H0LiCOW VI. Testing the fidelity of lensed quasar host galaxy reconstruction

Xuheng Ding,[1,2]⋆ Kai Liao,[2] Tommaso Treu,[2] Sherry H. Suyu,[3,4,5] Geoff C.-F. Chen, [6,4] Matthew W. Auger, [7] Philip J. Marshall,[8] Adriano Agnello,[2,9] Frederic Courbin, [10] Anna M. Nierenberg,[11] Cristian E. Rusu,[6] Dominique Sluse,[12] Alessandro Sonnenfeld,[13] Kenneth C. Wong[14,4]

[1]*Department of Astronomy, Beijing Normal University, Beijing 100875, China*
[2]*Department of Physics and Astronomy, University of California, Los Angeles, CA, 90095-1547, USA*
[3]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany*
[4]*Institute of Astronomy and Astrophysics, Academia Sinica, P.O. Box 23-141, Taipei 10617, Taiwan*
[5]*Physik-Department, Technische Universität München, James-Franck-Straße 1, 85748 Garching, Germany*
[6]*Department of Physics, University of California, Davis, CA 95616, USA*
[7]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[8]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA*
[9]*European Southern Observatories, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany*
[10]*Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[11]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus OH 43210, USA*
[12]*STAR Institute, Quartier Agora - Allée du six Août, 19c B-4000 Liège, Belgium*
[13]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[14]*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*

ABSTRACT

The empirical correlation between the mass of a super-massive black hole ($\mathcal{M}_{BH}$) and its host galaxy properties is widely considered to be evidence of their co-evolution. A powerful way to test the co-evolution scenario and learn about the feedback processes linking galaxies and nuclear activity is to measure these correlations as a function of redshift. Unfortunately, currently $\mathcal{M}_{BH}$ can only be estimated in active galaxies at cosmological distances. At these distances, bright active galactic nuclei (AGN) can outshine the host galaxy, making it extremely difficult to measure the host's luminosity. Strongly lensed AGNs provide in principle a great opportunity to improve the sensitivity and accuracy of the host galaxy luminosity measurements as the host galaxy is magnified and more easily separated from the point source, provided the lens model is sufficiently accurate. In order to measure the $\mathcal{M}_{BH}-L$ correlation with strong lensing, it is necessary to ensure that the lens modelling is accurate, and that the host galaxy luminosity can be recovered to at least a precision and accuracy better than that of the typical $\mathcal{M}_{BH}$ measurement. We carry out extensive and realistic simulations of deep *Hubble Space Telescope* observations of lensed AGNs obtained by our collaboration. We show that the host galaxy luminosity can be recovered with better accuracy and precision than the typical uncertainty on $\mathcal{M}_{BH}(\sim 0.5$ dex) for hosts as faint as $2-4$ magnitudes dimmer than the AGN itself. Our simulations will be used to estimate bias and uncertainties on the actual measurements to be presented in a future paper.

Key words: galaxies: evolution — black hole physics — galaxies: active

1 INTRODUCTION

In the past two decades, tight correlations have been discovered between supermassive black hole mass ($\mathcal{M}_{BH}$) and the

⋆ E-mail: dxh@astro.ucla.edu

arXiv:1610.08504v1 [astro-ph.GA] 26 Oct 2016

properties of the host galaxy, including luminosity ($L_{\rm bulge}$), stellar mass ($\mathcal{M}_*$) and stellar velocity dispersion (σ_*) (e.g., Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2001; Marconi & Hunt 2003; Häring & Rix 2004; Gültekin et al. 2009; Graham et al. 2011; Beifiori et al. 2012; Park et al. 2015; Kormendy & Ho 2013).

These tight correlations are usually considered to be the result of physical coupling between supermassive BH growth and galaxy evolution (see, however, Peng 2007, for a different view). The origin of this co-evolution is still an open question, central for our understanding of the role of active galactic nuclei (AGN) feedback in galaxy formation and evolution. One key to study this coupling is to trace the correlations to higher redshift, determining how and when they emerged and evolved (e.g., Treu et al. 2004; Salviander et al. 2006; Woo et al. 2006; Jahnke et al. 2009; DeGraf et al. 2015).

Unfortunately, there are challenges in mapping these correlations beyond the local universe (i.e. redshift $z \sim 0.1$). At cosmologically significant look-back times, the sphere of influence of the black hole cannot be directly resolved, forcing[1] one to study active black holes, where $\mathcal{M}_{\rm BH}$ can be estimated indirectly using the properties of the broad emission lines (see recent reviews by Shen 2013; Peterson 2014). However, the presence of a bright point source makes it very hard to study the properties of the host galaxy, which is typically barely resolved and faint owing to cosmological surface brightness dimming. For these reasons, much attention so far has been devoted to relatively faint AGN (Seyfert 1s) where the brightness of the point source is comparable to that of the host (Woo et al. 2006; Treu et al. 2007; Bennert et al. 2011; Park et al. 2015), and very little is known about the hosts of the most luminous AGNs, hosting the most massive black holes (see, e.g., Walter et al. 2004, for a few exceptions).

Strong gravitational lensing (see, e.g., Courbin et al. 2002; Schneider et al. 2006; Treu 2010, for reviews) provides a unique opportunity to make progress in the study of the luminosity of the host galaxies of the most luminous AGN in the distant universe. Even with lensing magnification, the AGN is unresolved at *Hubble Space Telescope* (*HST*) resolution, while the host galaxy is typically well resolved. Peng et al. (2006b) and Peng et al. (2006a) first studied the $\mathcal{M}_{\rm BH}$-$L_{\rm host}$ relation using gravitational lensed AGN image out to redshift $2 - 4$, finding that host galaxies were typically underluminous compared to the local universe, when taking the evolution of stellar populations into account, provided that systematic uncertainties can be controlled.

Given the advances in both the quality of the data and lensing modelling techniques in the past decade, it is time to revisit this measurement. In this paper we present extensive and realistic simulations of ultradeep *HST* images of lensed AGNs obtained with the goal of measuring cosmological parameters from gravitational time delays. We then analyze the simulated images with the state of the art code

GLEE[2] as if they were real images, in order to quantify realistic uncertainties and uncover biases and systematic errors associated with the measurement of the lensed host galaxy luminosity. The data analyzed here represent a dramatic improvement over the NICMOS data used a decade ago: not only is WFC3-IR at least a factor of 10 faster (in terms of time required to reach a certain depth), but our exposure times are typically $5-10$ longer for each object.

Likewise, GLEE has several advantages with respect to the lensing code used by Peng et al. (2006b). First, the host galaxy is not assumed to have a regular surface brightness profile, but is modelled as a grid of surface brightness pixels obtained through regularized linear inversion. Thus, it has the flexibility to describe the complex morphologies observed in some cases (Suyu et al. 2013, 2014). This allows the observer to choose the most appropriate parametrized form to describe the surface brightness distribution of each specific system. Second, the deflector mass distribution can be described with models that are more flexible than singular isothermal ellipsoid thus eliminating a potential source of bias in the host galaxy luminosity measurement (e.g. Marshall et al. 2007).

In addition to quantifying bias and uncertainties for each system in our sample, we study trends with lensing configuration, host galaxy brightness and host galaxy to AGN contrast, which should be general enough to provide guidance for future studies. We note that the $\mathcal{M}_{\rm BH}$ estimates are considered to be uncertain at the ~ 0.5 dex level even in the local universe (Vestergaard & Peterson 2006). Thus, the natural goal of this simulation exercise is to find out whether the random error and bias on $L_{\rm host}$ can be constrained or corrected to a level that is smaller than 1.25 magnitudes. In a future paper we will apply the results of our simulations to the actual data and combine them with spectroscopy to construct the $\mathcal{M}_{\rm BH}$-$L_{\rm host}$ relation.

The paper is organized as follows. We briefly summarize the properties of the data set in Section 2. We describe our simulation pipeline in Section 3. In section 4 we present our main results. Discussion and summary are presented in Section 5 and 6, respectively. Magnitudes are given in the AB System.

2   DATA

The simulations are intended to reproduce as closely as possible the deep *HST* observations available for a sample of eight lenses currently being analyzed by our team. More information and references for each system are given in Table 1. The real data are taken from three *HST* programs and are described in detail elsewhere. Briefly, the data for RXJ1131 were obtained as part of *HST* program (GO-9744; PI: Kochanek) and, together with data for three lenses from program *HST*-GO-12889 (PI: Suyu), they form the bulk of the H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW) sample. We refer to Suyu et al. (2013, 2016) for a description of these data. Four systems have been imaged in cycle 23 as part of Program *HST*-GO-14254 (PI: Treu). The cycle

---

[1] In very rare cases gravitational lensing can also be used to estimate the black hole mass using the properties of the central image (Quinn et al. 2016; Wong et al. 2015; Winn et al. 2004; Tamura et al. 2015).

[2] Developed by Suyu & Halkola (2010) based on Suyu et al. (2006) and Halkola et al. (2008)

23 sample was selected to have measured time delays from the COSMOGRAIL collaboration (Eigenbrod et al. 2005) [3], confirmed to have extended host galaxy features in near infrared adaptive optics images taken at the Keck Telescope (PI: Treu) (Agnello et al. 2016), before being imaged with *HST*. More details on these observations will be given by Agnello et al. (2017, in preparation). We note that we do not consider the system B1608+656 in this work even though it is an integral part of the H0LiCOW sample. That system is different from the other eight considered here in that the lensed source does not host a bright AGN, but it is rather an active nucleus without broad emission lines (Fassnacht et al. 1996). Therefore broad line based $\mathcal{M}_{\mathrm{BH}}$ estimates cannot be obtained for this system.

3 SIMULATIONS

The purpose of this section is to describe simulations of lens systems that have similar host luminosity and lensing parameters to the real data. Conceptually, the following steps need to be taken in order to carry out a realistic simulation of strongly lensed AGN.

(i) Compute high resolution simulated images of the lensed AGN, host, and deflector light.

(ii) Convolve with the telescope point spread function (PSF) and sample according to the pixel size of the camera.

(iii) Rebin the high resolution image to the actual data resolution.

(iv) Add noise.

(v) Repeat steps (iii) and (iv) at different positions to simulate dithering and drizzle the individual images to recover some of the resolution lost to undersampling as is done with the real observations. This step introduces correlated noise.

In the remainder of this section we describe these steps in detail, referring to the specific case of HE0435−1223 as an example, when necessary. More details about this system are available in five papers detailing the cosmographic analysis (Suyu et al. 2016; Sluse et al. 2016; Rusu et al. 2016a; Wong et al. 2016; Bonvin et al. 2016).

3.1 Simulating high resolution noiseless images

We start by generating a high resolution PSF using TINY TIM[4]. In practice, TINY TIM will not yield a sufficiently good model of the PSF for our application. However, for the purpose of this simulation exercise we assume that we know the PSF, and that the details of its shape are irrelevant as long as they possess a sufficiently realistic description of *HST*-quality PSF profiles. We have developed an iterative PSF modelling process which can accurately recover the PSF for real *HST* observations (Chen et al. 2016; Wong et al. 2016, Suyu et al. in preparation). This iterative process will be used when analyzing the real lens systems, and the residual uncertainties taken into account when measuring the host

galaxy luminosity. An alternative scheme for modelling the PSF in ground based data is described by Rusu et al. (2016b) and applied to adaptive optics observations of gravitationally lensed AGNs from SDSS Quasar Lens Search (SQLS).

In the case of HE0435, the high resolution PSF was oversampled by a linear factor of 4, i.e. using a pixel size of $0\overset{''}{.}13/4 = 0\overset{''}{.}0325$ in steps (i) and (ii) listed above. The same factor for pixel size is used for the other lenses imaged with WFC3-IR except for HE1104, for which we used $0\overset{''}{.}13/6 = 0\overset{''}{.}0217$, in order to simulate the actual dither pattern used during the observations (see Section 3.2 for details).

Once the PSF is generated, we simulate the lensing system itself. The surface brightness is described as a hybrid model (Suyu 2012) comprised of (1) point sources for the lensed AGN in the image plane, (2) a simply-parametrized model for the deflector surface brightness in the image plane, and (3) a simply-parametrized model for the host galaxy in the source plane.

For simplicity, in this study we adopt a single Sersic (1968) profile to describe the surface brightness profile of the host (step (i)) and of the deflector. This Sérsic profile is commonly used in the literature and provides a reasonably accurate description of the surface brightness of galaxies, ranging from exponential disks to de Vaucouleurs (1948) profiles. In future papers, when analyzing real systems, we will examine the pixellated reconstructed surface brightness profile of each host galaxy and identify the most appropriate parametrization. Likewise we will adopt different (or multiple) surface brightness profiles to describe the deflector light, if warranted by the data as in the case of RXJ1131 (Suyu et al. 2013, 2014). We note that choosing a simply-parametrized description of the surface brightness pixels is necessary in order to obtain results directly comparable with the rest of the literature and also to extrapolate the observed flux to an infinite aperture, as it is common practice (the extrapolation is typically several tens of percent in total luminosity, as the reconstructed source image does not extend much further than the effective radius).

The Sersic (1968) profile is parametrized as follows:

$$I(R) = I_e \exp\left[-k\left(\left(\frac{R}{R_{\mathrm{eff}}}\right)^{1/\mathrm{n}} - 1\right)\right], \qquad (1)$$

$$R(x, y, q) = \sqrt{qx^2 + y^2/q}. \qquad (2)$$

I_e is the pixel surface brightness at the effective radius R_{eff}. The Sérsic index n controls the shape of the radial surface brightness profile; a larger n corresponds to a steeper inner profile and a highly extended outer wing. The constant k depends on n so as to ensure that the isophote at $R = R_{\mathrm{eff}}$ encloses half of the total light, i.e. that R_{eff} is the half-light radius (Ciotti & Bertin 1999). The symbol q denotes the axis ratio.

The deflector mass density profile is described by a class of elliptical power-law models. The surface mass density is parametrized as

$$\Sigma(x, y) = \Sigma_{cr} \frac{3 - \gamma'}{1 + q_m} \left(\frac{\sqrt{X^2 + q_m^{-2}Y^2}}{R_{\mathrm{E}}}\right)^{1 - \gamma'}, \qquad (3)$$

where Σ_{cr} is the *critical density*, depending on the angular diameter distances to the deflector (D_{d}), to the source (D_{s}) and between the deflector and the source (D_{ds}). The Einstein radius R_{E} is chosen such that, in the spherical limit

Table 1. Summary of lensed AGN properties and observational setup.

Name	z_s	camera	filter	exposure time(s)	RA	DEC	*HST* program	PI	AGN image	Approximate Magnification	References
HE0435−1223	1.69	WFC3-IR	F160W	9940	04:38:14.90	-12:17:14.4	12889	S. H. Suyu	4	4.0	(1), (2)
HE1104−1805	2.32	WFC3-IR	F160W	14700	11:06:33.45	-18:21:24.2	12889	S. H. Suyu	2	2.2	(3), (4)
SDSS1206+4332	1.79	WFC3-IR	F160W	8460	12:06:29.65	+43:32:17.6	14254	T. Treu	2/4	3.0	(5), (6)
WFI2033−4723	1.66	WFC3-IR	F160W	14470	20:33:42.08	-47:23:43.0	12889	S. H. Suyu	4	3.2	(7), (2)
HE0047−1756	1.66	WFC3-UVIS	F814W	9710	00:50:27.83	-17:40:08.8	14254	T. Treu	2	2.8	(8), (9)
SDSS0246−0825	1.68	WFC3-UVIS	F814W	9680	02:46:34.11	-08:25:36.2	14254	T. Treu	2	2.8	(10), (11)
HS 2209+1914	1.07	WFC3-UVIS	F814W	14240	22:11:30.30	+19:29:12.0	14254	T. Treu	2	2.6	(12), (13)
RX J1131−1231	0.65	ACS	F814W	2085	11:31:51.6	-12:31:57	9744	C.S. Kochanek	4	4.3	(14), (15)

Note:− For conciseness, we abbreviate each lens name to four digits (e.g HE0435−1223 to HE0435). Approximate magnification is given in magnitudes.

References:− (1) Wisotzki et al. (2002); (2) Sluse et al. (2012); (3) Wisotzki et al. (1993); (4) Smette et al. (1995); (5) Oguri et al. (2005) (6) Eulaers et al. (2013); (7) Morgan et al. (2004); (8) Wisotzki et al. (2004); (9) Ofek et al. (2006); (10) Inada et al. (2005); (11) Eigenbrod et al. (2007); (12) Hagen et al. (1999); (13) Chantry et al. (2010); (14) Sluse et al. (2003); (15) Sluse et al. (2007).

($q_m = 1$), it encloses a mean surface density equal to Σ_{cr}. This is also the radius of a ring traced by the host of the AGN when this is exactly aligned with the lens galaxy. The exponent γ' is the so-called slope of the mass density profile, ≈ 2 for massive elliptical galaxies (Treu & Koopmans 2002, 2004). The projected axis ratio of the mass density profile is described by q_m. We refer the reader to the reviews by Schneider (2006); Bartelmann (2010); Treu (2010) for more details.

Having specified the intrinsic surface brightness of the host galaxy and the mass model of the deflector, the lensed images can be computed by solving the lens equation. In practice, this is done using the code Pylens[5] written by one of us (M.W.A.) and functionally tested extensively by Meng et al. (2015).

In order to create simulations that are as realistic as possible, we first obtained rough models of the real systems. This set of rough models enables us to set realistic values and ranges (see Table 2 and Section 4 as for details) to the input parameters needed for Pylens simulations of the host galaxy and deflector surface brightness. The AGN images are added directly in the image plane as PSFs scaled to match the observed flux. The procedure is illustrated in Fig. 1, panels *a* and *b*.

3.2   Simulating dithering

The *HST* PSF is undersampled by the camera setups that we simulate in this study. Therefore, a common strategy consists of dithering the observations by non-integer number of pixels in order to recover some of the information lost via a process known as drizzling[6].

In order to simulate the undersampling and dithering, we bin the high resolution image (4×4 for HE0345 as an example), varying the location of the binning grid so as to mimic the actual dithering pattern adopted in the *HST* observations (see Fig. 1-(c), 2 for an illustration).

3.3   Adding noise

The noise in each pixel is due as usual to read noise, background noise, and Poisson noise from the astronomical sources themselves. The read noise and background noise level are measured directly from empty regions of the real data while the Poisson noise level is related to the effective exposure time in each pixel. For the systems observed with WFC3-IR, three exposures were obtained at each dither position: one short exposure ($\sim 43s$) followed by two long exposures ($\sim 599s$). This strategy allowed us to obtain unsaturated images of the brightest pixels in the center of the AGN, while minimizing overheads and maximizing the signal to noise ratio on the low surface brightness parts of the images. In order to reproduce this strategy, for the central AGN area ($3-5$ pixels in the AGN center), we consider the noise corresponding to the short exposures. For the rest of the image we consider the full exposure time. By comparing the final simulation to the real sample, we find that the simulated noise properties match the noise properties of the observed data very well. (Fig. 1-(d) shows the image after adding noise).

3.4   Drizzling

The final step is to combine the 8 images by using MultiDrizzle[7]. The images are drizzled to a final pixel scale $0.''08$ pixel^{-1} as in real data. The final simulated image is shown on Fig. 1-(e). Of course, dithered images of the PSF are also drizzled following the same exact procedure in order to produce a realistic PSF for the final combined image.

3.5   Modifications for WFC3-UVIS and ACS

Our sample includes systems observed with WFC3-UVIS and the Advanced Camera for Surveys (ACS). The pipeline used to simulate the lensed AGN in these bands is similar to that used for WFC3-IR and described above. The main difference is that the pixel scale is $0.''03$ for WFC-UVIS and

---

[5] https://github.com/tommasotreu/HIGHRESOLUTIONIMAGING/tree/master/code/pylens
[6] http://www.stsci.edu/~fruchter/dither/#basics

[7] MultiDrizzle is a product of the Space Telescope Science Institute, which is operated by AURA for NASA, see http://www.stsci.edu/hst/wfpc2/analysis/drizzle.html for more information.

(a) Host arc and lens image, no AGN yet.

(b) Adding AGN images.

(c) Binning into 8 images.

(d) Adding noise.

(e) Final image, after drizzling.

(f) Simulated 48 images VS the real on the bottom right.
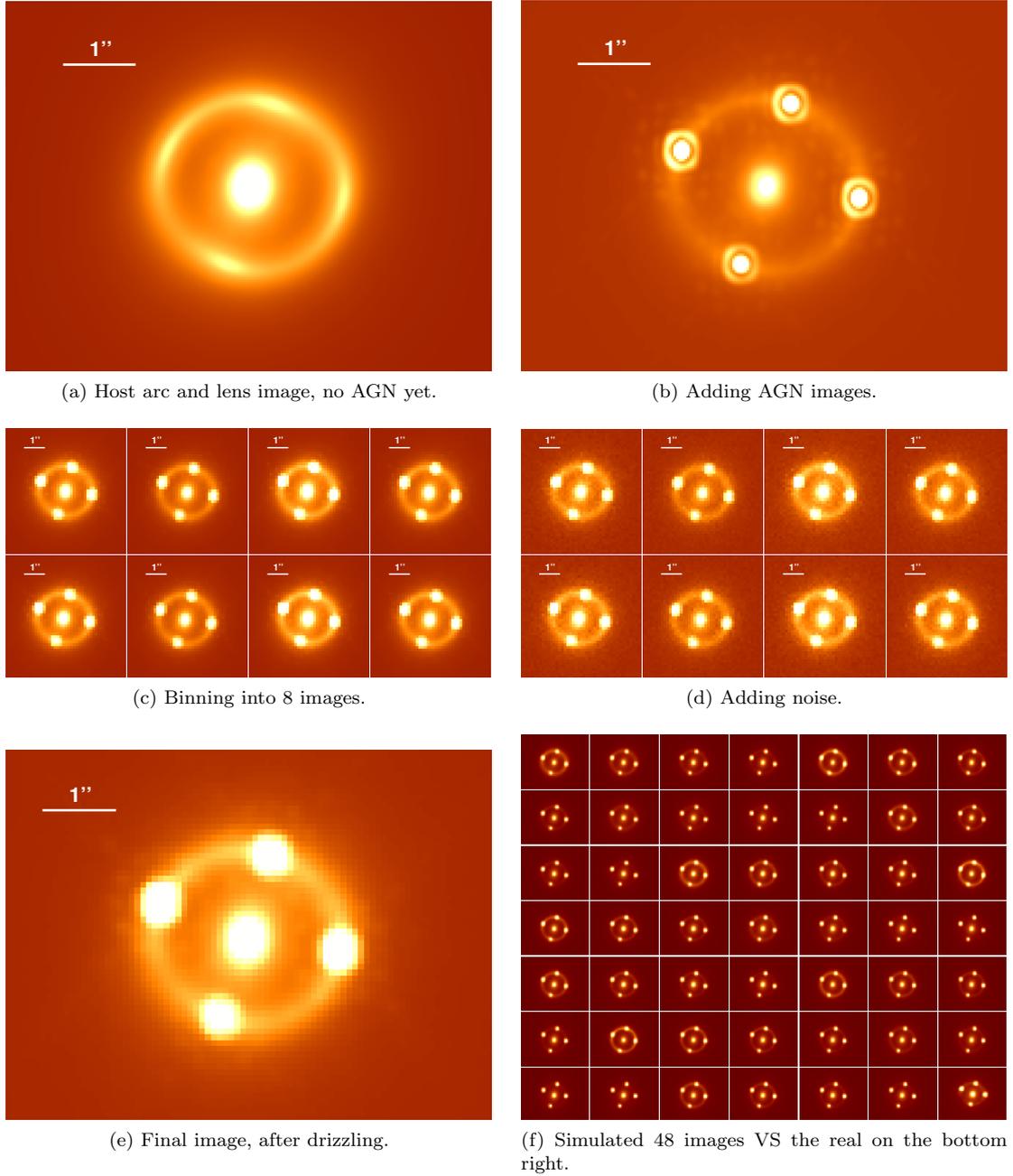
Figure 1. Illustration of the simulation process using a linear color scale. (a) Simulating the host arc and lens image with PYLENS, (b) Adding AGN image as appropriately normalized PSFs, (c) Binning into a lower resolution images in 8 different ways following the actual dithering pattern, (d) Including the noise information, (e) Drizzling 8 images together to obtain the final image, (f) Montage of 48 simulated images plus the real one in the bottom right panel, presented on an oversampled grid. It is clearly seen that the arc brightness is varying from case to case and brackets the real data.

$0.''05$ for ACS after drizzling. Also, the noise levels and dither patterns are modified to match the actual observations.

4 RESULTS

In this section we present the results of our simulations of eight systems. Four of the systems are imaged with WFC3-IR through filter F160W, three with WFC3-UVIS through filter F814W, and the remaining one with ACS through filter F814W. For each system we construct 48 realizations by changing the parameters describing the unknown properties of the host galaxy and mass distribution of the deflector, as summarized in Table 2. As mentioned above, the set of unknown parameters is chosen as to represent the range of expected true values in this particular data set. The parameters describing the surface brightness of the deflector and the brightness of the point sources are measured directly
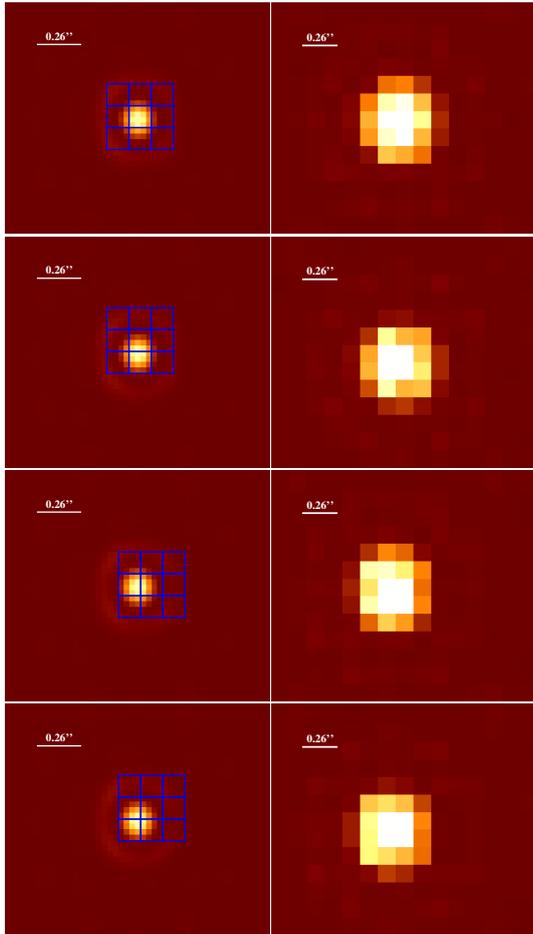
Figure 2. We illustrate 4 different ways to bin high resolution image pixels into lower resolution, using the same linear color stretch. Left panels show the high resolution images before binning. By combining all the pixel inside the blue box, we get the low resolution images shown in the right panels. To make the dither pattern more evident, we only show 9 blue boxes, rather than cover the entire plane. $0.''26$ corresponds to 8 and 2 pixels in left and right panels, respectively.

from the data, with negligible uncertainty for our purposes and are therefore kept fixed. Of course, the brightness of the AGNs are the key parameters for this study. However, the AGN flux can be measured with very high precision and residual uncertainties have negligible effect on the host galaxy reconstruction, provided a good PSF model is available. We will return to this issue and account for for residual uncertainties when we analyze the actual data in a future paper. Examples of simulated images are shown in Fig. 3.

4.1   Glee fitting

We use Glee to reconstruct the host galaxy properties from the simulated lensed AGN and analyze the resulting uncertainty and bias on the main parameters including magnitude, effective radius and sersic index of the host galaxy. We are mostly interested in magnitude, which is the key quantity entering the correlation with $\mathcal{M}_{BH}$.

The inputs to Glee are the simulated image, a noise map, a mask identifying the lensed image of the arc, and

Table 2. Parameter grid

parameter	simulated values
n	2, 4
l_{eta}	0.8, 1.0, 1.2
WFC-IR images	
S_{re} (arc second)	0.39, 0.78
HE1104 S_{Mag}	21.0, 21.5, 22.0, 22.5
HE0435 S_{Mag}	20.5, 21.0, 21.5, 22.0
WFI2033 S_{Mag}	20.5, 21.0, 21.5, 22.0
SDSS1206 S_{Mag}	20.0, 20.5, 21.0, 21.5
WFC-UVIS images	
S_{re} (arc second)	0.3, 0.6
SDSS0246 S_{Mag}	21.5, 22.0, 22.5, 23.0
HS2209 S_{Mag}	21.0, 21.5, 22.0, 22.5
HE0047 S_{Mag}	21.5, 22.0, 22.5, 23.0
ACS images	
S_{re} (arc second)	0.5, 1.0
RXJ1131 S_{Mag}	19.0, 19.5, 20.0, 20.5

Note:− Grid of simulated parameters. n, S_{re} and S_{Mag} are the Sérsic index, $R_{\rm eff}$ and the magnitude of the source brightness respectively. After preliminary tests we found that the fidelity of the reconstruction does not depend significantly on source ellipticity. So, we decided to set all ellipticities to 0.7 to save computation time. The magnitude ranges are selected to bracket the range of observed brightness. l_{eta} describes the slope of the mass density profile of the deflector ($l_{eta} = \gamma' - 1$, as defined by Eq. 3).

a PSF. The crucial point is that Glee describes the host galaxy as a set of surface brightness pixels, not as a Sérsic profile. This choice enables the extraction of more information, down to the noise level, even though one still needs to fit the reconstructed source plane surface brightness map with a Sérsic profile in order to obtain the desired parameters.

In principle, the mask containing the lensed host galaxy light should be selected to be as large as possible, including pixels with surface brightness close to the noise level. However, these masks would be extremely large, and thus too computationally expensive. Thus, in order to keep the run time manageable (of order a week per system on a multicore linux box), and in order to closely mimic what is done in the H0LiCOW analysis, we select a mask corresponding to a thick ring around the Einstein radius (see examples in Fig. 4).

Importantly, the image pixels used in modelling the AGN host are kept fixed to those in the selected arc mask. For a given set of lens mass parameters, these image pixels in the arc mask are casted back to the source plane, forming a region on the source plane where we would have information to reconstruct the AGN host surface brightness distribution. We then use a minimal rectangular grid with fixed dimensions that encapsulates this source reconstruction region for our AGN host surface brightness modelling. In this way, the data (image pixels) are kept fixed throughout the modelling as the lens mass parameters vary, which is important for sampling the posterior of the data and for deriving lens parameter constraints. Even though the source grid changes as the modelling proceeds, the source pixel parameters are
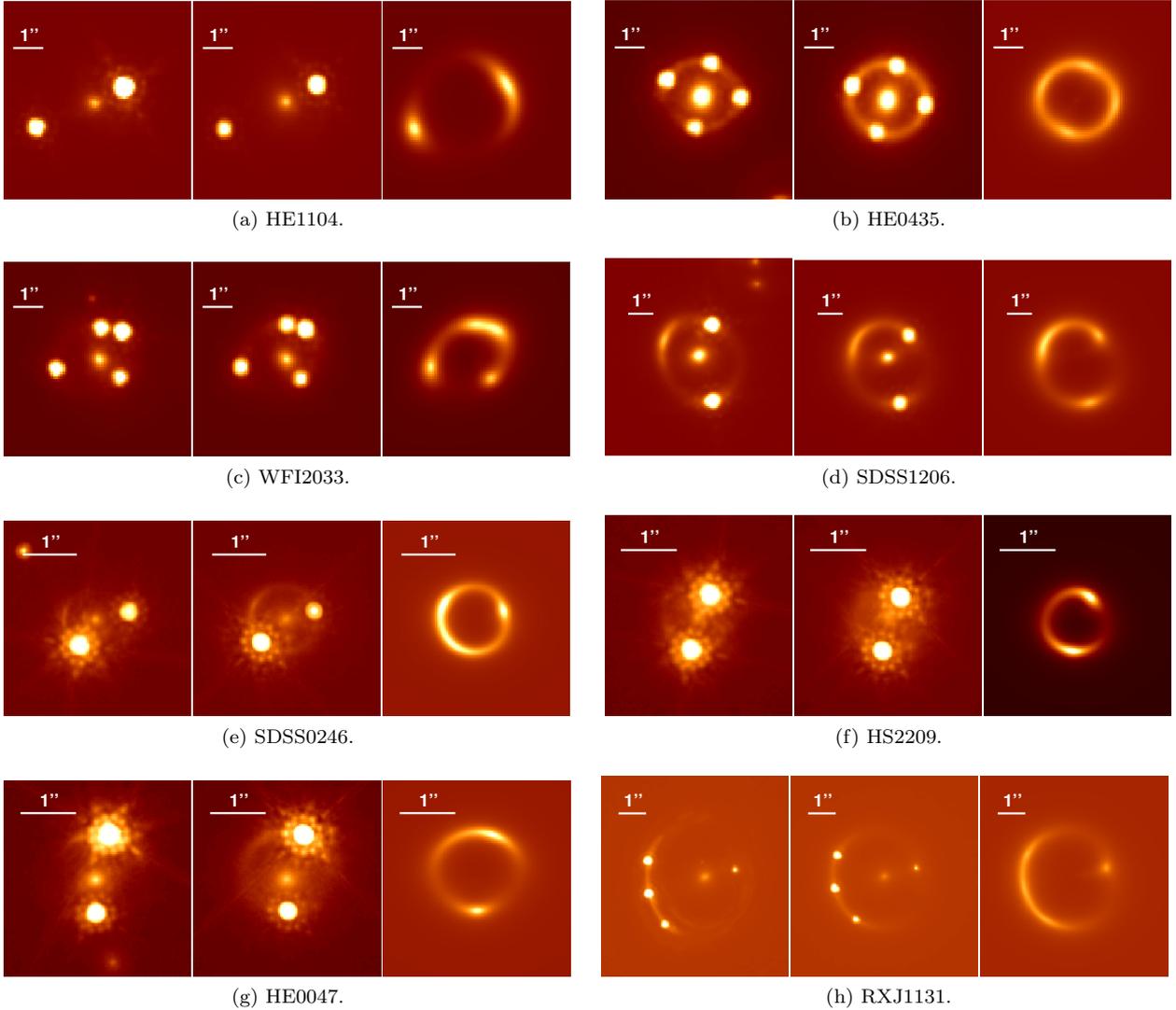
Figure 3. For each lens the left panel shows the observed image and the middle panel shows the simulation with the same stretch. (a)-(d) and (h) are in linear color scale; (e)-(g) are in log color scale. For clarity, we also show in the right panels the simulated arc image without the bright AGN and deflector, using a different stretch. (a)-(d) are from WFC3-IR observation, (e)-(g) are from WFC-UVIS and (h) is from ACS.

marginalized (see, e.g., Suyu & Halkola 2010) when computing the posterior of the data, so the constraints on the lens parameters account for the variable source region. To study the AGN host, we then consider the best-fitting source reconstructions from a set of lens mass models whose parameters are obtained from the modelling.

In order to avoid overfitting the noise, regularization is introduced in the framework of Bayesian statistics (Suyu 2012; Suyu et al. 2006). Following common practice (e.g. Suyu et al. 2013, 2014; Agnello et al. 2016), in order to account for systematic errors in the interpolation of the sub-sampled PSF, we boost the noise estimate in the central pixels of each AGN image to be effectively infinite. In practice this means that the reconstruction of the host galaxy brightness in the central pixels is dominated by the regularization. This procedure is designed to account for PSF modelling systematic errors and helps avoid underestimating the uncertainties on the final parameter estimates.

Briefly, GLEE models each system by simultaneously fitting the following ingredients through a Markov Chain Monte Carlo process:

(i) A set of scaled PSFs to represent the AGN images;

(ii) A Sérsic profile to represent the surface brightness of the deflector;

(iii) A grid of pixels in the source plane representing the surface brightness of the host galaxy;

(iv) A mass model for the deflector.

The modelled host image is generated onto a 30×30 pixel grid. The pixel resolution for each reconstruction is determined by the arc mask region and light profile of reconstruction. We have tested that the results of our analysis are the same if a 40x40 source grid is used instead of 30x30, using a subset of models. Similar tests on real data performed by our group show that the effect of discretization bias (Vegetti & Koopmans 2009; Nightingale & Dye 2015)

(a) Example of a simulated image of HE0435.

(b) Residual from best fit AGN and deflector brightness, showing the lensed host galaxy.

(c) Best fit model of the host galaxy in the image plane.

(d) Normalized residuals.

(e) Reconstructed host in the source plane.

(f) True host image in the source plane.
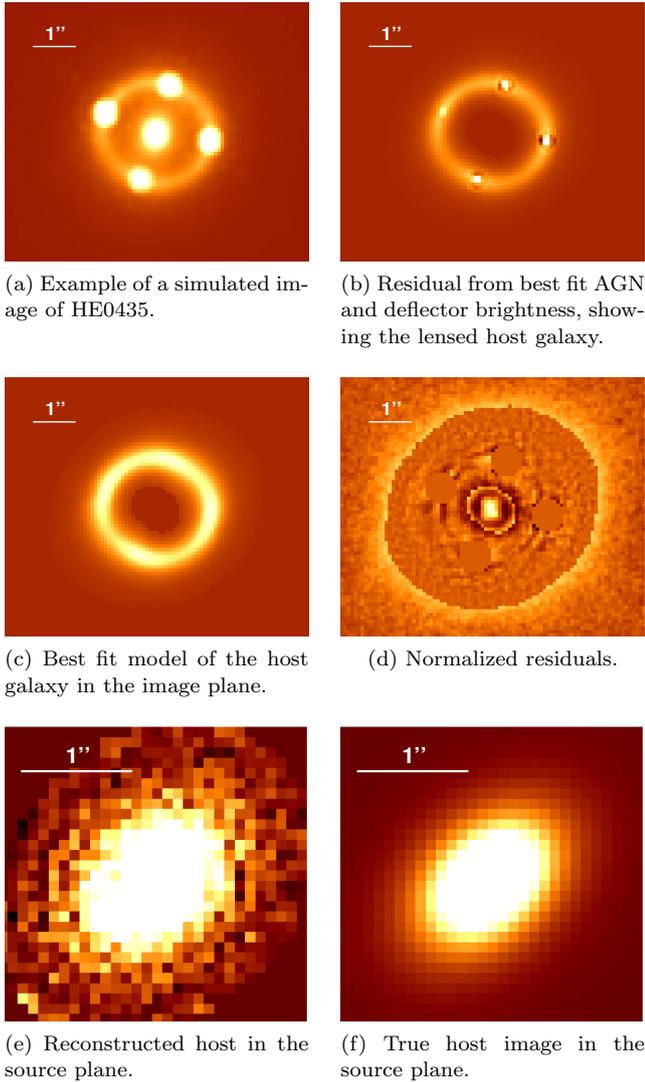
Figure 4. Illustration of the key ingredients in GLEE reconstruction of the host galaxy image in the source plane. The mask region can be clearly seen in (c) and (d).

is small provided that the mask is large enough and the source resolution is high enough (Suyu et al. 2013; Wong et al. 2016). Moreover, we have tested reconstructing the host galaxy by using the true slope of the mass density profile, and the results on the AGN host flux are unchanged.

We illustrate the steps of GLEE modelling in Fig. 4, taking HE0435 as an example. The normalized residual map (i.e. Fig. 4-(d)) shows that the light model for the deflector is imperfect near the center. This could be fixed by adopting a more extended mask, covering the central region of the image, at a cost of more computing time. In order to test whether this residual could cause systematic uncertainties we repeated the fit for a subset of objects zeroing the flux associated with the deflector and the results on the lensed arc are unchanged. Thus, we conclude that this residual mismatch is not important for the host galaxy property studies and we do not try to reduce it further in this work.

4.2 Analysis

In order to obtain a description of the total flux of the host galaxy, accounting for the flux lost by masking, we cannot rely on direct integration of the surface brightness map in the source plane. Instead we rely on the software GALFIT (Peng et al. 2002) to fit a Sérsic profile to the data, thus reproducing the procedure used for isolated galaxies. Of course, this approach will provide not only the total magnitude but also the effective radius S_{re} and Sérsic index n.

In some cases, the reconstructed images are very faint and cover only a small radial range across an arcsecond. This can lead to unphysical results from GALFIT fitting. Thus, we constrain n to be in the range $1-4$, typical of normal galaxies. We also test the effects of fixing n, as would be the case if we knew that the host galaxy of the bright AGN was a pure elliptical or disk galaxy. Errors on the GALFIT fit for each simulated image are obtained by repeating the fit on 30 sets of source reconstructions taken from the MCMC chain output by GLEE, and computing the mean and standard deviation of the results. We present four sets of comparisons between input and recovered values including total magnitude (n free and n fixed), effective radius S_{re} and Sérsic index n.

In the following subsections we describe the results for each system. The lens names are abbreviated in the rest of the paper. See Table 1 for the full name.

4.2.1   HE1104

The results are summarized in Fig. 5. Panels (a), (b), and (c) summarize the bias, i.e. the difference between inferred and input value of each parameter, allowing n to float in the range $1-4$ (filled circles). The bias when n is kept fixed to the input value is represented as filled stars. The serial number representing different parameter configurations in each panel is explained in Table 3. The random uncertainties are estimated as the scatter in the GALFIT inferred parameters for the 30 source reconstructions taken from the GLEE MCMC chain output. (This source of random error dominates over the random uncertainties associated with each pixel value, since it includes uncertainties in the lensing parameters.) As expected, given the high signal to noise ratio of the data, the random uncertainties are very small and the error bars are barely visible in Panels (a), (b), and (c).

The main result is that the inferred magnitudes tend to be somewhat larger (i.e. fainter) than the input value. The bias is found to be larger than the estimated noise (illustrated by the error bars). Crucially however, the bias is lower than 0.5 magnitude, which is within our target accuracy (i.e. 1.25 magnitude) to ensure a useful measurement of the $\mathcal{M}_{\mathrm{BH}}-L$ correlation. (0.5 and 1.25 magnitude area are filled with two colors in the figures.)

It is interesting to note that the bias does not depend strongly on the effective radius of the simulated host galaxy, though lower surface brightness galaxies are more difficult to detect, and a smaller fraction of the entire galaxy is actually reconstructed. This means the mask region we selected is large enough to enclose a sufficient portion of the light. Moreover, the inferred Sérsic index n tends to be smaller than the input value. This bias is likely due to regularization

Table 3. Combination of parameters for each configuration

Serial	n	S_{re}	l_{eta}	Serial	n	S_{re}	l_{eta}
$No.1:$	2	0.39	0.8	$No.2:$	4	0.39	0.8
$No.3:$	2	0.78	0.8	$No.4:$	4	0.78	0.8
$No.5:$	2	0.39	1.0	$No.6:$	4	0.39	1.0
$No.7:$	2	0.78	1.0	$No.8:$	4	0.78	1.0
$No.9:$	2	0.39	1.2	$No.10:$	4	0.39	1.2
$No.11:$	2	0.78	1.2	$No.12:$	4	0.78	1.2

Note:− List of parameter sets for each configuration. We use WFC3-IR images as an example. For WFC3-UVIS and ACS images the corresponding values for S_{re} are slightly different and given in Table 2.

that tends to suppress sharp features and therefore GLEE prefers a flatter profile, whenever possible.

As we will discuss later, HE1104 is the hardest system to model among the systems with WFC3-IR imaging, owing to the fact that the lens configuration is a double - and thus we have considerably less information than in quads - and that the AGN images outshine the host galaxy by $\sim 2-4$ magnitudes. In spite of all these challenges, the simulations here show us the residual uncertainty is within 0.5 magnitudes (<0.2 dex), whether fixing n or not, which is much smaller than the uncertainty on $\mathcal{M}_{BH}$ and thus acceptable for our purposes.

4.2.2 HE0435

The results for HE0435 are shown in Fig. 6, using the same format as in the previous section. We note that the bias is reduced relative to HE1104, for two reasons: i) the system has four lensed images and there is thus more information available to reconstruct the host galaxy surface brightness; ii) the contrast between host galaxy magnitude and AGN magnitude is reduced to $\sim -1-2$ magnitudes.

4.2.3 WFI2033

Fig. 7 summarizes the results for WFI2033. The host galaxy magnitude can be reliably inferred with a much smaller bias than our target accuracy.

4.2.4 SDSS1206

SDSS1206 is a particularly interesting system for the study of the lensed host. While most of the host galaxy is quadruply imaged, the AGN falls just outside the inner caustic and is thus only doubly imaged (Agnello et al. 2016). As expected, the accuracy of the inferred magnitudes and effective radius are similar to that obtained for quadruply imaged AGNs and the bias is smaller than our target accuracy. The results are summarized in Fig. 8.

4.2.5 SDSS0246

The system SDSS0246 was imaged with WFC3-UVIS, and it has a much fainter arc compared with other systems. Even though the pixels are smaller than with WFC3-IR and the resolution is higher, the available region for modelling the arc in the mask is quite limited in angular size. In spite of this serious challenge, GLEE successfully reconstructed the host

image within our target accuracy, at least for the brighter range of source magnitudes considered here. The final results are summarized in Fig. 9.

Not surprisingly, the bias is larger than for other systems. As we can see, for brighter host set (21.5, 22.0 and part of 22.5 mag), the inferred magnitudes are accurate within 1 mag, i.e. within our target of 1.25 mags.

For the fainter host configuration (e.g. 22.5 and 23 Mag with large S_{re}), the inferred magnitudes are much smaller than the input. The reasons can be easily understood in the following way. The reconstructions for these images are too faint with an irregular light distribution. In this case, GALFIT tries to fit an unrealistically large S_{re} (see Fig. 9-(b)) which corresponds to overestimating the brightness of the source. Indeed, this kind of modelling bias can be offset when the S_{re} are fixed to the truth or limited within $1''$. However, the final accuracy of our magnitude inference will be driven by our knowledge of the prior on S_{re}. Our conclusion is that if the host galaxy is inferred to be as faint or fainter than 22.5 in the final analysis, the inference cannot be considered reliable given the current data quality.

4.2.6 HS2209

It is worth noting that HS2209 is the most challenging system in our sample with the AGN brightness approaching 15 mag, in the image plane. The results of HS2209 is summarized in Fig. 10 with generally larger bias than SDSS0246. Nevertheless, in most cases, the bias levels are within our target.

4.2.7 HE0047

The results for HE0047 are summarized in Fig. 11. They are similar to those obtained for the other doubles SDSS0246 and HS2209, and show a larger offset between input and output model parameters than is seen in the quads. This is easily understood because the quads have more information and in general larger magnifications.

4.2.8 RXJ1131

RXJ1131 was observed with ACS through the F814W filter. A detailed lens model with GLEE has already been published by Suyu et al. (2013, 2014). The host galaxy is much brighter than in the other cases (see Table 2). Thus, the results of RXJ1131 have the smallest bias in the entire sample, as summarized in Fig. 12.

5 DISCUSSION

Having analyzed each individual system we can now study the trends across the sample to identify general lessons. We expect the magnitude bias to depend on the total brightness of the lensed arc in image plane, the contrast between the lensed arc and the AGN, and possibly the effective surface brightness of the host galaxy. The latter quantity is defined as the average surface brightness inside the effective radius in the source plane, which should be approximately independent of lensing magnification.

This intuition is supported by our results across the
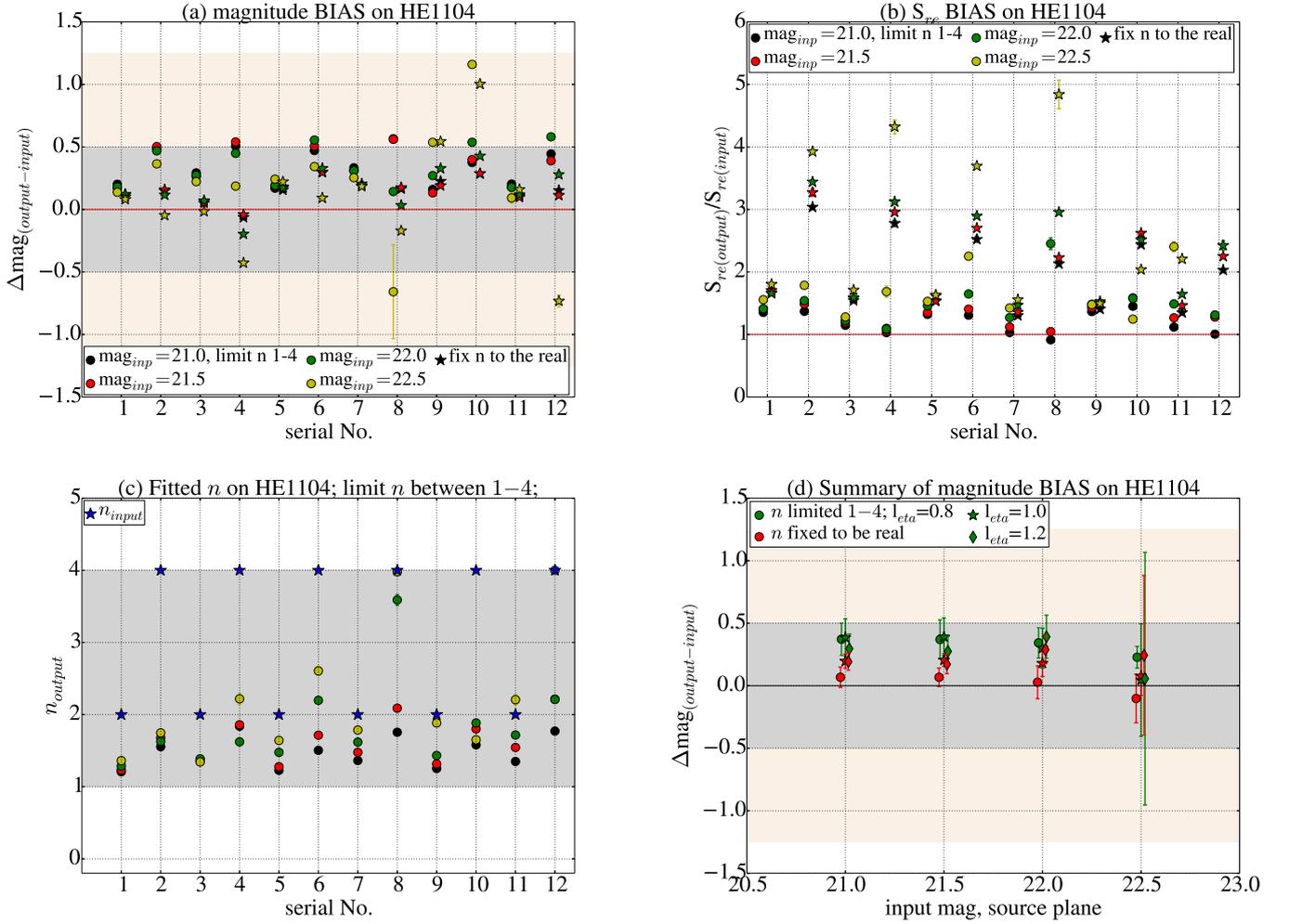
Figure 5. Summary of the difference between input (true) and inferred model parameters describing the lensed host galaxy light profile for HE1104. For panels (a), (b) and (c), the serial numbers correspond to different parameter configurations, as listed in Table 3. The y axis shows the modelling results obtained with GALFIT. Different colors represent different input magnitudes. The filled circles are the results when allowing n to float in the range $1 - 4$; filled stars are obtained by fixing n to the input value. For panel (d), different colors represent different priors on n; different shapes correspond to different l_{eta} inputs.

sample, as illustrated in Fig. 13 in which we relate the magnitude bias level to the brightness of the host arc. (The total magnitude of the host galaxy in the image plane is calculated by re-simulating the images without the AGN and deflector light (i.e. right panels in Fig. 3).) Indeed, the bias gets larger than our target as magnitudes get fainter than ~ 20, and it is well within 0.5 mags for arc magnitudes brighter than $\sim$19. It is also clear that the quads yield better results than the doubles, as expected. By comparing the six panels, it appears that magnitude of the host galaxy and contrast between host and AGN are more important than effective surface brightness of the host in driving the precision of the recovery of the total magnitude. We note that except for RXJ1131 the exposure times are comparable for each system so that differences across the sample are not driven by varying depth of the data, but rather by the different properties of each system.

Besides the intrinsic bias from the source reconstruction and fitting, additional bias can also be potentially in-

troduced by the following factors. First, the mask regions for modelling the arc are selected manually. Of course, a different choice of mask region will produce a different reconstructed image. However, we find that this difference is negligible once the mask region is sufficiently large. Second, the noise level in central AGN area was boosted to account for PSF mismatch. This in general leads to underestimating the Sérsic index n. In practice, we find that this bias is not sufficient to affect the inferred magnitudes beyond our target precisions, provided that n is allowed to be within a physically plausible range.

We note that in this work we have assumed that the PSF is known. The analysis of the real systems to be presented in future work will have to take into account additional uncertainties related on the host galaxy magnitude arising from residual uncertainties in the PSF reconstruction. This can be achieved by marginalizing over the distribution of acceptable PSF reconstructions.

It is difficult to compare with Peng et al. (2006b) given
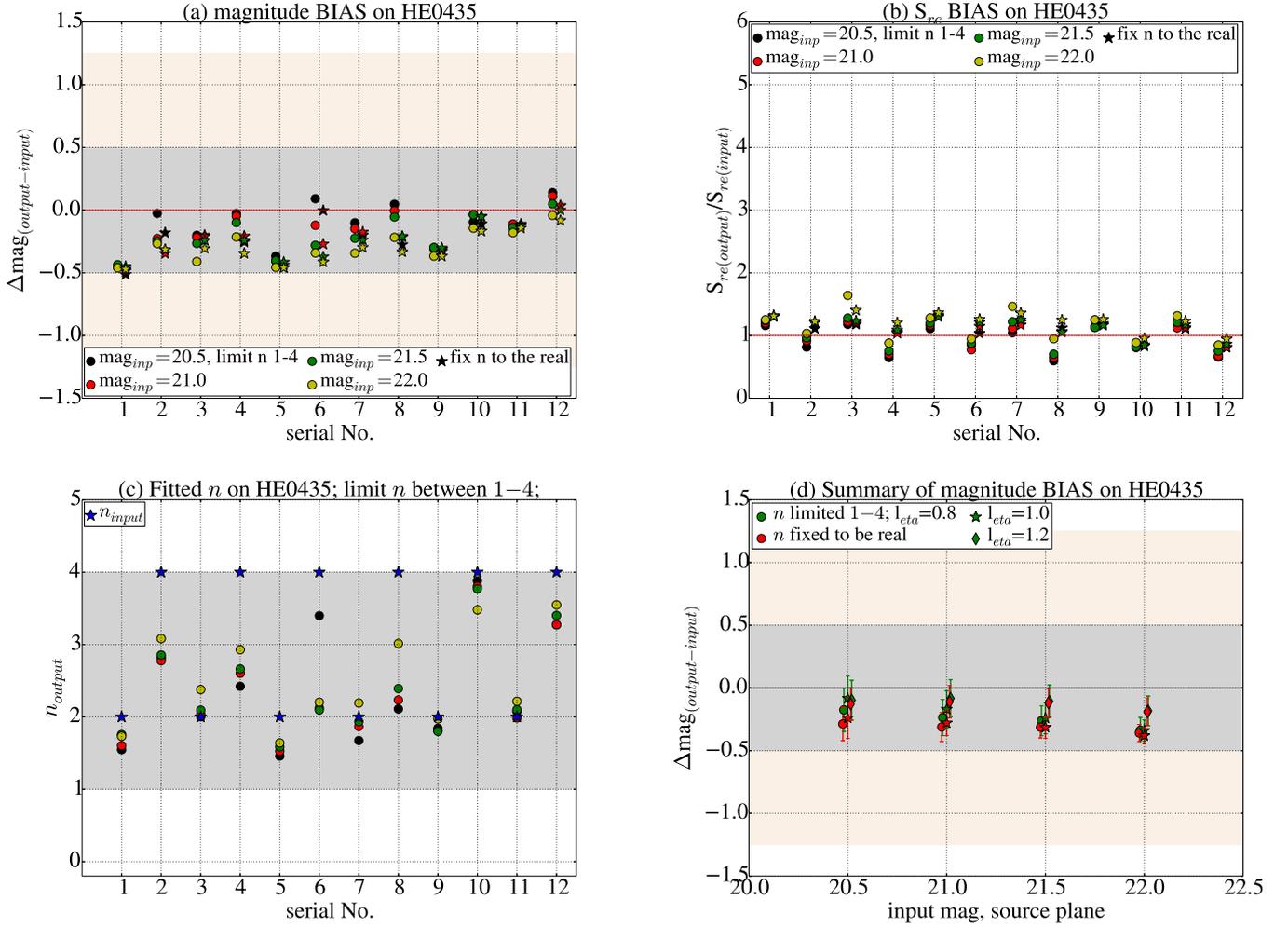
Figure 6. Results for HE0435. See caption of Fig. 5 for details.

the very different data quality and inference methods. Three systems (HE0435, HE1104, RXJ1131) are in common with their study and our simulations suggest that indeed systematic errors should be smaller than their reported 0.3 magnitudes uncertainty given our data quality. Moreover, we find that their measurements agree with our estimated range of magnitudes. The host magnitudes in the source plane for HE1104 and HE0435 in Peng et al. (2006b) are measured to be 21.4 and 21.3 mag respectively (converting from Vega to AB), which are both within our estimated range $21 \sim 22$ mag, taking into account our approximate magnification[8] (see Table 2). The measured brightness of the AGN in the source plane in their paper are 19.5 and 20.9 mag respectively while our simulations range between $\sim$19.5 and 22.25. The comparison for RXJ1131 is not straightforward because the analysis is carried out in two different filters.

In this paper we have focused on the accuracy of the

reconstruction of the lensed host galaxy properties. A topic of great interest, but beyond the scope of this paper, is how well the parameters of the deflector are recovered. As shown by our analysis of HE0435 (Wong et al. 2016) and RXJ1131 (Suyu et al. 2013), the mass model is highly constrained by the high signal to noise ratio data available, especially when stellar velocity dispersion, time delays, and other ancillary data are included (Dobke & King 2006). Thus residual uncertainties in the lens model are unlikely to induce major uncertainties in the lensed galaxy brightness. However, a full assessment of this effect requires more extensive tests and simulations and this is left for future work, when a blind lens modelling challenge will be presented.

Finally, we conclude by noting that the selection function of our sample of lensed quasars is non-trivial in terms of black hole mass and host galaxy properties. Following the practice established for non-lensed samples (e.g., Treu et al. 2007; Lauer et al. 2007; Park et al. 2015), it will have to be modelled in order to reach general conclusions about the overall population of AGN hosts from this sample.

---

[8] The approximate total magnification in magnitudes (see Table 1) is estimated by calculating the magnification of the point source for an isothermal mass model. For SDSS1206 the magnification of the host galaxy is likely to be significantly larger, since it is quadruply imaged.
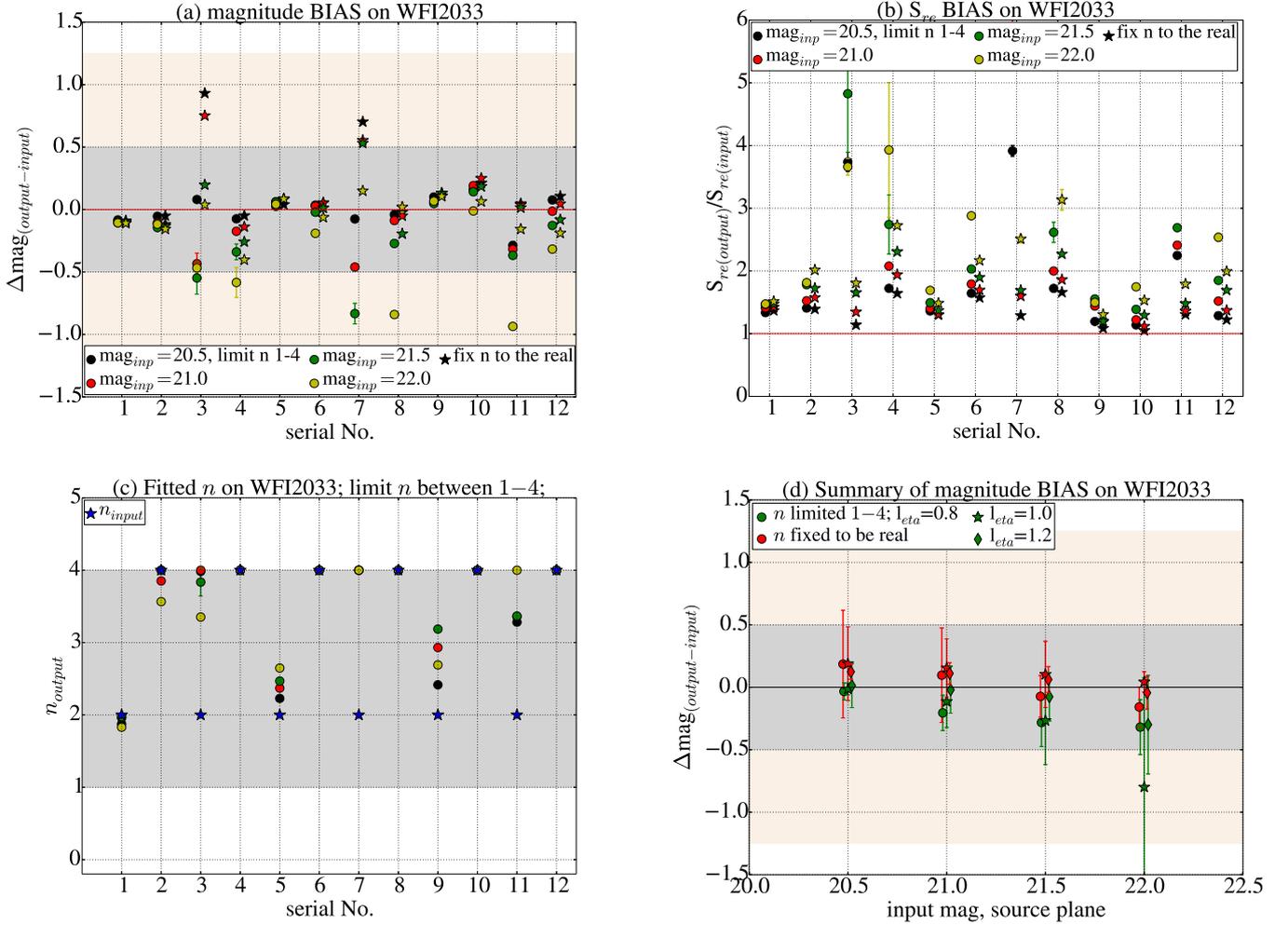
Figure 7. Results for WFI2033. See caption of Fig. 5 for details.

6 SUMMARY

We carried out extensive and realistic simulations of lensed AGN images taken with the *Hubble Space Telescope*, in order to study the fidelity of AGN host galaxy reconstruction using state of the art data and lens reconstruction software. We simulated eight lensed AGN systems (see Table 1 for detailed information), with 48 sets of parameters each covering the plausible physical range for the host galaxy and lens galaxy properties. Reproducing as closely as possible the analysis procedure of real data, we inferred the magnitude of the AGN host galaxy and compared it to the input value. Our main conclusions are:

(i) Systematic errors in the reconstructed host galaxy luminosity are typically larger than random errors. However, the level of systematic errors and biases are smaller than the uncertainty of single epoch $\mathcal{M}_{\rm BH}$ estimates (~ 0.5 dex, i.e. 1.25 mag). In fact, for brighter host arc images ($F160W \lesssim 19-20$ in the image plane) the level of bias and systematic error is significantly smaller (within 0.5 mag).

(ii) The magnitude bias depends on both the magnitude of the host arc, and the contrast between the host arc and the AGN, as expected.

(iii) The magnitude bias does not depend on the Sérsic index of the host; there is no significant difference between different Sérsic index (i.e. 2 and 4) inputs, in the range that may be expected for bright AGN hosts.

(iv) The quadruply imaged systems tend to have smaller bias than the doubles, consistent with their larger magnification and larger amount of available information.

(v) For the fainter host galaxies ($F160W \gtrsim 20$ mag in the image plane) systematic errors and bias can be larger than 1.25 magnitudes. Such cases may be difficult to analyze without introducing informative priors to constrain the shape and size of the surface brightness profile.

We conclude that by using state of the art data and analysis tools the magnitudes of lensed AGN host galaxies can be studied with uncertainties that are smaller than those stemming from single epoch black hole mass estimates. We will present the actual measurements in forthcoming papers.
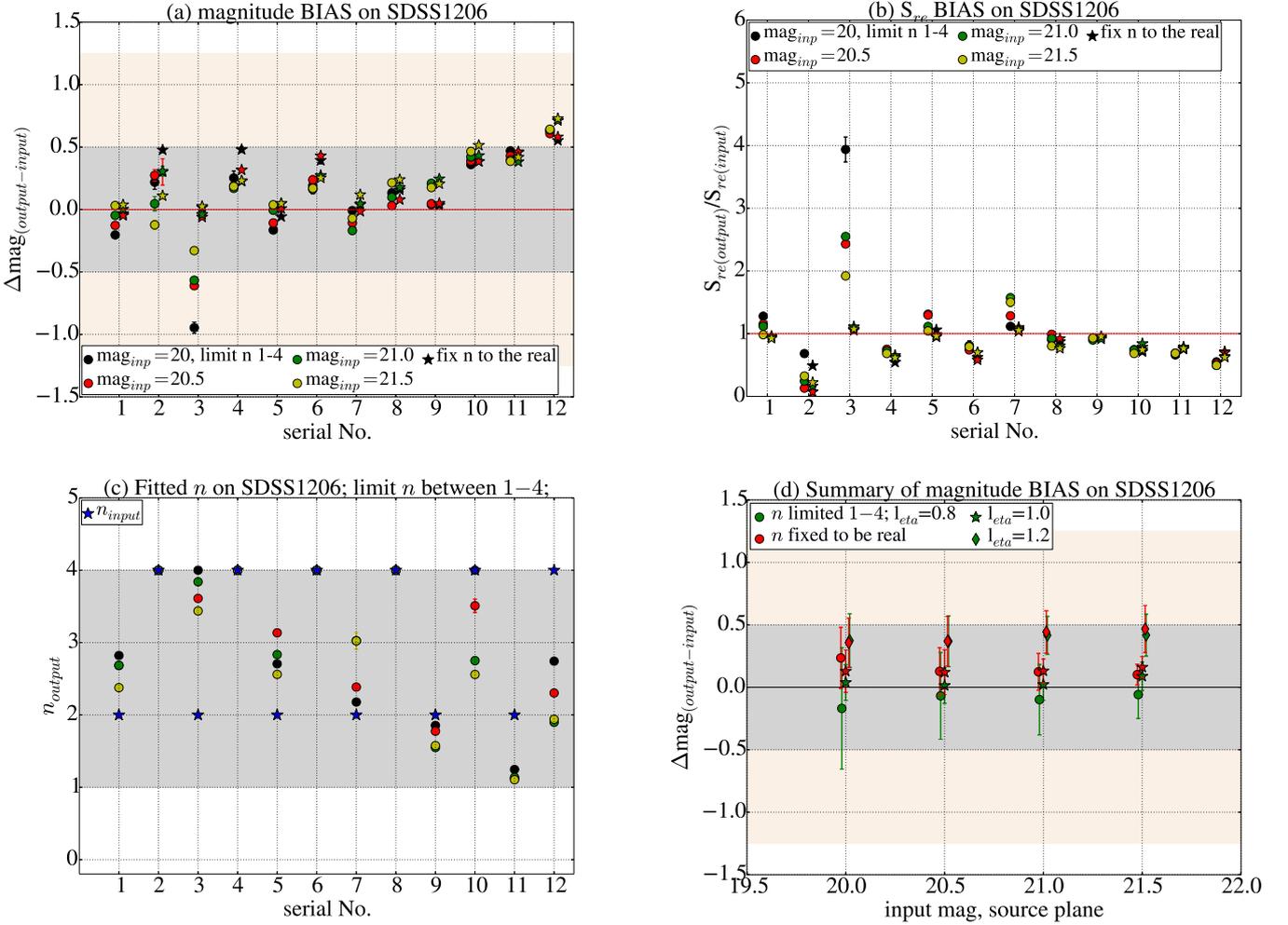
Figure 8. Results for SDSSJ1206. See caption of Fig. 5 for details.

REFERENCES

Agnello A., Sonnenfeld A., Suyu S. H., Treu T., Fassnacht C. D., Mason C., Bradač M., Auger M. W., 2016, MNRAS, 458, 3830
Bartelmann M., 2010, Classical and Quantum Gravity, 27, 233001
Beifiori A., Courteau S., Corsini E. M., Zhu Y., 2012, MNRAS, 419, 2497
Bennert V. N., Auger M. W., Treu T., Woo J.-H., Malkan M. A., 2011, ApJ, 726, 59
Bonvin V., et al., 2016, preprint, (arXiv:1607.01790)
Chantry V., Sluse D., Magain P., 2010, A&A, 522, A95
Chen G. C. F., et al., 2016, preprint, (arXiv:1601.01321)
Ciotti L., Bertin G., 1999, A&A, 352, 447
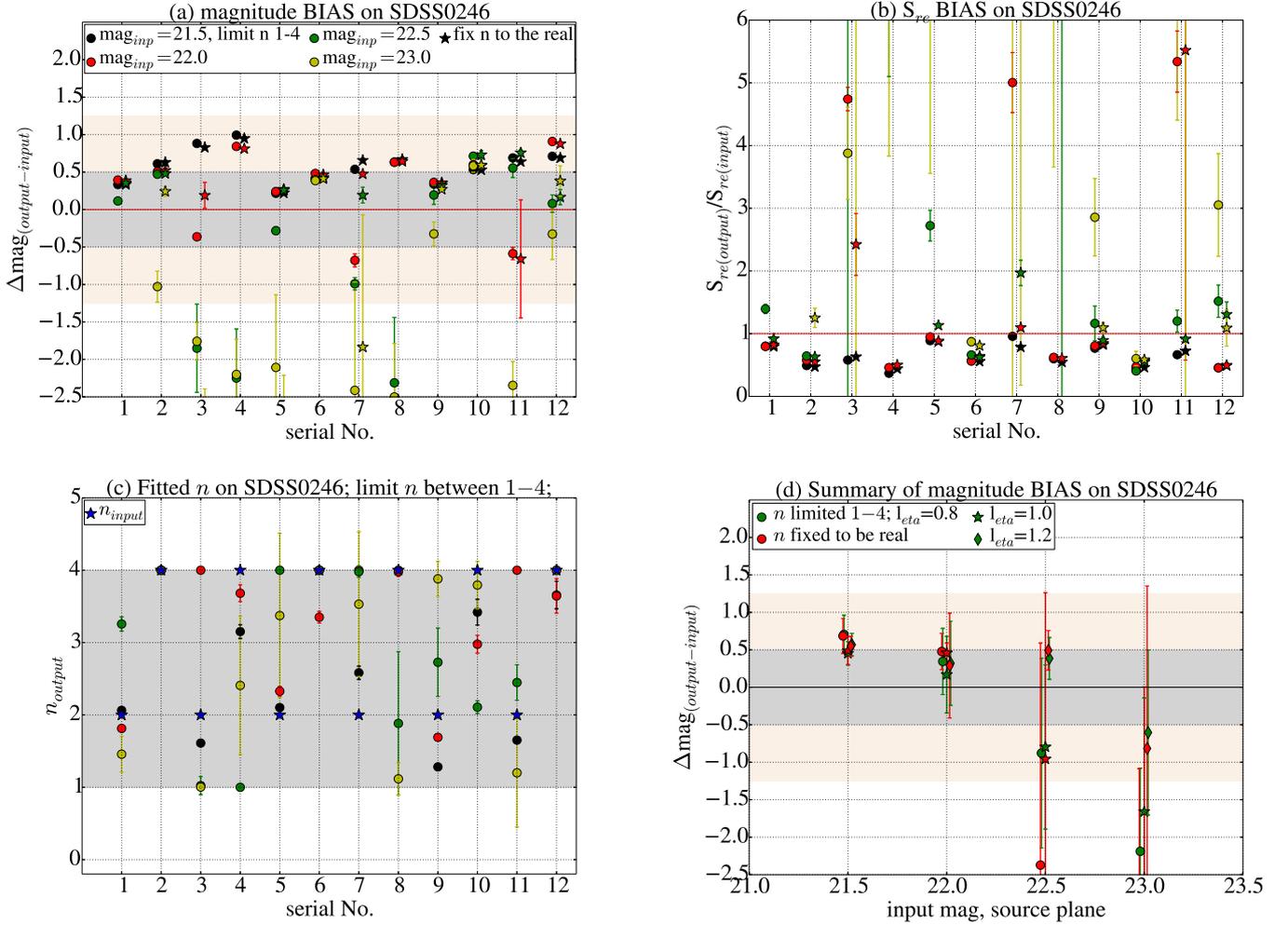
Figure 9. Results for SDSSJ0246. See caption of Fig. 5 for details.

Courbin F., Saha P., Schechter P. L., 2002, in Courbin F., Minniti D., eds, Lecture Notes in Physics, Berlin Springer Verlag Vol. 608, Gravitational Lensing: An Astrophysical Tool. p. 1 (arXiv:astro-ph/0208043)

DeGraf C., Di Matteo T., Treu T., Feng Y., Woo J.-H., Park D., 2015, MNRAS, 454, 913

Dobke B. M., King L. J., 2006, A&A, 460, 647

Eigenbrod A., Courbin F., Vuissoz C., Meylan G., Saha P., Dye S., 2005, A&A, 436, 25

Eigenbrod A., Courbin F., Meylan G., 2007, A&A, 465, 51

Eulaers E., et al., 2013, A&A, 553, A121

Fassnacht C. D., Womble D. S., Neugebauer G., Browne I. W. A., Readhead A. C. S., Matthews K., Pearson T. J., 1996, ApJ, 460, L103+

Ferrarese L., Merritt D., 2000, ApJ, 539, L9

Gebhardt K., et al., 2001, ApJ, 555, L75

Graham A. W., Onken C. A., Athanassoula E., Combes F., 2011, MNRAS, 412, 2211

Gültekin K., et al., 2009, ApJ, 698, 198

Hagen H.-J., Engels D., Reimers D., 1999, A&AS, 134, 483

Halkola A., Hildebrandt H., Schrabback T., Lombardi M., Bradač M., Erben T., Schneider P., Wuttke D., 2008, A&A, 481, 65

Häring N., Rix H.-W., 2004, ApJ, 604, L89

Inada N., et al., 2005, AJ, 130, 1967

Jahnke K., et al., 2009, preprint, (arXiv:0907.5199)

Kormendy J., Ho L. C., 2013, eprint arXiv, 1304, 7762

Lauer T. R., Tremaine S., Richstone D., Faber S. M., 2007, ApJ, 670, 249

Magorrian J., et al., 1998, AJ, 115, 2285

Marconi A., Hunt L. K., 2003, ApJ, 589, L21

Marshall P. J., et al., 2007, ApJ, 671, 1196

Meng X.-L., Treu T., Agnello A., Auger M. W., Liao K., Marshall P. J., 2015, J. Cosmology Astropart. Phys., 9, 059

Morgan N. D., Caldwell J. A. R., Schechter P. L., Dressler A., Egami E., Rix H.-W., 2004, AJ, 127, 2617

Nightingale J. W., Dye S., 2015, MNRAS, 452, 2940

Ofek E. O., Maoz D., Rix H.-W., Kochanek C. S., Falco E. E., 2006, ApJ, 641, 70

Oguri M., et al., 2005, ApJ, 622, 106

Park D., Woo J.-H., Bennert V. N., Treu T., Auger M. W., Malkan M. A., 2015, ApJ, 799, 164

Peng C. Y., 2007, The Astrophysical Journal, 671, 1098

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Peng C. Y., Impey C. D., Ho L. C., Barton E. J., Rix H.-W., 2006a, ApJ, 640, 114

Peng C. Y., Impey C. D., Rix H.-W., Kochanek C. S., Keeton C. R., Falco E. E., Lehár J., McLeod B. A., 2006b, ApJ, 649, 616

Peterson B. M., 2014, Space Sci. Rev., 183, 253
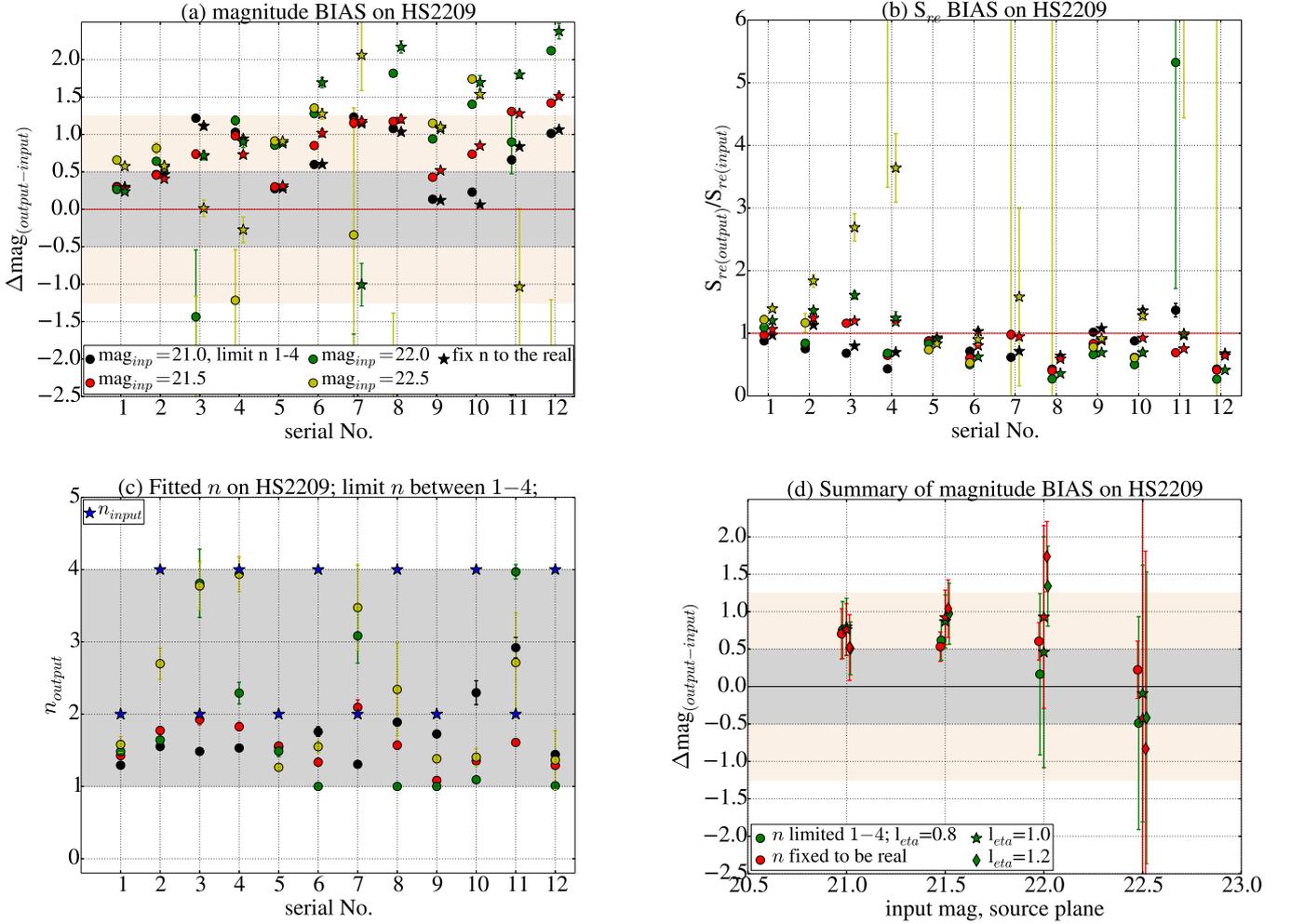
**Figure 10.** Results for HS2209. See caption of Fig. 5 for details.

Quinn J., et al., 2016, MNRAS, 459, 2394

Rusu C. E., et al., 2016a, preprint, (arXiv:1607.01047)

Rusu C. E., et al., 2016b, MNRAS, 458, 2

Salviander S., Shields G. A., Gebhardt K., Bonning E. W., 2006, New Astronomy Review, 50, 803

Schneider P., 2006, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C. S., Wambsganss J., eds, Saas-Fee Advanced Course 33: Gravitational Lensing: Strong, Weak and Micro. pp 1–89

Schneider P., Kochanek C. S., Wambsganss J., 2006, Gravitational Lensing: Strong, Weak and Micro, doi:10.1007/978-3-540-30310-7.

Sersic J. L., 1968, Atlas de galaxias australes. Cordoba, Argentina: Observatorio Astronomico

Shen Y., 2013, eprint arXiv, 1302, 2643

Sluse D., et al., 2003, A&A, 406, L43

Sluse D., Claeskens J.-F., Hutsemékers D., Surdej J., 2007, A&A, 468, 885

Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, A&A, 544, A62

Sluse D., et al., 2016, preprint, (arXiv:1607.00382)

Smette A., Robertson J. G., Shaver P. A., Reimers D., Wisotzki L., Koehler T., 1995, A&AS, 113, 199

Suyu S. H., 2012, MNRAS, 426, 868

Suyu S. H., Halkola A., 2010, A&A, 524, A94

Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983

Suyu S. H., et al., 2013, ApJ, 766, 70

Suyu S. H., et al., 2014, ApJ, 788, L35

Suyu S. H., et al., 2016, preprint, (arXiv:1607.00017)

Tamura Y., Oguri M., Iono D., Hatsukade B., Matsuda Y., Hayashi M., 2015, PASJ, 67, 72

Treu T., 2010, ARA&A, 48, 87

Treu T., Koopmans L. V. E., 2002, ApJ, 575, 87

Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739

Treu T., Malkan M. A., Blandford R. D., 2004, ApJ, 615, L97

Treu T., Woo J.-H., Malkan M. A., Blandford R. D., 2007, ApJ, 667, 117

Vegetti S., Koopmans L. V. E., 2009, MNRAS, 392, 945

Vestergaard M., Peterson B. M., 2006, ApJ, 641, 689

Walter F., Carilli C., Bertoldi F., Menten K., Cox P., Lo K. Y., Fan X., Strauss M. A., 2004, The Astrophysical Journal, 615, L17

Winn J. N., Rusin D., Kochanek C. S., 2004, Nature, 427, 613

Wisotzki L., Koehler T., Kayser R., Reimers D., 1993, A&A, 278, L15

Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, A&A, 395, 17

Wisotzki L., Schechter P. L., Chen H.-W., Richstone D., Jahnke K., Sánchez S. F., Reimers D., 2004, A&A, 419, L31
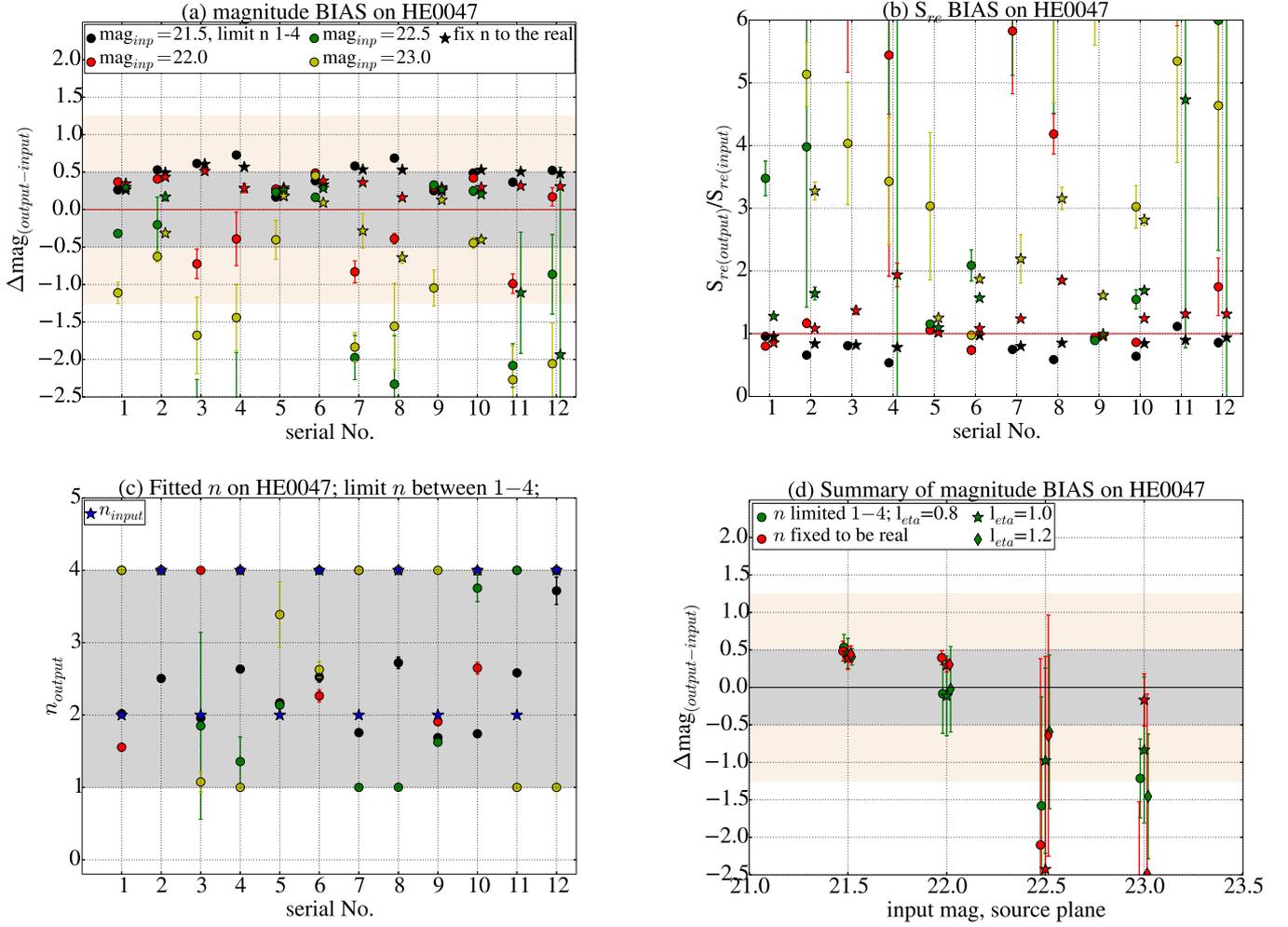
Figure 11. Results for HE0047. See caption of Fig. 5 for details.

Wong K. C., Suyu S. H., Matsushita S., 2015, ApJ, 811, 115

Wong K. C., et al., 2016, preprint, (arXiv:1607.01403)

Woo J., Treu T., Malkan M. A., Blandford R. D., 2006, ApJ, 645, 900

de Vaucouleurs G., 1948, Annales d'Astrophysique, 11, 247

This paper has been typeset from a TEX/LATEX file prepared by the author.
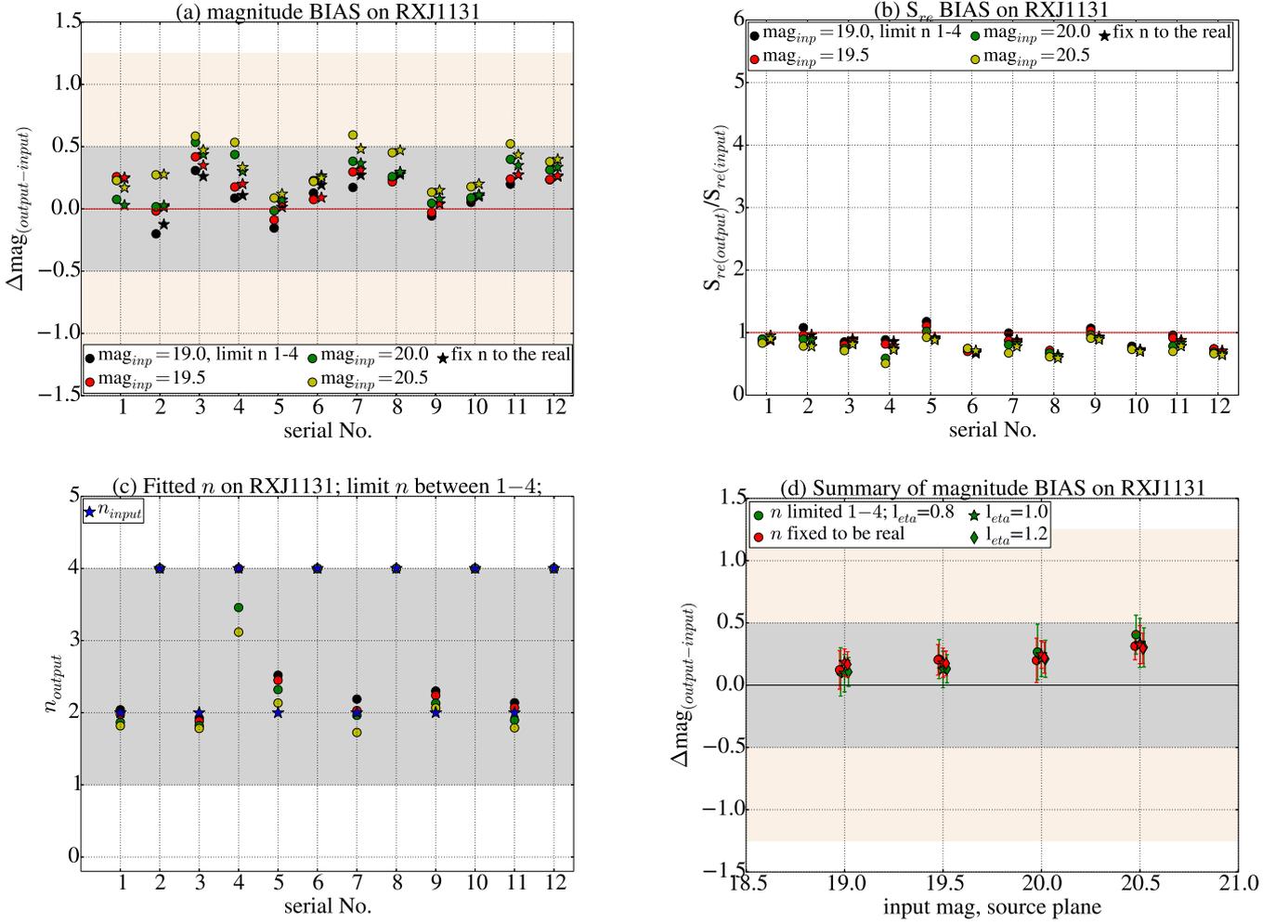
**Figure 12.** Results for RXJ1131. See caption of Fig. 5 for details.
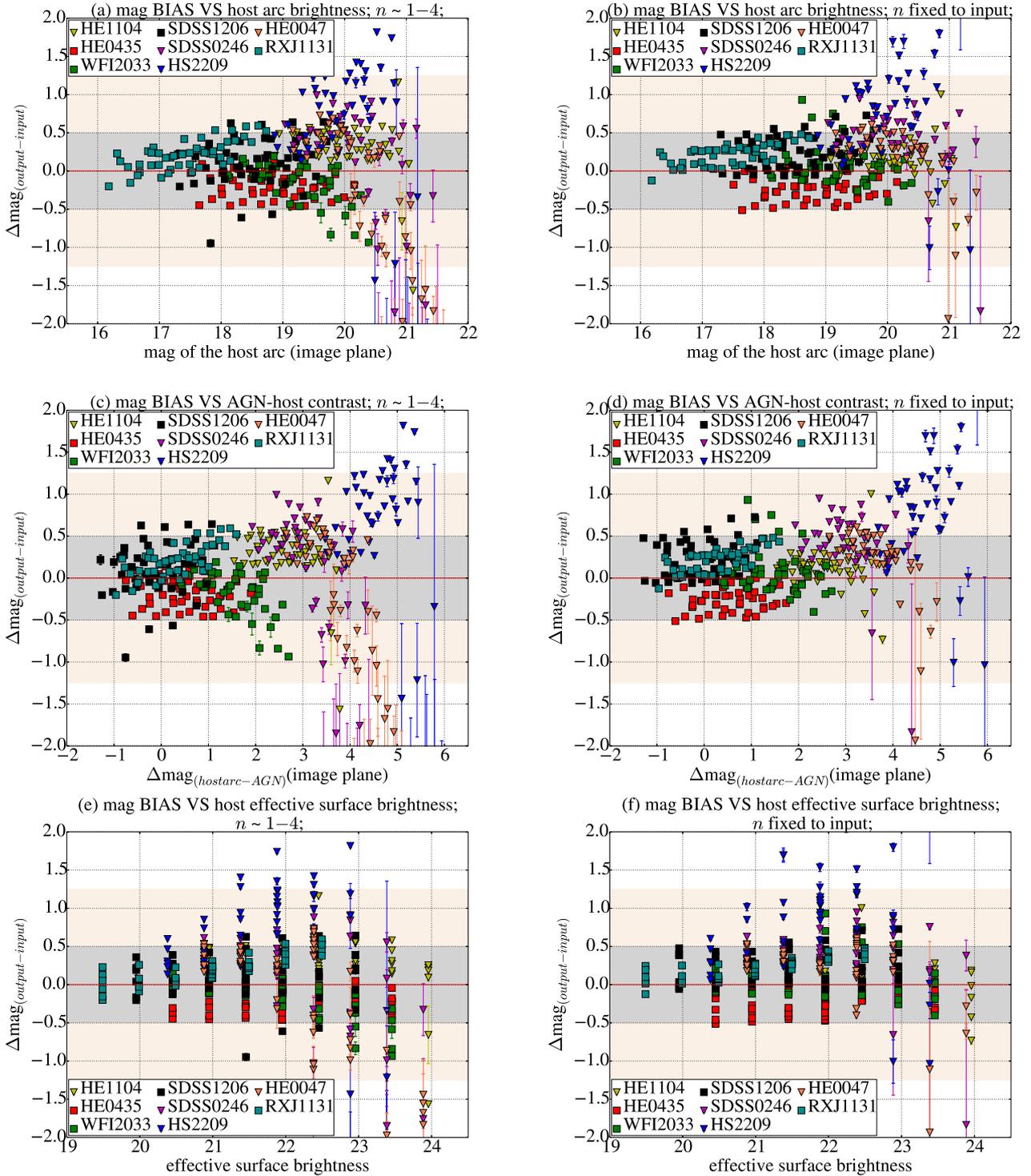
Figure 13. Magnitude bias across the sample as a function of host arc magnitude in image plane (top row) and as a function of contrast between the AGN and the host arc brightness (middle row), and effective surface brigthenss (bottom row). Different colors represent different systems in the sample. The square symbols represent quadruply imaged sources while the triangles represent doubly-imaged system. In the left panels n is left free within the range $1-4$ during the fit; in the right panels it is kept fixed to the input value.