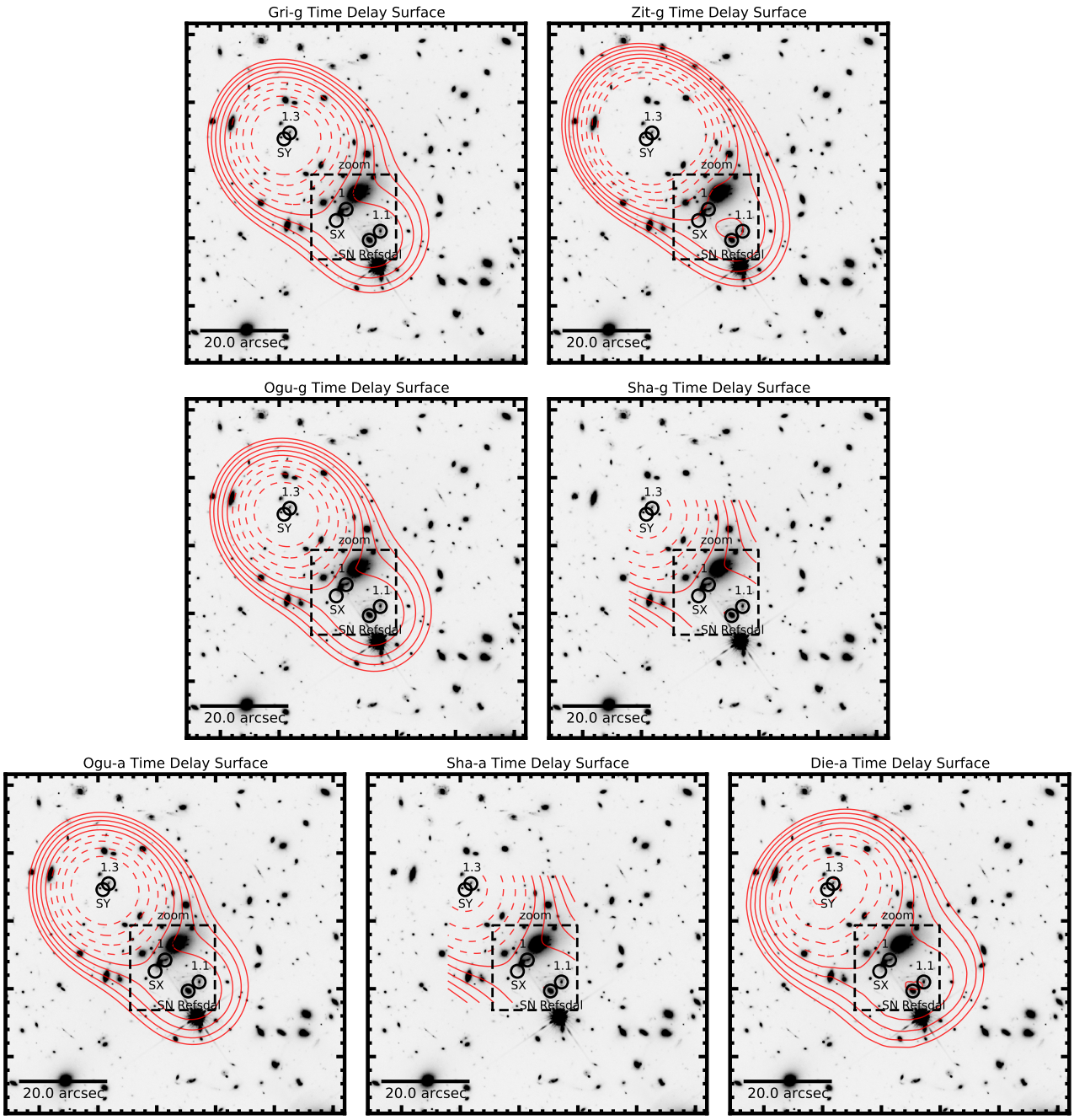

Figure 6. As Figure 4 for time delay surfaces. The dashed boxes mark the location of the zoom-ins shown in Figure 7. Contour levels show the time delay from -12 to 12 years in increments of 3 years, relative to S1. For the Sha-a and Sha-g models the time delay surfaces were only calculated in the region shown. Negative levels are marked by dashed contours. The gray-scale background image shows the HFF F140W epoch2 version 1.0 mosaic.

tematic uncertainties arising from milli-lensing and microlensing effects than time delays.

Overall, the Zit-g model stands apart from the rest, predicting significantly different time delays and magnification ratios, and larger uncertainties. This qualitative difference is consistent with the different topology of the time-delay surface highlighted in the previous section. Quantitatively, however, the Zit-g model predictions are in broad agreement with the measurements if one considers the 95% credible interval. Collectively, the "simply-parametrized" models seem to predict smaller uncertainties than the others, especially the Ogu-g and Ogu-a ones. This is expected, considering that they have less flexibility than the free form model. What is surprising, however, is that they also obtain the smaller r.m.s. residual scatter in the predicted vs observed image positions (Table 4). The Zit-g light traces mass model is perhaps the least flexible, in the sense that it cannot account for systematic variations in the projected mass-to-light ratio. This appears to be reflected in its overall largest r.m.s. residual scatter. When comparing the Die-a to the Zit-g model, we note that the former uses

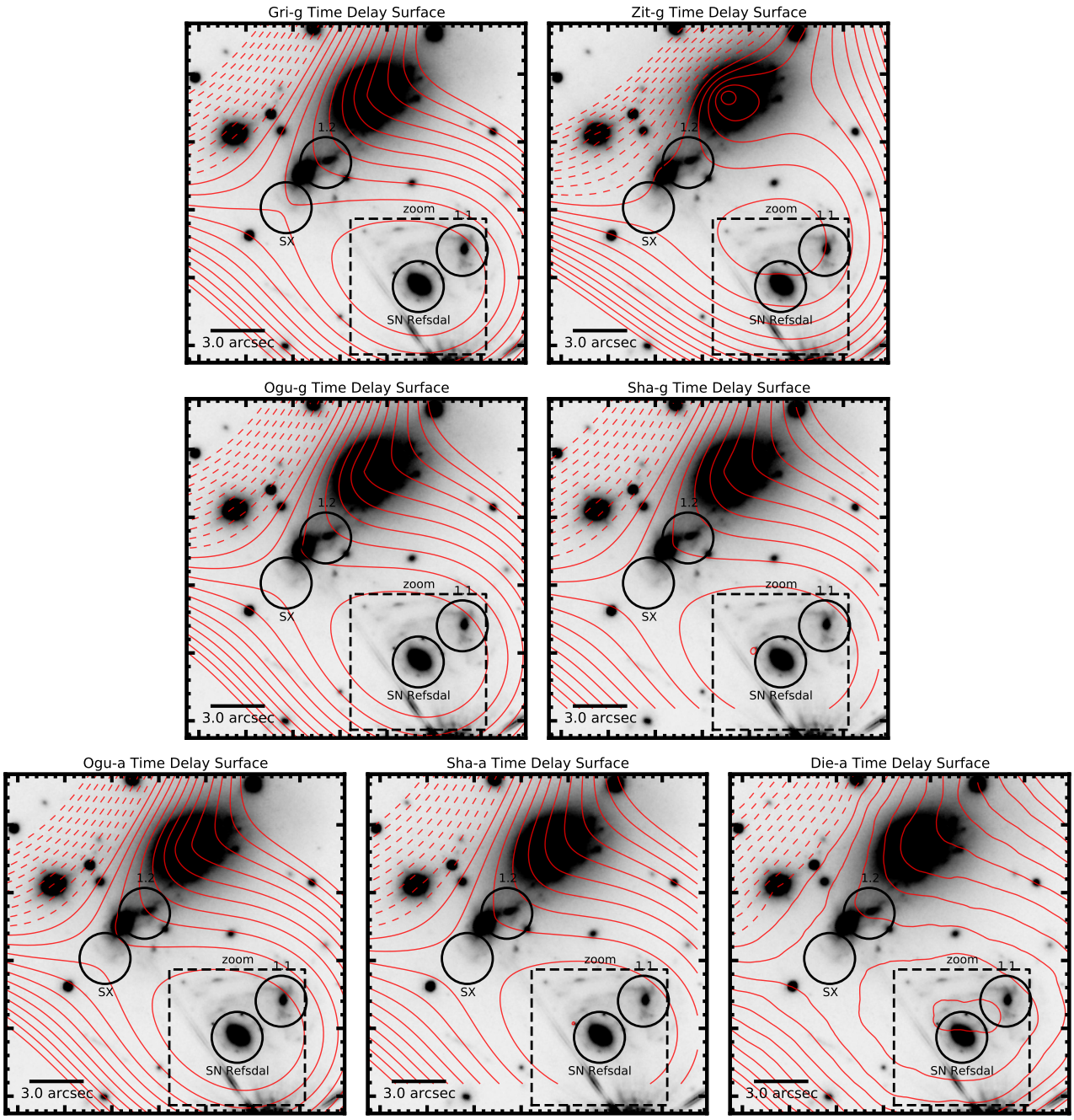

Figure 7. The time delay surface details in the region marked in Figure 6. The dashed boxes mark the location of the zoom-ins shown in Figure 8. Contour levels show the time delay from -5 to 5 years in increments of 0.5 years, relative to S1. Negative levels are marked by dashed contours. The gray-scale background image shows the HFF F140W epoch2 version 1.0 mosaic.

significantly more constraints than the latter. This may explain why, even though Die-a is in principle more flexible, it ends up estimating generally smaller uncertainties than Zit-g.

5.3. *Forecasts for 'Refsdal': peak appearance and brightness*

Figure 11 compares the prediction for the next appearance of SN 'Refsdal', near image 1.2 of the spiral galaxy (hereafter SX/1.2). All the models considered here predict the image to peak between the end of 2015 and the first half of 2016. We note that S1 was first discovered six months before its peak with F160W AB magnitude ∼25.5 (Kelly et al. 2015), and peaked at about F160W∼24.5 AB (Kelly et al. 2015, in preparation; Strolger et al. 2015, in preparation). Image SX/1.2 is predicted to be approximately 1/3 as bright as image S1 (Figure 12), so it should be approximately ∼26.7 six months before peak and 25.7 at peak. No image is detected in the vicinity of SY/1.3 in data taken with *HST* up until MACSJ1149.5+2223 became unobservable at the end of July, allowing us to rule out predicted peak times until January 2016.

Remarkably, the models are in excellent mutual agreement regarding the next appearance of 'Refsdal'. All the

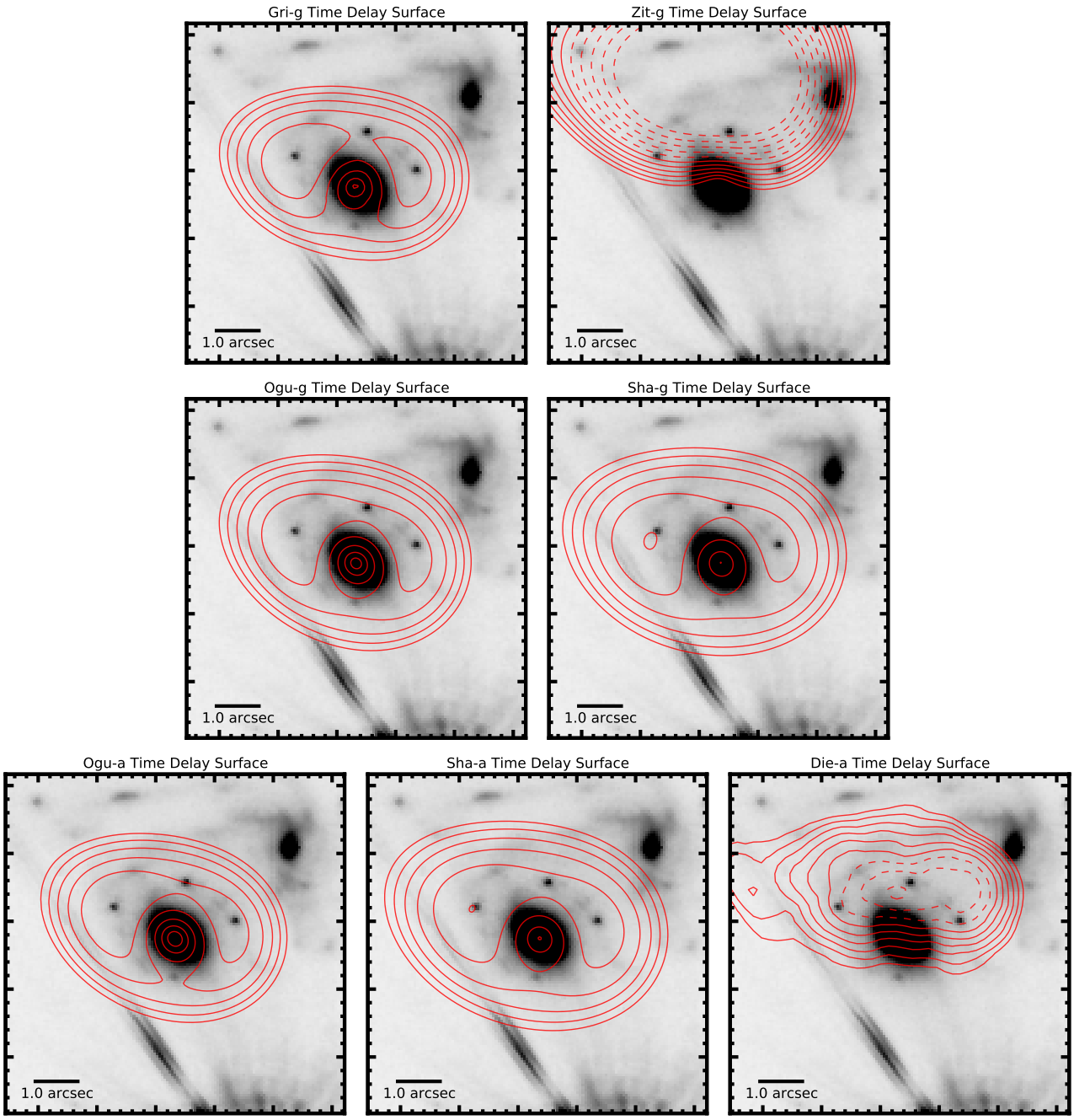

Figure 8. The time delay surface details in the region marked in Figure 7. Contour levels show the time delay from -50 to 50 days in increments of 10 days, relative to S1. Negative levels are marked by dashed contours. The gray-scale background image shows the HFF F140W epoch2 version 1.0 mosaic.

predictions agree on the first trimester of 2016 as the most likely date of the peak. Sha-a is the only one that predicts a slightly fainter flux with a magnification ratio ($0.19^{+0.01}_{-0.04}$) as opposed to the $\sim$1/3 value predicted by the other models. Interestingly, Zit-g has the largest uncertainty on time delay, but not on magnification ratio. As in the case of the cross configuration, the "simply-parametrized" models yield the smallest uncertainties.

Unfortunately, the model-based estimates of the past appearance of 'Refsdal' cannot be tested by observations. The image near 1.3 (herafter SY/1.3) is estimated to have been significantly fainter than S1, and thus undetectable from the ground, at a time when WFC3-IR was not available. The images of MACSJ1149.5+2223 taken in the optical with ACS in April 2004 (GO: 9722, PI: Ebeling; 3-σ limit F814W AB = 27.0) are not sufficiently deep to set any significant constraints, considering the peak brightness of S1 in F814W was approximately $\sim$27, and we expect SY/1.3 to be 0.75 to 2 magnitudes fainter. As a purely theoretical exercise it is interesting to notice that the time delay varies dramatically between models, differing by almost 10 years between the Zit-g and the Sha-a, Sha-g and Die-a models. Remarkably, and similarly to what was seen for the cross configuration, the

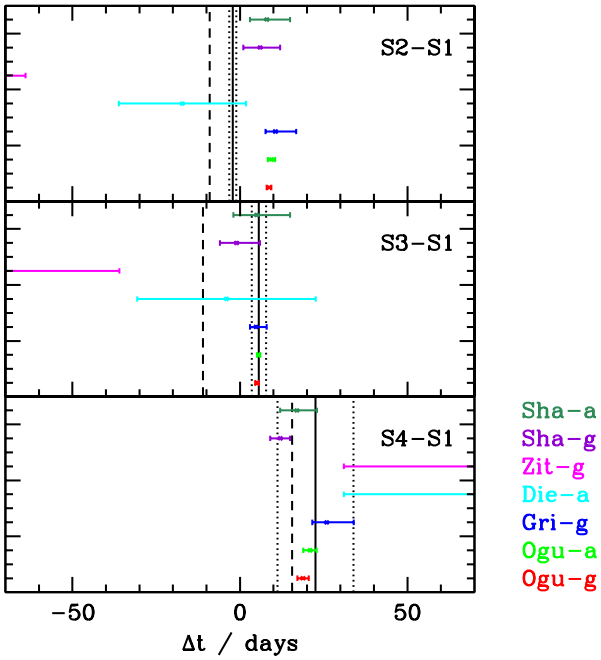

Figure 9. Observed (solid vertical line with dotted lines represents the measurement and uncertainty by Kelly et al.; dashed vertical line represents the measurement by Strolger et al.) and predicted (points with error bars) time delays for the images in the cross configuration, relative to S1. Uncertainties represent the 68% confidence interval.

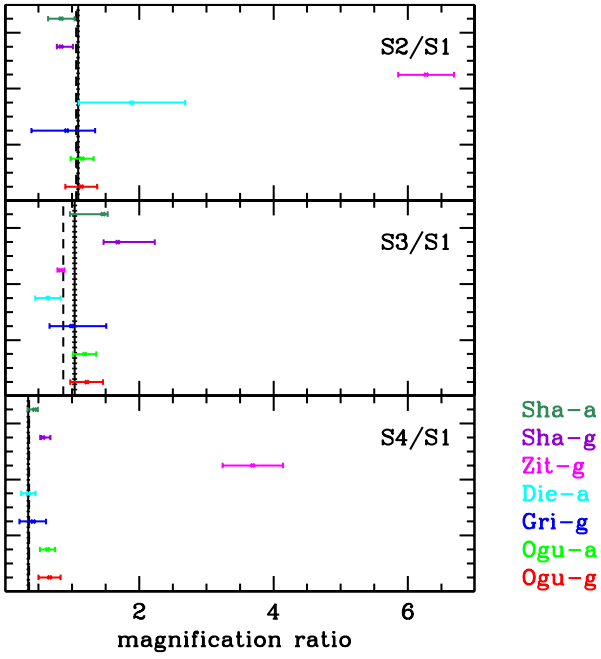

Figure 10. Observed (lines as in Figure 9) and predicted (points with error bars) magnification ratios (absolute values) for the images in the cross configuration, relative to S1.

magnifications are in significantly better agreement.

6. DISCUSSION

In this section we briefly discuss our results, first by recapitulating the limitations of our analysis (Section 6.1), and then by comparing them with previous work (Section 6.2).

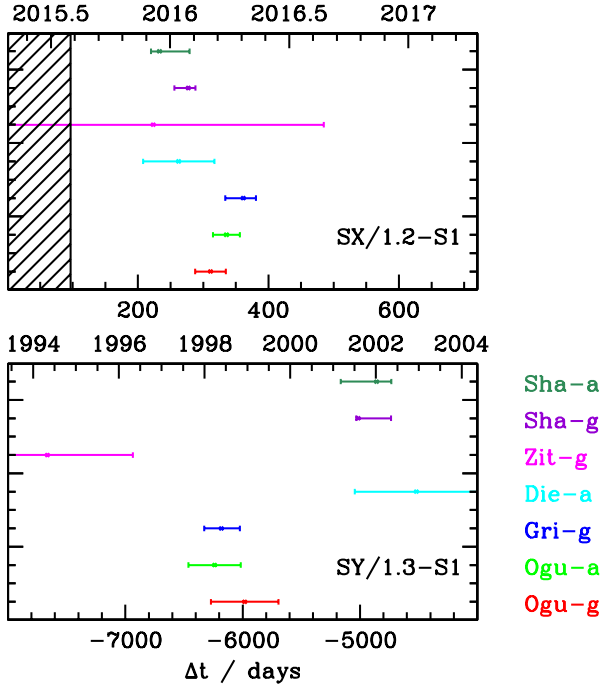

Figure 11. Predicted time delays for the more distant images, relative to S1. The top scale gives the expected date of the peak brightness of the image - with an uncertainty of ±20 days given by the uncertainty on the date of the peak of the observed images. The hatched region is ruled out by past observations with the Hubble Space Telescope.

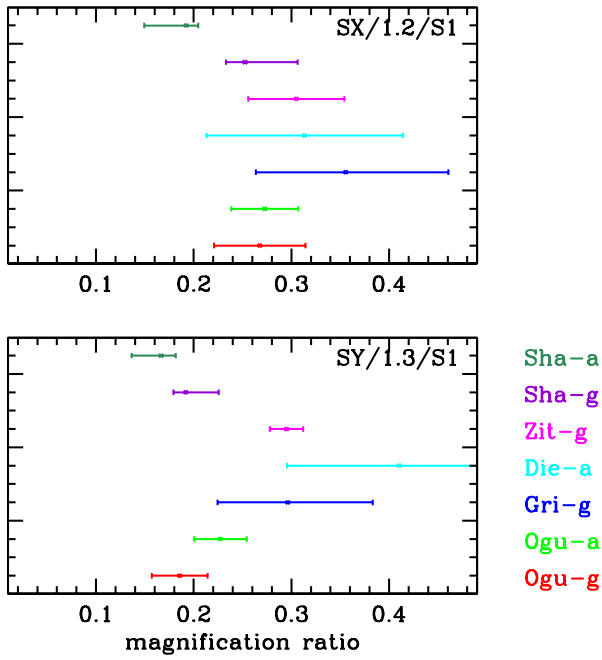

Figure 12. Predicted magnification ratios (absolute values) for the more distant images, relative to S1.

6.1. *Limitation of the blind test and of the models*

SN 'Refsdal' gives us a unique opportunity to test our models blindly. However, in order to draw the appropriate conclusions from this test, we need to be aware of the limitations of both the test and the models.

The first limitation to keep in mind, is that this test

is very specific. We are effectively testing point-like predictions of the lensing potential and its derivatives. Similarly to the case of SN 'Tomas' (Rodney et al. 2015), it is very hard to generalize the results of this test even to the strong lensing area shown in our maps. More global metrics should be used to infer a more global assessment of the quality of the models. An example of such metric is the r.m.s. scatter between the image positions given in Table 4, even though of course even this metric does not capture all the features of a model. For example the r.m.s. does not capture how well the model reproduces time delays and magnifications, in addition to positions, and one could imaging trading one for the other.

It is also important to remind ourselves that whereas the magnification and time delays at specific points may vary significantly between models, other quantities that are more relevant for statistical use of clusters as cosmic telescopes, such as the area in the source plane, are much more stable (e.g., Wang et al. 2015). And, of course, other quantities such as colors and line ratios, are not affected at all by gravitational lensing. It would be interesting to find ways to carry out true observational tests of more global predictions of lens models. One way to achieve this would be to carry out tests similar to those afforded by 'Tomas' and 'Refsdal' on a large sample of clusters. Another possibility could be to reach sufficiently deep that the statistical properties of the background sources (e.g. the luminosity function) are measured with sufficient precision and sufficiently small cosmic variance to allow for meaningful tests of model uncertainties. Alternatively, tests against simulated data are certainly informative (e.g. Meneghetti et al. 2015, in preparation), although their results should also be interpreted with great care, as they depend crucially on the fidelity of the simulated data and the cross-talk between methods used to simulate the data and those used to carry out the inference.

The second limitation to keep in mind is that the uncertainties listed in this paper are purely statistical in nature. As for the case of image positions – where the r.m.s. scatter is typically larger than the astrometric precision of the image positions themselves (consistent with the fact that there are residual systematics in cluster lens modeling due to known effects such as substructure, e.g., Bradač et al. 2009) – we should not expect the time delays and magnifications to be perfectly reproduced by the models either. The spread between the different model predictions gives us an idea of the so-called model uncertainties, even though unfortunately they cannot be considered an exact measurement. The spread could be exaggerated by inappropriate assumptions in some of the models, or underestimated if common assumptions are unjustified.

As already mentioned in the introduction, other potential sources of uncertainty are related to the mass sheet degeneracy and its generalizations (Falco et al. 1985; Schneider & Sluse 2013, 2014), the effects of structure along the line of sight (Dalal et al. 2005), and multiplane lensing (Schneider 2014; McCully et al. 2014). All the models considered here are single plane lens models. They break the mass sheet degeneracy by assuming that the surface mass density profile goes to zero at infinity with a specific radial dependency.

On the scale of the known images of 'Refsdal', the measured time delays and magnification ratios give us a way to estimate these residual uncertainties. The reasonably good agreement between the model prediction and measurements shows that these (systematic) 'unknown unknowns' are not dominant with the respect to the (statistical) 'known unknowns'. However, since the agreement is not perfect, we conclude that the 'unknown unknowns' are not negligible either. We can perhaps use the experience gathered in the study of time delays of lensed quasars to estimate the amplitude of the line of sight uncertainties. On scales similar to that of the known images of 'Refsdal' they are believed to be up to $\sim 10\%$ before corrections for galaxies not in clusters (Suyu et al. 2010; Greene et al. 2013; Collett et al. 2013; Suyu et al. 2014). In numerical simulations the line of sight effects appear to increase with the measured overdensity of galaxies (Greene et al. 2013), so it is possible that they are larger for an overdense region like that of MACSJ1149.5+2223. On galaxy scales, breaking the mass-sheet degeneracy using stellar kinematics and physically motivated galaxy models appears to produce results consistent with residual uncertainties of the order of a few per cent (Suyu et al. 2014). On cluster scales, the degeneracy is partly broken by the use of multiple images at different redshifts (e.g., Bradač et al. 2004a). However, in the absence of non-lensing data, we cannot rule out that the residual mass-sheet degeneracy is the dominant source of systematic uncertainty. Assessing the uncertainties related to multiplane effects would require knowledge of the mass distribution in three-dimensions and is beyond the scope of the present work. Thus multiplane lensing cannot be ruled out as a significant source of systematic uncertainty for the prediction of the time delay and magnification ratios of the known images of 'Refsdal'. As far as the future image of 'Refsdal' is concerned, future observations will tell us how much our uncertainties are underestimated due to unknown systematics.

Finally, we remind the reader that although for this analysis we kept fixed the cosmological parameters, they are a (subdominant) source of uncertainty. To first order, the time delay distance is proportional to the Hubble Constant, so there is at least a 3% systematic uncertainty (Riess et al. 2011; Freedman et al. 2012) on our predicted time delays (and substantially more when considering all the other parameters, depending on assumptions and priors; Suyu et al. 2013, 2014).

6.2. *Comparison with previous models*

We can get a quantitative sense of the improvement of the mass models as a result of the new data by comparing how the prediction of the time delay and magnification ratios have changed for the teams who had previously publish predictions.

6.2.1. *Previous models by members of our team*

The Zit-g model updates the models developed by A.Z. for the 'Refsdal' discovery paper (Kelly et al. 2015). The Zit-g model supercedes the estimates of time delays and magnifications given in the original paper by providing predictions as well as quantitative uncertainties.

The update of the Oguri (2015) model presented here changes the time delays for S2, S3, S4, SX, SY from 9.2, 5.2, 22.5, 357, -6193 days to 8.7 ± 0.7, 5.1 ± 0.5,

18.8 ± 1.7, 311 ± 23.6, -5982 ± 287 days, respectively (for the Ogu-g model, see plot for Ogu-a). Thus, the predicted time delays have changed by less than 1 or 2 σ, with the inclusion of additional data. The magnification ratios have been similarly stable. The main effect of the additional data has been to reduce the uncertainties.

The update of the Sharon & Johnson (2015) model presented here changes the time delays for S2, S3, S4, SX, SY from 2.0^{+10}_{-6}, -5.0^{+13}_{-7}, 7.0^{+16}_{-3}, 237^{+37}_{-50} -4251^{+369}_{-373} days to 6 ± 6, -1^{+7}_{-3}, 12 ± 3, 277^{+11}_{-21}, -5016^{+281}_{-15} days, respectively (for the Sha-g model, see plot for Sha-a). Thus, the predicted time delays have changed by less than 1 or 2 σ, with the inclusion of additional data, especially the new spectroscopic redshifts (the list of multiple images is very similar). The magnification ratios have been similarly stable. The main effect of the additional data has been to reduce the uncertainties.

Diego et al. (2015) do not give time delays for the cross configuration, owing to the limitations inherent to keeping the M/L of the galaxy in the middle of the cross fixed to the global value. Their predictions for the long delays SX and SY have changed with the inclusion of new data from 375 ± 25 to 262 ± 54 days, and from -3325 ± 762 to -4521 ± 524. Interestingly the uncertainties on the future delay have increased with the new data, which may be due to the correction of previously erroneous inputs, like the redshift of system 3, and also to the increased range of models and grid parameters considered here. The fact that the predictions changed by more than the estimated uncertainties is consistent with our previous conclusion that the statistical uncertainties underestimate the total uncertainty.

6.2.2. *Jauzac et al.*

During the final stages of the preparation of this manuscript, Jauzac et al. (2015) posted on the arxiv another independent model of MACSJ1149.5+2223. Their model is based on a subset of the data presented here, and different sets of multiple images and knots in the spiral host galaxy. Comparing only the systems with spectroscopic redshifts, our analyses agree on systems 1,2,3,4,5. We do not use systems 9 and 22, for which they obtain spectroscopic redshift of 0.981 and 3.216 respectively. We obtain spectroscopic redshifts for systems 13 and 14 (1.24 and 3.70), which contradict their model redshifts of 1.34 and 2.88. We also have measured a spectroscopic redshift for system 110, which they do not include in their analysis. Their catalog comprises 57 spectroscopically confirmed cluster members, while ours consists of 170. Thus, the Jauzac et al. (2015) model is not directly comparable to the models presented here. However, it provides a useful additional comparison for this forecast. We note that Jauzac et al. (2015) include the main developers of "Lenstool", the "simply-parametrized" lens modeling software used by Sharon et al. for the analysis presented in this paper. The difference in the predictions between the two teams highlights how systematic differences can arise from input data and modelers choices, as well as from assumptions of each modeling method.

The Jauzac et al. (2015) model[31] predicts time delays and magnification ratios that are significantly different from the ones actually observed from the cross configuration. Their predicted time delays for S2-S1 S3-S1 and S4-S1 are 90 ± 17, 30 ± 35, and -60 ± 41 days respectively, to be compared with the measured values given in Table 1. Considering their model uncertainties, which are much larger than the measurement uncertainties, the time delays are within 5.4σ, 1σ, and 2σ the Kelly et al. measurements. The disagreement with the S2-S1 time delay is especially remarkable considering that all the other models predict the two images to be almost simultaneous. The flux ratios (0.86 ± 0.13, 0.89 ± 0.11, and 0.42 ± 0.05 for S2/S1, S3/S1, S4/S1, respectively) are also somewhat in tension with the measured values, although the disagreement is in line with that of the models presented in this paper. It would be interesting to update the Jauzac et al. (2015) model, correcting the redshifts of systems 13 and 14 to see if those misidentifications could be at the root of the discrepancy. Overall it is interesting to note that Jauzac et al. (2015) predict the magnifications with higher accuracy than the time-delays, even though magnifications are potentially more sensitive to local substructure (milli-lensing) and microlensing effects.

The time delay predicted by Jauzac et al. (2015) for image SX/1.2 is significantly longer than for the models presented here, pushing the next appearance of the peak to the middle of 2016. The time delay of image SY/1.3 is shorter than for most models presented here, but unfortunately not short enough to be testable with archival observations. Incidentally, Jauzac et al. (2015) also predict SY/1.3 to be fainter than in the models presented here (0.16 ± 0.02 of the brightness of S1), which makes this prediction even more difficult to test with archival data.

7. SUMMARY

SN 'Refsdal' gives us a unique opportunity to carry out a truly blind test of cluster-scale gravitational lens models. In order to make the most of this opportunity we have used an unprecedented combination of imaging and spectroscopic data as input for 7 lens models, based on 5 independent techniques. The models have been tested against independent measurements of time delays and magnification ratios for the known images of 'Refsdal' and used to predict its future (and past) appearance. Our main results can be summarized as follows:

1. We have collected 429 spectroscopic redshifts in the field of MACSJ1149.5+2223 from VLT-MUSE (Grillo et al. 2015, in prep.) and HST-WFC3 (Brammer et al. 2015, in prep. Schmidt et al. 2014; Treu et al. 2015) observations. These include 170 spectroscopic cluster members and 23 multiple images of 10 different galaxies.

2. We have collected two independent measurements of time delays and magnification ratios for the known images of 'Refsdal' (Kelly et al. 2015, in preparation; Strolger et al. 2015, in preparation).

3. We have compiled and expanded a list of candidate multiply imaged galaxies and multiply imaged

[31] We refer here to version 3 of the Jauzac et al. (2015) paper, which appeared on the arxiv on 2015 October 13. The predictions have changed significantly since versions 1 (2015 September 29), and 2 (2015 October 1).

knots in the host galaxy of 'Refsdal'. All images have been vetted by a group of expert classifiers resulting in a list of 'gold' and 'silver' quality images.

4. The seven lens models have remarkably good fidelity with residual r.m.s. scatter between observed and predicted image positions ranging between 0.″16 and 1.″3.

5. The model predictions agree reasonably well with the observed delays and magnifications of 'Refsdal' (within 68-95% uncertainty, or 10 days in the case of S2-S1), showing that unknown systematics are comparable or smaller than the calculated statistical uncertainties.

6. All models predict that an image of 'Refsdal' will appear near the SX/1.2 location between the submission of this paper and the beginning of 2016. The most likely time for the peak is the first trimester of 2016. Given the long light curve of 'Refsdal' and the predicted brightness of SX/1.2, the image could be visible as soon as MACSJ1149.5+2223 is visible again by HST-WFC3 at the end of October 2015.

7. The past appearance of 'Refsdal' near position SY/1.3 would have been too faint to be detectable in archival images, and thus cannot be tested.

There are two possible outcomes to the work presented in this paper. First, our predictions could be proven correct. This outcome would be an encouraging sign that all the efforts by the community to gather data and improve lens modeling tools are paying off. If, alternatively, our predictions turn out to be wrong, we will have to go back to the drawing board having learned an important lesson about systematic uncertaintes.

The authors thank Raphael Gavazzi for insightful comments on this manuscript. Support by NASA through grants HST-GO-13459 and HST-GO-14041 from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. We are very grateful to the staff of the Space Telescope for their assistance in planning, scheduling, and executing the observations all the observations used in this work. We thank the STSCI and ESO directors for granting Directory Discretionary time to allow for timely follow-up of Refsdal. T.T. gratefully acknowledges the hospitality of the American Academy in Rome and of the Observatorio di Monteporzio Catone, where parts of this manuscript were written. TT acknowledges support by the Packard Foundation in the form of a Packard Research Fellowship. J.M.D acknowledges support of the consolider project CSD2010-00064 and AYA2012-39475-C02-01 funded by the Ministerio de Economia y Competitividad. C.G. acknowledges support by VILLUM FONDEN Young Investigator Programme through grant no. 10123. Support for A.Z. was provided by NASA through Hubble Fellowship grant #HST-HF2-51334.001-A awarded by STScI, which is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555. The work of MO was supported in part by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and Grant-in-Aid for Scientific Research from the JSPS (26800093). Financial support for this work was provided to S.A.R. by NASA through grants HST-HF-51312 and HST-GO-13386 from STScI, which is operated by Associated Universities for Research in Astronomy, Inc. (AURA), under NASA contract NAS 5-26555. A.H. is supported by NASA Headquarters under the NASA Earth and Space Science Fellowship Program - grant ASTRO14F-0007.

REFERENCES

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blandford, R., & Narayan, R. 1986, ApJ, 310, 568
Bolton, A. S., & Burles, S. 2003, ApJ, 592, 17
Bradač, M., Lombardi, M., & Schneider, P. 2004a, A&A, 424, 13
Bradač, M., Schneider, P., Lombardi, M., et al. 2004b, A&A, 423, 797
Bradač, M., Treu, T., Applegate, D., et al. 2009, ApJ, 706, 1201
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, The Astrophysical Journal Supplement, 200, 13
Broadhurst, T., Benítez, N., Coe, D., et al. 2005, ApJ, 621, 53
Coe, D., Bradley, L., & Zitrin, A. 2015, ApJ, 800, 84
Collett, T. E., Marshall, P. J., Auger, M. W., et al. 2013, Monthly Notices of the Royal Astronomical Society, 432, 679
Dalal, N., Hennawi, J. F., & Bode, P. 2005, ApJ, 622, 99
De Marchi, G., & Panagia, N. 2014, MNRAS, 445, 93
Diego, J. M., Protopapas, P., Sandvik, H. B., & Tegmark, M. 2005, MNRAS, 360, 477
Diego, J. M., Tegmark, M., Protopapas, P., & Sandvik, H. B. 2007, MNRAS, 375, 958
Diego, J. M., Broadhurst, T., Chen, C., et al. 2015, ArXiv e-prints, arXiv:1504.05953
Dobler, G., & Keeton, C. R. 2006, ApJ, 653, 1391
Ebeling, H., Barrett, E., Donovan, D., et al. 2007, ApJ, 661, L33
Elíasdóttir, Á., Limousin, M., Richard, J., et al. 2007, ArXiv e-prints, arXiv:0710.5636
Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, ApJ, 289, L1
Fitzpatrick, E. L., & Walborn, N. R. 1990, AJ, 99, 1483
Freedman, W. L., Madore, B. F., Scowcroft, V., et al. 2012, ApJ, 758, 24
Goobar, A., Mörtsell, E., Amanullah, R., & Nugent, P. 2002, A&A, 393, 25
Greene, Z. S., Suyu, S. H., Treu, T., et al. 2013, ApJ, 768, 39
Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, ApJ, 800, 38
Hamuy, M., & Suntzeff, N. B. 1990, AJ, 99, 1146
Holz, D. E. 2001, ApJ, 556, L71
Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2015, ApJ, 799, 12
Jauzac, M., Richard, J., Limousin, M., et al. 2015, ArXiv e-prints, arXiv:1509.08914
Johnson, T. L., Sharon, K., Bayliss, M. B., et al. 2014, ApJ, 797, 48
Jullo, E., & Kneib, J.-P. 2009, MNRAS, 395, 1319
Jullo, E., Kneib, J.-P., Limousin, M., et al. 2007, New Journal of Physics, 9, 447
Karman, W., Grillo, C., Balestra, I., et al. 2015a, ArXiv e-prints, arXiv:1509.07515
Karman, W., Caputi, K. I., Grillo, C., et al. 2015b, A&A, 574, A11
Kassiola, A., & Kovner, I. 1993, ApJ, 417, 450
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Science, 347, 1123
Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, PASP, 121, 1028
Kolatt, T. S., & Bartelmann, M. 1998, MNRAS, 296, 763
Limousin, M., Kneib, J.-P., & Natarajan, P. 2005, MNRAS, 356, 309
Linder, E. V., Wagoner, R. V., & Schneider, P. 1988, ApJ, 324, 786

McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2014, MNRAS, 443, 3631
Meylan, G., Jetzer, P., North, P., et al., eds. 2006, Gravitational Lensing: Strong, Weak and Micro
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2015, ArXiv e-prints 1510.02106, arXiv:1510.02106
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Oguri, M. 2010, PASJ, 62, 1017
—. 2015, MNRAS, 449, L86
Oguri, M., Bayliss, M. B., Dahle, H., et al. 2012, MNRAS, 420, 3213
Oguri, M., & Kawano, Y. 2003, MNRAS, 338, L25
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Oguri, M., Suto, Y., & Turner, E. L. 2003, ApJ, 583, 584
Oguri, M., Schrabback, T., Jullo, E., et al. 2013, MNRAS, 429, 482
Pastorello, A., Pumo, M. L., Navasardyan, H., et al. 2012, A&A, 537, A141
Popper, K. R. 1992, The logic of scientific discovery (London: Routledge, —c1992)
Postman, M., Coe, D., Benítez, N., et al. 2012, The Astrophysical Journal Supplement, 199, 25
Quimby, R. M., Oguri, M., More, A., et al. 2014, Science, 344, 396
Refsdal, S. 1964, MNRAS, 128, 307
Riess, A. G., Macri, L., Casertano, S., et al. 2011, ApJ, 730, 119
Rodney, S. A., Patel, B., Scolnic, D., et al. 2015, ApJ, 811, 70
Schmidt, K. B., Treu, T., Brammer, G. B., et al. 2014, The Astrophysical Journal Letters, 782, L36
Schneider, P. 1985, A&A, 143, 413
—. 2014, A&A, 568, L2
Schneider, P., & Sluse, D. 2013, A&A, 559, A37
—. 2014, A&A, 564, A103
Sendra, I., Diego, J. M., Broadhurst, T., & Lazkoz, R. 2014, MNRAS, 437, 2642
Shapiro, I. I. 1964, Physical Review Letters, 13, 789
Sharon, K., & Johnson, T. L. 2015, ApJ, 800, L26
Sirianni, M., Jee, M. J., Benítez, N., et al. 2005, PASP, 117, 1049
Smith, G. P., Ebeling, H., Limousin, M., et al. 2009, ApJ, 707, L163
Sullivan, M., Ellis, R., Nugent, P., Smail, I., & Madau, P. 2000, MNRAS, 319, 549
Suyu, S. H., & Halkola, A. 2010, A&A, 524, A94
Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201
Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10
Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, ApJ, 766, 70
Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJ, 788, L35
Taddia, F., Stritzinger, M. D., Sollerman, J., et al. 2012, A&A, 537, A140
Tewes, M., Courbin, F., & Meylan, G. 2013a, A&A, 553, A120
Tewes, M., Courbin, F., Meylan, G., et al. 2013b, A&A, 556, A22
Treu, T. 2010, ARA&A, 48, 87
Treu, T., & Ellis, R. S. 2015, Contemporary Physics, 56, 17
Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, ApJ, 812, 114
Wang, X., Hoag, A., Huang, K.-H., et al. 2015, ApJ, 811, 29
Xu, D., Sluse, D., Schneider, P., et al. 2015, ArXiv e-prints, arXiv:1507.07937
Zheng, W., Postman, M., Zitrin, A., et al. 2012, Nature, 489, 406
Zitrin, A., & Broadhurst, T. 2009, The Astrophysical Journal Letters, 703, L132
Zitrin, A., & Broadhurst, T. 2009, ApJ, 703, L132
Zitrin, A., Broadhurst, T., Bartelmann, M., et al. 2012a, MNRAS, 423, 2308
Zitrin, A., Broadhurst, T., Umetsu, K., et al. 2009, MNRAS, 396, 1985
Zitrin, A., Rosati, P., Nonino, M., et al. 2012b, ApJ, 749, 97
Zitrin, A., Meneghetti, M., Umetsu, K., et al. 2013, ApJ, 762, L30
Zitrin, A., Fabris, A., Merten, J., et al. 2015, ApJ, 801, 44