






H0LiCOW XII. Lens mass model of WFI2033 – 4723 and blind measurement of its time-delay distance and H_0

Cristian E. Rusu ^{1,2,3}★†, Kenneth C. Wong,^{1,4} Vivien Bonvin,⁵ Dominique Sluse,⁶ Sherry H. Suyu,^{7,8,9} Christopher D. Fassnacht,³ James H. H. Chan ⁵, Stefan Hilbert,^{10,11} Matthew W. Auger,¹² Alessandro Sonnenfeld ^{4,13} Simon Birrer,¹⁴ Frederic Courbin,⁵ Tommaso Treu,¹⁴ Geoff C.-F. Chen ³, Aleksi Halkola,¹⁵ Léon V. E. Koopmans,¹⁶ Philip J. Marshall ¹⁷ and Anowar J. Shajib¹⁴

Affiliations are listed at the end of the paper

Accepted 2019 December 5. Received 2019 December 5; in original form 2019 May 22

ABSTRACT

We present the lens mass model of the quadruply-imaged gravitationally lensed quasar WFI2033 – 4723, and perform a blind cosmographical analysis based on this system. Our analysis combines (1) time-delay measurements from 14 yr of data obtained by the COSmological MONitoring of GRAvItational Lenses (COSMOGRAIL) collaboration, (2) high-resolution *Hubble Space Telescope* imaging, (3) a measurement of the velocity dispersion of the lens galaxy based on ESO-MUSE data, and (4) multi-band, wide-field imaging and spectroscopy characterizing the lens environment. We account for all known sources of systematics, including the influence of nearby perturbers and complex line-of-sight structure, as well as the parametrization of the light and mass profiles of the lensing galaxy. After unblinding, we determine the effective time-delay distance to be 4784^{+399}_{-248} Mpc, an average precision of 6.6 per cent. This translates to a Hubble constant $H_0 = 71.6^{+3.8}_{-4.9}$ km s⁻¹ Mpc⁻¹, assuming a flat Λ CDM cosmology with a uniform prior on Ω_m in the range [0.05, 0.5]. This work is part of the H_0 Lenses in COSMOGRAIL’s Wellspring (H0LiCOW) collaboration, and the full time-delay cosmography results from a total of six strongly lensed systems are presented in a companion paper (H0LiCOW XIII).

Key words: gravitational lensing; strong – cosmological parameters – distance scale.

1 INTRODUCTION

The flat Λ CDM cosmological model, characterized by spatial flatness, dark energy in the form of a cosmological constant, and cold dark matter, is considered to be the standard cosmological model today. Although this model is known as the concordance model, sources of tension have none-the-less begun to appear as the results of different cosmological experiments have grown in precision. Most notably, the tension between the measurements of the Hubble constant from the analysis of the cosmic microwave background (CMB) by the *Planck* mission (under the strict assumption of flat Λ CDM) and of Type Ia supernovae standard candles calibrated using the local distance ladder by the Supernovae, H_0 , for the Equation of State of Dark Energy collaboration (SH0ES; Riess et al. 2016) has recently increased from 3.4σ (Planck Collaboration XIII 2015; Riess et al. 2016) to 4.4σ (Planck Collaboration VI 2018; Riess et al. 2019). The latest results are $H_0 = 67.4 \pm 0.5$ km s⁻¹ Mpc⁻¹ from *Planck*, and $H_0 = 74.03 \pm 1.42$ km s⁻¹ Mpc⁻¹ from SH0ES.

At present, sources of systematic error in either of these measurements that are significant enough to explain the discrepancy

have not been demonstrated to exist. This opens up the intriguing possibility of having to extend the standard cosmological model by allowing for curvature, more general dark energy, or increasing the number of neutrinos (see fig. 13 in Riess et al. 2019), or to consider exotic alternatives, such as a vacuum phase transition (Di Valentino, Linder & Melchiorri 2018), early dark energy models (Poulin et al. 2018), self-interacting neutrinos (Kreisch, Cyr-Racine & Doré 2019) or decaying dark matter (Vattis, Koushiappas & Loeb 2019). The various parameters of such extensions are highly degenerate with the value of H_0 , and therefore a high-precision determination, with a technique independent of, and therefore not subject to the same systematics of either *Planck* or SH0ES, is in demand (e.g. Hu 2005; Suyu et al. 2012a; Weinberg et al. 2013). Some proposed independent methods, such as water masers (e.g. Gao et al. 2016; Braatz et al. 2018), extragalactic background light attenuation (e.g. Domínguez et al. 2019), and gravitational waves (e.g. Feeney et al. 2019), etc. have yet to resolve the H_0 discrepancy, as their precision is not yet comparable to *Planck* or SH0ES.

The time-delay cosmography technique uses gravitational lens time delays to measure H_0 . This technique rests on the fact that light rays from a multiply-imaged source will take different paths as they propagate through space–time, with different geometrical lengths and gravitational potential depths. This will introduce an offset in

* E-mail: cerusu@naoj.org

† Subaru Fellow.

arrival times, which can be measured through monitoring, if the source brightness varies in time. The measured time delays are used to infer the ‘time-delay distance’, primarily sensitive to H_0 , which therefore provides a one-step way of measuring H_0 (e.g. Vanderriest et al. 1989; Keeton & Kochanek 1997; Oguri 2007; Suyu et al. 2010). Although proposed more than half a century ago by Refsdal (1964) in the context of lensed supernovae, the original idea has only recently been implemented (Grillo et al. 2018). Far more common is the use of gravitationally lensed quasars, given the sample of 250 such systems known to date (e.g. Lemon, Auger & McMahon 2019).

In practice, an accurate measurement of H_0 through this method requires extensive observational data for each system, as well as the development of advanced modelling techniques (Suyu & Halkola 2010; Suyu et al. 2012b; Tewes, Courbin & Meylan 2013; Birrer, Amara & Refregier 2015; Bonvin et al. 2016; Birrer & Amara 2018), and has only become feasible in the current decade. Our collaboration, H_0 Lenses in COSMOGRAIL’s Wellspring (HOLiCOW; Suyu et al. 2017, hereafter HOLiCOW I) is designed to perform such measurements. We have precise, long-term time-delay measurements from the COSmological MONitoring of GRAVItational Lenses (COSMOGRAIL; Courbin et al. 2005; Eigenbrod et al. 2005; Bonvin et al. 2018) project. We use deep high-resolution imaging from the *Hubble Space Telescope* (HST) or adaptive optics that provide constraints on the lens model not only from the point-source positions, but also from the extended arcs of the lensed quasar host galaxy. Finally, we have velocity dispersion measurements of the lens galaxies and characterizations of their environments and line of sight (LOS; e.g. Collett et al. 2013; Greene et al. 2013; McCully et al. 2014, 2017; Tihhonova et al. 2018), in order to reduce the mass-sheet degeneracy (e.g. Falco, Gorenstein & Shapiro 1985; Schneider & Sluse 2013).

With four lenses, we measured $H_0 = 72.5^{+2.1}_{-2.3}$ km s⁻¹ Mpc⁻¹ with a precision of 3.0 percent (Birrer et al. 2019, hereafter HOLiCOW IX) including systematic uncertainties, achieving our previous goal of the program of reaching < 3.5 per cent precision from the five separate lenses in the base HOLiCOW sample (see HOLiCOW I) and finding good agreement with SH0ES. We have thus shown that we are on track to measure H_0 with a precision of 1 percent from a future sample of ~40 lenses with comparable precision per system (e.g. Treu & Marshall 2016; Shajib, Treu & Agnello 2018), a result which will have significant implications for understanding the current tension with the CMB value. Time-delay cosmography is therefore a very effective technique, in the sense that only a relatively small number of systems is required to achieve a tight precision. The efficiency is similar to that expected for gravitational wave detections with optical counterparts (Chen, Fishbach & Holz 2018b). As we work towards the 1 percent precision goal from a sample of lenses, it is important to keep systematics in the inference of H_0 from individual systems within the 1 percent threshold, in order to insure accuracy, and also to test for biases by using multiple codes (Birrer et al. 2019) and data challenges (Liao et al. 2015; Ding et al. 2018).

In this paper, we present the results of a detailed lens modelling analysis of the gravitational lens WFI2033 – 4723 (J2000: 20^h33^m41^s.9, –47°23′43″.4), a quadruply-lensed quasar discovered by Morgan et al. (2004). The source redshift is $z_s = 1.662$ (Sluse et al. 2012), and the main deflector is a massive elliptical galaxy at a redshift of $z_d = 0.6575 \pm 0.0002$ (Sluse et al. 2019, hereafter HOLiCOW X), updating the $z_d = 0.661 \pm 0.001$ measurement from Eigenbrod et al. (2006). Bonvin et al. (2019b) (hereafter COSMOGRAIL XVIII) measure the time delays between the quasar

images based on 14 yr of monitoring, and HOLiCOW X study the environment and LOS to the lens, based on multiband imaging and targeted spectroscopy. Our work supersedes the models presented in Vuissoz et al. (2008) (hereafter COSMOGRAIL VII), which are based on monitoring of shorter duration and constrained only by the positions of the quasar images.

This is the fifth HOLiCOW system analysed in this manner, following B1608 + 656 (Suyu et al. 2010), RXJ1131 – 1231 (Suyu et al. 2013, 2014), HE 0435 – 1223 (Wong et al. 2017, hereafter HOLiCOW IV), and SDSS 1206 + 4332 (HOLiCOW IX), with a sixth lens, PG 1115 + 080, analysed simultaneously (Chen et al. 2019). A HOLiCOW milestone paper (Wong et al. 2019) presents the results of a conjoined cosmographical analysis of these lenses.

This paper is organized as follows. We give a brief overview of using time-delay lenses for cosmography in Section 2. In Section 3, we describe the observational data used in our analysis. We describe our lens modelling procedure in Section 4. In Section 5 we quantify the effect of the lens environment in terms of an external convergence. The time-delay distance results and their implications for cosmology are presented in Section 6. We summarize our main conclusions in Section 7.

2 SUMMARY OF TIME-DELAY COSMOGRAPHY

2.1 Time-delay distance

When a source is gravitationally lensed by a foreground mass, the arrival time of photons travelling from the source to the observer depends on both the path-length and the gravitational potential traversed by the light rays. For a single lens plane, the excess time-delay of an image at a position $\theta = (\theta_1, \theta_2)$ on the sky with a corresponding source position $\beta = (\beta_1, \beta_2)$ relative to the case of no lensing is

$$t(\theta, \beta) = \frac{D_{\Delta t}}{c} \left[\frac{(\theta - \beta)^2}{2} - \psi(\theta) \right], \quad (1)$$

where $D_{\Delta t}$ is the time-delay distance and $\psi(\theta)$ is the lens potential. The time-delay distance $D_{\Delta t}$ (Refsdal 1964; Schneider, Ehlers & Falco 1992; Suyu et al. 2010) is defined¹ as

$$D_{\Delta t} \equiv (1 + z_d) \frac{D_d D_s}{D_{ds}}, \quad (2)$$

where z_d is the lens redshift, D_d , D_s , and D_{ds} are the angular diameter distances between the lens and the observer, the source and the observer, and the lens and the source, respectively. $D_{\Delta t}$ has units of distance and is inversely proportional to H_0 , with weak dependence on other cosmological parameters.

The time-delay between two images, i and j , of a lensed source is the difference of their excess time delays,

$$\Delta t_{ij} = \frac{D_{\Delta t}}{c} \left[\frac{(\theta_i - \beta)^2}{2} - \psi(\theta_i) - \frac{(\theta_j - \beta)^2}{2} + \psi(\theta_j) \right], \quad (3)$$

where θ_i and θ_j are the positions of images i and j , respectively, in the image plane. If the source is variable on short time-scales (on the order of weeks to months), it is possible to monitor the

¹For historical reasons, the time-delay distance is written in terms of angular diameter distances. A more natural definition is $D_{\Delta t} \equiv \hat{D}_d \hat{D}_s / \hat{D}_{ds}$ where \hat{D} are the proper distances that the photons have travelled.

lensed image fluxes at positions θ_i and θ_j and measure the time-delay, Δt_{ij} , between them (e.g. Vanderriest et al. 1989; Schechter et al. 1997; Fassnacht et al. 1999, 2002; Kochanek et al. 2006; Courbin et al. 2011). The lens potentials at the image positions, $\psi(\theta_i)$ and $\psi(\theta_j)$, as well as at the source position, β , can be determined by modelling the system. In this way, lenses with measured time delays and accurate lens models can constrain $D_{\Delta t}$, and in turn, H_0 .

If there are multiple deflectors at different redshifts, the observed time delays depend on combinations of the angular diameter distances among the observer, the multiple deflectors, and the source. The observed image positions are determined by the multiplane lens equation (e.g. Blandford & Narayan 1986; Kovner 1987; Schneider et al. 1992; Petters, Levine & Wambsganss 2001; Collett & Auger 2014; McCully et al. 2014), but there is no longer a unique time-delay distance associated with the system. However, if the lensing is dominated by the mass in a single redshift plane, the observed time delays are mostly sensitive to the time-delay distance (equation 2), with the deflector redshift set to the redshift of the main lens plane. This approximation is valid for WFI2033 – 4723 (see Appendix B), and thus our results can be interpreted in terms of the ‘effective’ time-delay distance, $D_{\Delta t}(z_d, z_s)$. Hereafter, $D_{\Delta t}$ refers to the effective time-delay distance unless otherwise indicated.

A complicating factor in determining the time delay is the ‘microlensing time delay’, an effect first described by Tie & Kochanek (2018). Stars and compact objects in the lens galaxy can act as microlenses, which causes a differential magnification of the accretion disc of the lensed quasar. Since the microlensing effect is different at the positions of the various lensed images and varies over time as the microlenses move, this may create an additional bias and scatter in the measurement of the time delay between different images. The microlensing time delay depends on a number of assumptions about the accretion disc size, its orientation, and inclination, and the propagation of radiation through the disc. The effect tends to be small, of the order of \sim days or shorter, and can be modelled and accounted for under proper assumptions (Bonvin et al. 2018). This effect can also be mitigated by using the relative offsets between the measured time delays and those expected from lens modelling (Chen et al. 2018a).

Another difficulty is due to the fact that external perturbations from mass along the LOS can affect the lens potential that light rays pass through. These perturbations not only can affect the lens model of the system, but also lead to additional focusing and defocusing of the light rays, which also affect the measured time delays (e.g. Seljak 1994). If unaccounted for, these perturbers can lead to biased inferences of $D_{\Delta t}$. If the effects of LOS perturbers are small enough that higher order terms are unimportant (Keeton 2003; McCully et al. 2014), they can be approximated by an external convergence term in the lens plane. The true $D_{\Delta t}$ is related to the $D_{\Delta t}^{\text{model}}$ inferred from a mass model by

$$D_{\Delta t} = \frac{D_{\Delta t}^{\text{model}}}{1 - \kappa_{\text{ext}}} \rightarrow H_0 = (1 - \kappa_{\text{ext}})H_0^{\text{model}}. \quad (4)$$

Here, κ_{ext} cannot, in general, be constrained from the lens model due to the mass-sheet degeneracy (e.g. Falco et al. 1985; Gorenstein, Shapiro & Falco 1988; Saha 2000), in which the addition of a uniform mass sheet and a rescaling of the source plane coordinates can affect the inferred $D_{\Delta t}$ but leaves other observables unchanged.

This degeneracy can be substantially mitigated by estimating the mass distribution along the LOS (e.g. Fassnacht et al. 2006; Momcheva et al. 2006, 2015; Williams et al. 2006; Wong et al.

2011) and assuming that the physical mass of the deflector profile goes to zero at large radius. However, perturbers that are very massive or projected very close to the lens may need to be included explicitly in the mass model since their higher order effects need to be accounted for (McCully et al. 2017). In contrast, the lens profile is also degenerate with the time-delay distance in that the radial profile slope is tightly correlated with the time-delay distance (e.g. Kochanek 2002; Wucknitz 2002; Suyu 2012). This degeneracy can affect models with the same form of mass density profile (e.g. a power-law density profile), as well as models with different forms of density profiles (described analytically or not). Furthermore, this degeneracy can mimic the effects of the mass-sheet degeneracy because different profiles can approximate or exactly match mass-sheet transformations of one form or another (e.g. Schneider & Sluse 2013, 2014; Unruh, Schneider & Sluse 2017). These degeneracies can be reduced by combining the lensing data with stellar kinematics information (e.g. Treu & Koopmans 2002; Koopmans et al. 2003; Auger et al. 2010; Suyu et al. 2014; Yıldırım, Suyu & Halkola 2019), and by making reasonable assumptions about the mass profile. Including a velocity dispersion measurement in the modelling helps constrain any internal uniform mass component from a local galaxy group that the dynamics is sensitive to (Koopmans 2004).

2.2 Joint inference

Our inference of $D_{\Delta t}$ generally follows that of previous H0LiCOW analyses (Suyu et al. 2013, H0LiCOW IV, IX). Our observational data are denoted by \mathbf{d}_{HST} for the *HST* imaging data, $\Delta \mathbf{t}$ for the time delays, σ for the velocity dispersion of the lens galaxy, and \mathbf{d}_{LOS} for the LOS mass distribution determined from our photometric and spectroscopic data. We want to determine the posterior probability distribution function (PDF) of the model parameters ξ given the data, $P(\xi | \mathbf{d}_{HST}, \Delta \mathbf{t}, \sigma, \mathbf{d}_{LOS}, \mathbf{A})$. The vector ξ includes the lens model parameters ν , the cosmological parameters π , and nuisance parameters representing the external convergence (κ_{ext} ; Section 5) and anisotropy radius for the lens stellar velocity ellipsoid (r_{ani} ; Section 4.3). \mathbf{A} denotes a discrete set of assumptions about the form of the model, which includes the data modelling region, the source reconstruction grid, the treatment of the various deflector mass distributions, etc. In general, \mathbf{A} is not fully captured by continuous parameters. From Bayes’ theorem, we have

$$P(\xi | \mathbf{d}_{HST}, \Delta \mathbf{t}, \sigma, \mathbf{d}_{LOS}, \mathbf{A}) \propto P(\mathbf{d}_{HST}, \Delta \mathbf{t}, \sigma, \mathbf{d}_{LOS} | \xi, \mathbf{A}) P(\xi | \mathbf{A}), \quad (5)$$

where $P(\mathbf{d}_{HST}, \Delta \mathbf{t}, \sigma, \mathbf{d}_{LOS} | \xi, \mathbf{A})$ is the joint likelihood function and $P(\xi | \mathbf{A})$ is the prior PDF for the parameters given our assumptions. Since the data sets are independent, the likelihood can be separated,

$$P(\mathbf{d}_{HST}, \Delta \mathbf{t}, \sigma, \mathbf{d}_{LOS} | \xi, \mathbf{A}) = P(\mathbf{d}_{HST} | \xi, \mathbf{A}) \times P(\Delta \mathbf{t} | \xi, \mathbf{A}) \times P(\sigma | \xi, \mathbf{A}) \times P(\mathbf{d}_{LOS} | \xi, \mathbf{A}). \quad (6)$$

We can calculate the individual likelihoods separately and combine them as in equation (6) to get the final posterior PDF for a given set of assumptions.

For each of our main lens models in Sections 4.2.1 and 4.2.2, we have a range of systematics tests (Section 4.2.3) where we vary the content of \mathbf{A} and repeat the inference of ξ . These tests are important for checking the magnitude of various known but unmodelled systematic effects, but leave us with the question of how to combine the

results. We follow H0LiCOW IX in using the Bayesian Information Criterion (BIC) to weight the various models in our final inference (Section 4.4). This effectively combines our various assumptions \mathbf{A} using the BIC so that we obtain $P(\xi|\mathbf{d}_{HST}, \Delta t, \sigma, \mathbf{d}_{LOS})$. We can further marginalize over the non-cosmological parameters (ν , κ_{ext} , r_{ani}) and obtain the posterior probability distribution of the cosmological parameters π :

$$P(\pi|\mathbf{d}_{HST}, \Delta t, \sigma, \mathbf{d}_{LOS}) = \int d\nu d\kappa_{\text{ext}} dr_{\text{ani}} P(\xi|\mathbf{d}_{HST}, \Delta t, \sigma, \mathbf{d}_{LOS}). \quad (7)$$

In the lens model, we actually vary H_0 , keeping other parameters fixed at $w = -1$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. This assumes a fixed curvature of the expansion history of the Universe, but not the absolute scale (represented by H_0 or $D_{\Delta t}$). This is done because there is not a unique $D_{\Delta t}$ when accounting for multiple lens planes, but we convert this to an ‘effective’ $D_{\Delta t}$ that is insensitive to assumptions of the cosmological model (see Appendix B). Specifically, given the lens/quasar redshifts and π (i.e. H_0 and the other fixed cosmological parameters), we can compute the effective time-delay distance $D_{\Delta t}(\pi, z_d, z_s)$ to obtain the posterior probability distribution of $D_{\Delta t}$, $P(D_{\Delta t}|\mathbf{d}_{HST}, \Delta t, \sigma, \mathbf{d}_{LOS})$.

3 DATA

The data we use to infer $D_{\Delta t}$ consists of (1) the *HST* imaging used for lens modelling, which we present in Section 3.1; (2) the spectroscopy of the lensing galaxy, used to measure its stellar velocity dispersion, and (3) targeted spectroscopy of the LOS environment, both of which we present in Section 3.2; (4) wide-field multiband imaging, which we present in Section 3.3 and we use to infer κ_{ext} in Section 5; and (5) the time delays measured by COSMOGRAIL, presented in Section 3.4.

3.1 *HST* imaging

The *HST* images we use to model WFI2033 – 4723 consist of Wide Field Camera 3 (WFC3) F160W band observations (Program #12889; PI: Suyu), as well as archival Advanced Camera for Surveys (ACS) observations in the F814W filter (Program #9744; PI: Kochanek). The latter program also contains imaging in the F555W filter, which we do not use, because the signal-to-noise ratio from the lensed images is low and does not add much information to the lens model.

The details of the observations from Program #12889 are presented in H0LiCOW I. Using a combination of short (74 s) and long (599–699 s) exposures, we obtain the brightness distribution of the lens system covering a large dynamic range (of the bright lensed AGN, its much fainter host galaxy, and the foreground lens galaxy). The WFC3 images are drizzled using DRIZZLEPAC² to a final pixel scale of $0''.08$, whereas the ACS images are reduced using MULTIDRIZZLE³ on to a final pixel scale of $0''.05$. More details of the reduction are presented in H0LiCOW IV.

We create cutouts of the reduced *HST* images and define an arc-mask around the lens in each of the two filters, which encloses the region where we reconstruct the lensed arc from the extended quasar

host galaxy. We expand the cutout to the west of the lens to include the nearby galaxy G2, which is a bright perturber at $z = 0.7450$ whose light profile needs to be modelled, as it may contaminate the signal within the arc-mask. The cutout region is $10''.4 \times 6''.4$, which corresponds to a 208×128 pixel cutout for the F814W image and a 130×80 pixel cutout for the F160W image. These cutouts are shown in Fig. 1.

The reconstruction of the point spread function (PSF) for each *HST* exposure, as well as of the weight images and bad pixel masking for each cutout, are analogous to the procedure described in H0LiCOW IV. As detailed in that paper, we note that in order to avoid biasing the modelling due to large residuals from PSF mismatch near the AGN image centres, we rescale the weights in those regions by a power-law model such that a pixel originally given a noise value of p_i is rescaled to a noise value of $A \times p_i^b$. The constants A and b are chosen for each band such that the normalized residuals in the AGN image regions are approximately consistent with the normalized residuals in the rest of the arc region. For completeness, we show the residuals for models using the weight images without this power-law weighting in Appendix A. The strong residuals in these images motivates our decision to adopt this rescaling.

We note that although the background noise for the WFC3 IR camera depends on the number of non-destructive reads, we check that the number of reads in the lensed arc region is the same as for the blank sky patch used for estimating the background noise, so this procedure is valid. Since most of the lens model constraints come from the parts of the lensing arcs away from the centres of the AGN images, we check that these arcs do not have pixels that were flagged as bad in too many exposures, which would otherwise affect our lens mass model.

3.2 Spectroscopic data

Our spectroscopic observations, presented in H0LiCOW X, reveal that the lens is part of a galaxy group at $z_{\text{grp}} = 0.6588$ with a velocity dispersion of $\sigma = 500 \pm 80 \text{ km s}^{-1}$ measured from 22 member galaxies, which is independently confirmed by Wilson et al. (2016) based on a spectroscopic study by Momcheva et al. (2006, 2015).

We summarize hereafter the characteristics of the spectroscopic data used. A more exhaustive description of the data acquisition and analysis is provided in H0LiCOW X. WFI2033 – 4723 was observed with the ESO-MUSE integral field spectrograph (Bacon et al. 2010) during several observing runs between 2014 June 19 and 2016 July 20. The velocity dispersion measurement of the lensing galaxy was based on a total of 3×2400 s exposures with the lensing galaxy located close to the centre of the $1\text{arcmin} \times 1\text{arcmin}$ field of view (FOV). The data cubes are characterized by a $0''.2 \times 0''.2$ spatial sampling, a wavelength coverage in the optical range from 4800 to 9350 Å, a spectral sampling of 1.25 Å per pixel, and a resolving power $R \sim 1800\text{--}3600$ (i.e. 2.5 Å spectral resolution; Richard, R. & J. 2017). The analysis has been carried out on the combined datacube characterized by a median seeing of 1 arcsec. To deblend the lensing galaxy and the quasar images, we modelled each monochromatic slice with a model of the system composed of four Moffat (Moffat 1969) components for the quasar lensed images, and one de Vaucouleurs (de Vaucouleurs 1948) model for the lensing galaxy. After removing the quasar images from the datacube, we extracted the lensing galaxy spectrum within a square aperture of nine pixels = $1''.8$ side-length.

²DRIZZLEPAC is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

³MULTIDRIZZLE is a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

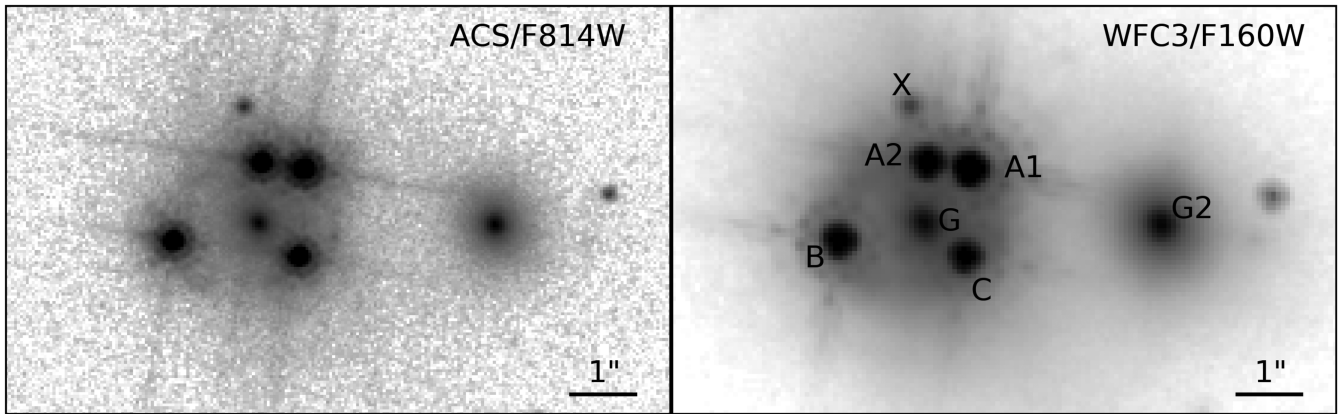


Figure 1. *HST* images of WFI2033 – 4723. Shown are cutouts of the lens system used for lens modelling in the ACS/F814W (left-hand panel) and WFC3/F160W (right-hand panel) bands. The images are $10''.4 \times 6''.4$. The scale is indicated in the bottom right of each panel. The main lens galaxy (G), lensed quasar images (A1, A2, B, and C), satellite galaxy (X), and nearby perturber (G2) are marked. The small object to the west of G2 is a foreground star.

The velocity dispersion was obtained following the same procedure as Suyu et al. (2010, 2013), resulting in an inference of $\sigma_{\text{LOS}} = 250 \text{ km s}^{-1}$ with a statistical uncertainty of $\approx 10 \text{ km s}^{-1}$. The order of the polynomial continuum and spectral regions masked for the fit introduce additional systematic uncertainties. The various choices we made have been treated as nuisance parameters over which we have marginalized to derive our final velocity dispersion PDF (see H0LiCOW X). The overall uncertainty, accounting for the random and systematic errors, reaches $\sigma_{\sigma_{\text{LOS}}} = 19 \text{ km s}^{-1}$. We integrate this measurement in our cosmographic inference in Section 4.3.

In addition to ESO MUSE spectroscopy of the galaxies located in the vicinity of the lens, we have also obtained multiobject spectroscopy of the galaxies in the FOV with the ESO FORS (Appenzeller et al. 1998) and the Gemini GMOS (Hook et al. 2004) instruments. In total, we used 10 masks, with about 35 long-slits (6 arcsec length) per mask positioned on targets located within 2 arcmin from the lens. For each mask, we obtained 40 min long exposures, and used a setup allowing to cover most of the optical wavelength range (typically 4500–9000 Å) with a resolving power of ≈ 440 (FORS) / 1100 (GMOS).

3.3 Photometric data

Our photometric data consist of wide-field optical wavelength data from the Dark Energy Survey⁴ (DES) and in particular the Data Release 1 (Abbott et al. 2018), ultraviolet data from the DES Camera (Flaugher et al. 2015) on the Blanco Telescope, VLT/HAWK-I (Pirard et al. 2004; Kissler-Patig et al. 2008) near-infrared data, and archival IRAC (Fazio et al. 2004) infrared data from the *Spitzer Space Telescope*. These data and their products, consisting of the galaxy–star classification, photometric redshifts, and stellar masses of all galaxies with $i < 23$ mag within a 120 arcsec radius around WFI2033 – 4723, are described in H0LiCOW X. In Section 5, where we measure the relative density of the environment of WFI2033 – 4723, we use a conservative cut of $i < 22.5$ mag in order to ensure that the galaxy catalogue, with a 5σ limiting magnitude of ~ 23.13 , is complete.⁵

⁴<https://www.darkenergysurvey.org>

⁵While shallow magnitude limits may bias the κ_{ext} distribution we determine in Section 5, fig. 6 in Collett et al. (2013) shows that the expected bias is at a

We show the 4 arcmin \times 4 arcmin FOV, with the galaxy catalogue overlapped, in Fig. 2.

3.4 Time delays

3.4.1 Time-delay measurements

COSMOGRAIL XVIII presents the most comprehensive analysis of the time delays of WFI2033 – 4723 so far, with the analysis of four different data sets spanning across 14 yr of monitoring, for a total of ~ 447 h of observations. The data were acquired in the scope of the COSMOGRAIL collaboration, using three different telescopes in the Southern hemisphere; the C2 and ECAM instruments mounted on the 1.2 m Leonhard Euler Swiss telescope and the WFI instrument mounted on the ESO/MPIA 2.2 m telescope, both located at La Silla Observatory in Chile, and the 1.3 m Small and Moderate Aperture Research System (SMARTS) at the Cerro Tololo Inter-American Observatory (CTIO) in Chile.

The data are split in four data sets, one per instrument (C2 and ECAM on the Euler telescope, WFI on the 2.2 m telescope and SMARTS), each being reduced independently. The photometry of the four images of WFI2033 – 4723 is recovered using the MCS deconvolution scheme (Magain, Courbin & Sohy 1998; Cantale et al. 2016). The light curves obtained are presented in Fig. 2 of COSMOGRAIL Paper XVIII. For three of the four data sets (C2, ECAM, and SMARTS), the deconvolution scheme is not able to properly resolve the flux coming from the A1 and A2 images. Thus, the A1 and A2 fluxes are summed into a virtual light-curve A, under the assumption that the time delay between A1 and A2 is zero. The WFI data set being composed of exposures of better quality, the deconvolution scheme manages to properly resolve the A1 and A2 images. A virtual light-curve $A = A1 + A2$ is also constructed for WFI in order to compare it to the other data sets.

The time-delay measurements between each pair of light curves are made with the PYCS software (Tewes et al. 2013; Bonvin et al. 2016) and follow the formalism introduced by Bonvin et al. (2018). We use two different curve shifting techniques. Both techniques share

level of ~ 0.25 per cent, which is acceptable given our goal of inferring H_0 with biases below the 1 per cent level.

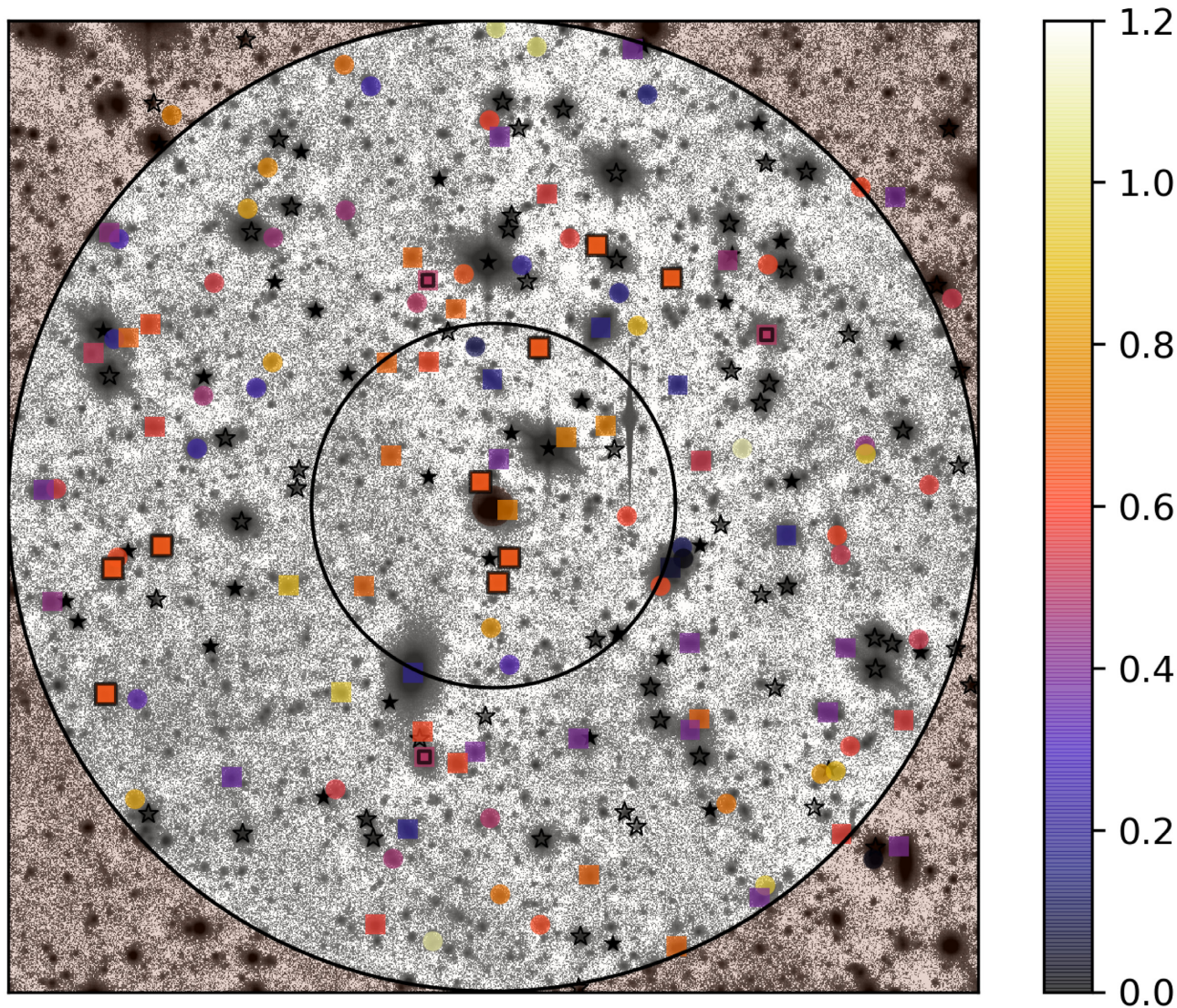


Figure 2. 240 arcsec \times 240 arcsec region around WFI2033 – 4723, overlaying the catalogue data from H0LiCOW X on top of the deepest image available, WFI R -band (see Section 3.4 and COSMOGRAIL XVIII for details). The ≤ 5 arcsec- and ≥ 120 arcsec-radius apertures are masked. The 45 arcsec - and 120 arcsec - radius apertures are marked by black circles. Detected sources with $i \leq 22.5$, corresponding to the limit used in our weighted number counts analysis, are marked: stars are marked with black star symbols, filled if confirmed spectroscopically and empty otherwise; galaxies are marked with squares if spectroscopic redshifts are available, and with circles otherwise. The colour scale corresponds to the spectroscopic redshift, if available, and to the photometric redshift, otherwise. Galaxies spectroscopically confirmed to be members of the galaxy group which includes the lensing galaxy are marked with squares with black contours, and those part of the group at $z = 0.49$ are marked with smaller square contours. For a larger FOV and more details on the available LOS spectroscopy, see Fig. 2 in H0LiCOW X.

a common framework to assess their own uncertainties, based on a statistical analysis of the residuals of the real data that prevents, by construction, involuntary fine-tuning of the curve-shifting technique parameters to recover a biased value of the time delays.

Each data set is analysed independently. The time-delay estimates obtained are in good agreement with each other and a Bayes factor analysis states that they can be combined without loss of consistency. In this work, we use the combined time-delay estimates with respect to image B. For our fiducial set of models, we use the B-A1 and B-A2 time delays estimated from the WFI data set (see Fig. 4 of COSMOGRAIL XVIII), and the B-C time delay estimated

by combining all the data sets together (labelled ‘PyCS-mult’ on Fig. 3 of COSMOGRAIL XVIII). They read $\Delta t_{B-A1} = -36.2^{+1.6}_{-2.3}$, $\Delta t_{B-A2} = -37.3^{+2.6}_{-3.0}$ and $\Delta t_{B-C} = -59.4 \pm 1.3$. Although using different time delays from different combinations of data sets might appear subjective, we recall that (i) only the WFI data set is of good enough quality to resolve the A1 and A2 images, thus bringing an additional independent constraint to the modelling and solving the potential issue of where to anchor a time-delay estimate related to a virtual image A, and (ii) all the time-delay estimates and combination of time-delay estimates are statistically consistent with each other.

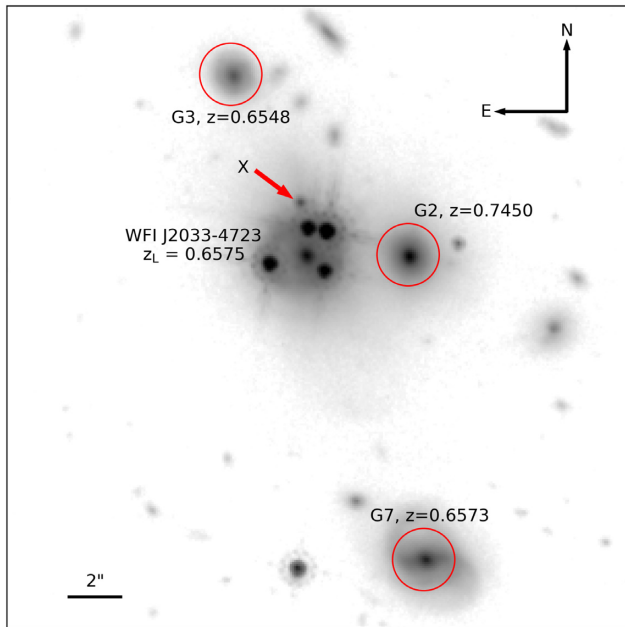


Figure 3. *HST*/WFC3 F160W image of a 24 arcsec \times 24 arcsec field around WFI2033 – 4723. The angular scale is indicated in the bottom left corner. The three most significant nearby perturbers are marked with red circles, and the redshifts of the perturbers are indicated. G2, G3, and G7 are included explicitly in our model, as they are the most massive and nearest in projection to WFI2033 – 4723. The small object X is indicated by a red arrow, and is assumed to be at the lens redshift in our models.

3.4.2 Microlensing time-delay

Our time-delay measurements do not include the contribution from the microlensing time delay (Tie & Kochanek 2018; Bonvin et al. 2019a), a time-dependent reweighting of the geometrical delay (originating from the extended spatial structure of the source) by the microlensing pattern affecting each image independently. As a result, an excess microlensing time delay adds to the excess cosmological time delay of each lensed image, and the measured time delays between pairs of images can deviate from the cosmological time delays by a non-negligible amount. The amplitude of the effect depends mainly on the mass of the central black hole of WFI2033 – 4723 (Sluse et al. 2012; Motta et al. 2017), and its estimation relies on the assumption that the accretion disc can be modelled as a thin disc (Shakura & Sunyaev 1973) – which, so far, is disfavoured by the data (see e.g. Morgan et al. 2018) – and that the emission of the accretion disc follows an idealized lamp-post model (Cackett, Horne & Winkler 2007; Starkey, Horne & Villforth 2016).

In Fig. 6 of COSMOGRAIL XVIII, we compute the amplitude of the microlensing time delay for various disc sizes. Although the measured time delays do not show any discrepancies that would be evidence for a microlensing time delay, it cannot be ruled out either. We thus chose to include it by default in our models, noting that the effect is much smaller than our other uncertainties. We follow the framework presented in Chen et al. (2018a) and assume the accretion disc size of Morgan et al. (2018) with $r = R_0$. We also test the effect of ignoring the microlensing time delay for one of our models, finding that it changes the $D_{\Delta t}$ accuracy by < 1 per cent (Section 4.2.3).

4 LENS MODELLING

In this section, we describe our procedure to simultaneously model the images in the two *HST* bands, and the time delays, in order to infer the lens model parameters and $D_{\Delta t}$.

4.1 Overview

We perform our lens modelling using GLEE, a software package developed by S. H. Suyu and A. Halkola (Suyu & Halkola 2010; Suyu et al. 2012b). The lensing mass distribution is described by a parametrized profile. The extended host galaxy of the source is modelled separately on a 50×50 pixel grid with curvature regularization (Suyu et al. 2006). The lensed quasar images are modelled as point sources on the image plane convolved with the PSF. The quasar image amplitudes are allowed to freely vary and are independent from the extended host galaxy light distribution to allow for variability due to microlensing, time delays, and substructure. The lens galaxy light distribution is modelled using either Sérsic profiles or Chameleon profiles. The Sérsic profile is defined as

$$I(\theta_1, \theta_2) = A \exp \left[-k \left(\left(\frac{\sqrt{\theta_1^2 + \theta_2^2/q_L^2}}{r_{\text{eff}}} \right)^{1/n} - 1 \right) \right], \quad (8)$$

where A is the amplitude, k is a constant such that r_{eff} is the effective (half-light) radius, q_L is the axial ratio, and n is the Sérsic index. The Chameleon profile (also known as the pseudo-Jaffe profile) is defined as the difference of two non-singular r^{-2} elliptical profiles (Kassiola & Kovner 1993; Dutton et al. 2011), which are a good approximation to Sérsic profiles.

We represent the galaxy light distribution as the sum of two Sérsic (or two Chameleon) profiles plus a point source (to account for possible AGN emission from the lens galaxy) with a common centroid. Since the light of G2 can also influence the model, we represent its light distribution as a single Sérsic profile plus a point source with a common centroid, although we mask its central regions since we only care about light from G2 that could affect the lens galaxy or arc light. There is a small nearby perturber (‘X’ in Fig. 1), which we also represent as a single Sérsic profile plus a point source with a common centroid. Model parameters of the lens and source are constrained through Markov Chain Monte Carlo (MCMC) sampling.

In accounting for perturbers at different redshifts from the main lens galaxy, we use the full multiplane lens equation (e.g. Blandford & Narayan 1986; Kovner 1987; Schneider et al. 1992; Petters et al. 2001; Collett & Auger 2014; McCully et al. 2014) in our modelling. We vary H_0 directly in our models and use this distribution to calculate the effective model time-delay distance $D_{\Delta t}^{\text{model}}$. In calculating $D_{\Delta t}^{\text{model}}$, we assume $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $w = -1$. Relaxing these assumptions by allowing these cosmological parameters to vary freely shifts the resulting $D_{\Delta t}^{\text{model}}$ distributions by < 1 per cent in previous analyses (see H0LiCOW IV), and we also verify that this is true for WFI2033 – 4723 (Appendix B). Thus, this approximation has no measurable effect on the inferred time-delay distance, which can then be applied to constrain any arbitrary cosmology.

4.2 Mass models

Our primary mass models for the lens galaxy are a singular power-law elliptical mass distribution (SPEMD; Barkana 1998), and a model

consisting of a baryonic component that traces the light distribution plus a separate dark matter component (hereafter the ‘composite’ model). We also include an external shear in the strong lens plane. Non-linear couplings due to multiplane effects are small and thus ignored.

We explicitly include the nearby perturber X in the lens model, linking its mass centroid to that of its light. Although we do not have a spectroscopic confirmation of the redshift of X, it is likely a satellite galaxy that is physically associated with the lens galaxy, given its small size and proximity. We also see evidence in the F160W image for possible tidal features emanating from X in the direction away from the lens galaxy, suggesting that it may be an infalling satellite. We therefore assume that X is at the same redshift as the lens and parametrize it as a singular isothermal sphere (SIS). In our models, X generally has a much smaller mass than the main lens, and therefore it has a minor influence on the potential, even if it is located at a different redshift.

We also explicitly include three nearby massive perturbing galaxies (denoted G2, G3, and G7, following the naming convention of Vuissoz et al. 2008) in Fig. 3 that are projected close to the lens. G2 is close enough that its influence may not be adequately described by external shear (H0LiCOW II; see also McCully et al. 2017), and H0LiCOW X showed that G3 ($z = 0.6548$) and G7 ($z = 0.6573$) may have a non-negligible higher order influence on the model as well. Our updated estimation of the influence of these galaxies, computed in terms of the flexion shift considered in H0LiCOW X but taking into account the galaxy morphologies and velocity dispersions measured in that paper, shows that G2, with $\log M_* \sim 11.15$, is in fact the only galaxy with significant impact on the modelling. Nonetheless, based on their proximity to the lensing galaxy, we choose to explicitly model G3 ($\log M_* \sim 10.17$) and G7 ($\log M_* \sim 11.16$) as well. G2 is modelled as a singular isothermal ellipsoid, which is a reasonable assumption since higher order moments of the potential will have a small effect at the position of the main lens galaxy. G3 and G7 are modelled as SIS. The relative Einstein radii of G2, G3, and G7 are calculated from their measured velocity dispersions (H0LiCOW X), assuming isothermal profiles. The ratio of their Einstein radii is fixed, but with a global scaling allowed to vary freely, as in H0LiCOW IV, IX. This is done to prevent the model from optimizing the perturbers’ Einstein radii in a way that would be inconsistent with their measured redshifts and velocity dispersions. The centroid of G2 is linked to the centroid of its light distribution in the F160W band in the modelling, while the centroids of G3 and G7 are fixed to their measured positions in the F160W image. We set the redshifts of G3 and G7 equal to the lens redshift of $z = 0.6575$ in our model, as their redshifts are consistent with this value within the range allowed by peculiar velocities. All masses are treated using the full multiplane lens equation, as detailed by Suyu et al. (in preparation).

Our constraints on the primary lens model include the positions of the lensed quasar images, the measured time delays, and the surface brightness of the pixels in the ACS/F814W and WFC3/F160W images that are fit simultaneously. The quasar positions are fixed to the positions of the point sources on the image plane (after they have stabilized) and are given a Gaussian uncertainty of width $0''.004$ to account for offsets due to substructure in the lens or LOS, which is small enough to satisfy astrometric requirements for cosmography (Birrer & Treu 2019). The quasar flux ratios are not used as constraints, as they can be affected by microlensing. We first model the lens separately in each band to iteratively update the respective PSFs using the lensed AGN images themselves, similar to Chen et al. (2016), but with the PSF corrections and source intensity reconstructed simultaneously in our case (H0LiCOW IV,

IX) rather than separately. We keep these ‘corrected’ PSFs fixed and use them in our final models that simultaneously use the surface brightness distribution in both bands as constraints. We then use the positions of the quasar images to align the images in the two *HST* bands. We do not enforce any similarity of pixel values at the same spatial position across different bands (i.e. the flux at any position in one band is independent of the other band). We also directly include the effect of microlensing time delays, as described in Section 3.4.2, although our tests show that this has a very small effect on our results (Section 4.2.3). In our MCMC sampling, we vary the light parameters of the lens galaxy, G2, X, and quasar images, the mass parameters of the lens galaxy, X, G2, G3, and G7, the external shear, and H_0 . The quasar image positions are linked across both bands, but the other light parameters are allowed to vary independently.

4.2.1 Power-law model

Our fiducial SPMD model uses the double Sérsic parametrization for the lens galaxy light and has the additional free parameters:

- (i) position (θ_1, θ_2) of the centroid (allowed to vary independently from the centroid of the light distribution)
- (ii) Einstein radius θ_E
- (iii) minor-to-major axial ratio q and associated position angle θ_q
- (iv) 3D slope of the power-law mass distribution γ'
- (v) position of X, linked to its light centroid
- (vi) Einstein radius of X
- (vii) position of G2, linked to its light centroid
- (viii) global scaling parameter that controls the Einstein radii of G2, G3, and G7
- (ix) minor-to-major axial ratio q and associated position angle θ_q of G2
- (x) external shear γ_{ext} and associated position angle θ_γ ⁶
- (xi) the Hubble constant, H_0 .

We conservatively assume uniform priors on the model parameters over a wide physical range. Although the lens is not drawn from a random population, but rather with some selection function that could, in principle, bias the inferred time-delay distance, this selection function is not well known and these biases are negligible for this type of analysis (e.g. Collett & Cunningham 2016). The parameters that are exceptions to our choice of uniform priors are that the global scaling parameter for the Einstein radii of the perturbers is given a Gaussian prior such that the expected mean and uncertainty of G2’s Einstein radius is constrained by its measured velocity dispersion, and that the position angle θ_q of G2’s is given a Gaussian prior based on the fit of its light profile. We anchor the scaling parameter to G2 as it is the perturber with the most precisely measured velocity dispersion, and its proximity to the lens makes it the most significant of the three massive perturbing galaxies.

Fig. 4 shows the data and the lens model results in both bands for our fiducial SPMD model, as well as the source reconstructions. Our model reproduces the surface brightness structure of the lensed AGN and host galaxy in both bands simultaneously.

⁶ θ_γ is defined to be the direction of the shear itself, i.e. orthogonal to the direction of the mass producing the shear.

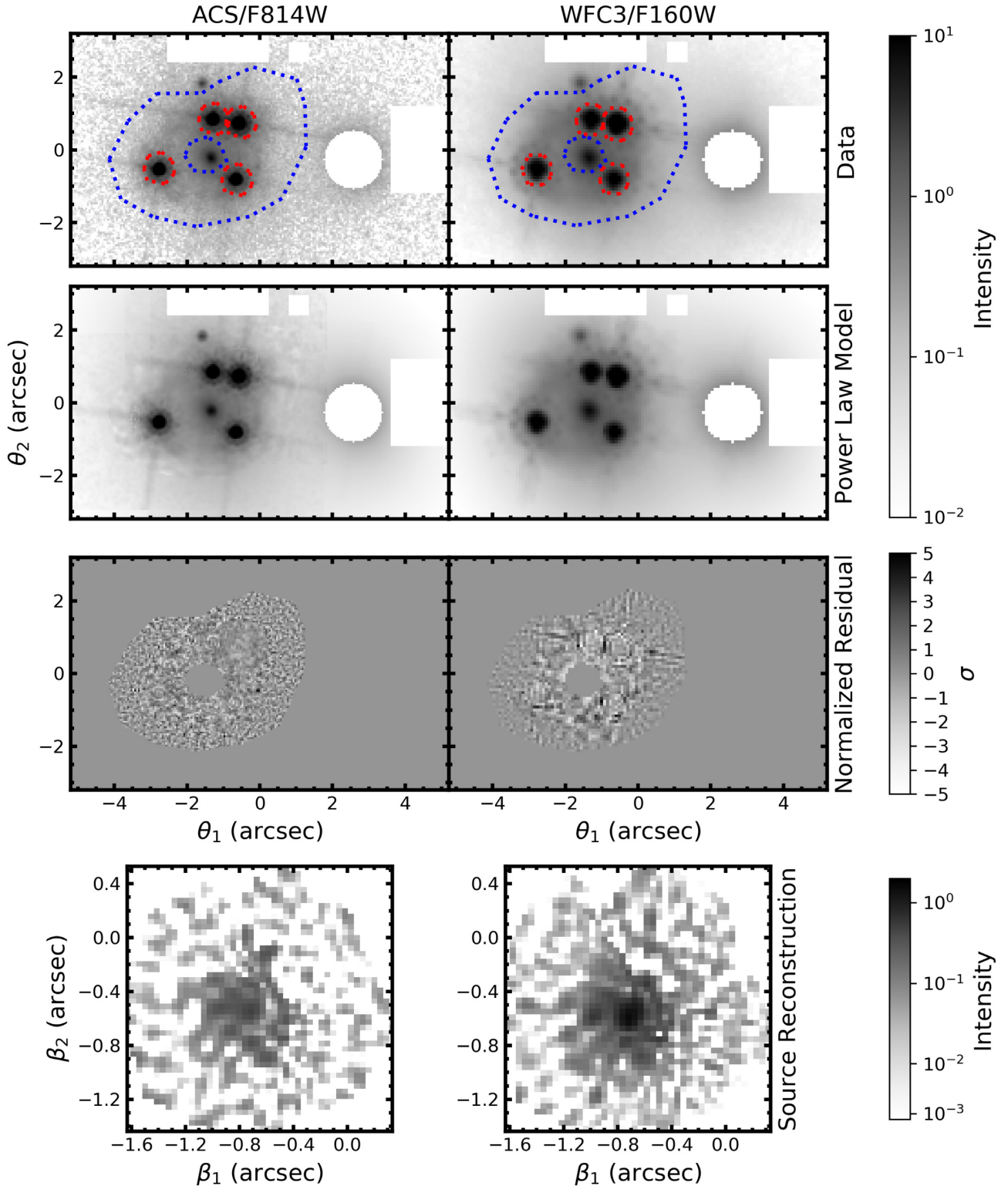


Figure 4. SPEMD lens model results for ACS/F814W (left-hand panel) and WFC3/F160W (right-hand panel). Shown are the observed image (top row), the reconstructed image predicted by the model (second row), the normalized residual within the arcmask region (defined as the difference between the data and model, normalized by the estimated uncertainty of each pixel; third row), and the reconstructed source (bottom row). This uses the weight image with the power-law rescaling near the AGN images. We show the normalized residuals without this rescaling in Appendix A. In the top row, the blue dotted lines indicate the arcmask (donut-shaped) region used for fitting the extended source, the red dotted lines indicate the AGN mask region where the power-law weighting is applied, and the region outside the blue dotted arcmask is used to further constrain the foreground lens light and (partly) the AGN light (but not the AGN host galaxy light since its corresponding lensed arcs are below the noise level in this outer region). The white regions indicate areas of the image that are masked out during the modelling. The colour bars show the scale in the respective panels. The results shown here are for the fiducial SPEMD model, but the results for the other systematics tests (Section 4.2.3) are qualitatively similar.

4.2.2 Composite model

We follow Suyu et al. (2014) and H0LiCOW IV to construct the composite model, consisting of a baryonic component linked to the light profile of the lens galaxy, plus a dark matter component. The composite model assumes the double Chameleon light profile for the lens galaxy in the WFC3/F160W band scaled by an overall mass-to-light (M/L) ratio. We use the Chameleon light profiles for the composite model because it is straightforward to link the parameters describing the light distribution to those of the mass distribution, as they are fundamentally just a combination of isothermal profiles. We use the F160W band because it probes the rest-frame near-infrared and thus should be the best tracer of stellar mass. Although we include a point source in the light profile, we assume that this is due to low-level AGN emission from the lens galaxy, and do not associate it with a massive component in the model. This point source is roughly ~ 2 per cent of the total light in the F160W band, so its inclusion would have a minor impact on our results. We keep the double Sérsic parametrization for the lens galaxy light in the F814W band to maintain consistency with the SPEMD models. The dark matter component is modelled as an elliptical NFW (Navarro, Frenk & White 1996) potential with the centroid linked to the light centroid in the F160W band, as non-contracted NFW profiles are a good representation of the dark matter haloes of massive elliptical galaxies (Dutton & Treu 2014).

Our fiducial composite model has the same free parameters (v) to (xi) as the SPEMD model in Section 4.2.1, as well as the additional parameters:

- (1) M/L ratio for the baryonic component
- (2) NFW halo normalization $\kappa_{0,h}$ (defined as $\kappa_{0,h} \equiv 4\kappa_s$; Golese & Kneib 2002)
- (3) NFW halo scale radius r_s
- (4) NFW halo minor-to-major axial ratio q and associated position angle θ_q

We set a Gaussian prior of $r_s = 11''.9 \pm 1''.6$ based on the results of Gavazzi et al. (2007) for lenses in the Sloan Lens ACS Survey (SLACS; Bolton et al. 2006) sample, which encompasses the redshift and stellar mass of WFI2033 – 4723. All other parameters are given uniform priors, again with the exception of the Gaussian prior on global scaling parameter based on G2’s Einstein radius, as well as the Gaussian prior on G2’s position angle. The relative amplitudes of the two Chameleon profiles that represent the stellar light distribution of the lens galaxy can vary, but the relative amplitudes of these two components in the mass profiles are fixed. To account for this, we iteratively run a series of MCMC chains and update the relative amplitudes of the two mass components to match that of the light components after each chain. We iterate until the inferred H_0 stabilizes, then combine the chains after this point into a single distribution to represent the fiducial composite model. The other composite models use fixed relative amplitudes of the mass components based on the latest iteration of the fiducial composite model.

Fig. 5 shows the data and the lens model results in both bands for the fiducial composite model described in this section, as well as the source reconstruction.

4.2.3 Systematics tests

In this section we describe a range of tests of the effects of various systematics in our modelling, stemming from different assumptions

in the way we constructed the model that might affect the posterior. In addition to the basic fiducial models described above, we perform inferences for both the SPEMD and composite models given the following sets of assumptions:

- (i) A model with the arc-mask region increased by one pixel on both the inner and outer edges. To compensate for the larger arc-mask region, we increase the source plane resolution to 60×60 pixels in all bands.
- (ii) A model where the region near the AGN images scaled by the power-law weighting is increased by one pixel around the outer edge. Increasing these regions by more pixels would start to greatly reduce the area of the arc-mask where we fit the extended source.
- (iii) A model where the regions near the AGN images are given zero weight rather than being scaled by a power-law weighting.
- (iv) A model that includes the group at $z = 0.6588$ (of which the lens galaxy is a member) as a spherical NFW halo. The halo centroid and mass are given Gaussian priors based on the calculations of H0LiCOW X. The scale radius is given a Gaussian prior of $r_{s,g} = 32''.0 \pm 8''.0$ from a calculation of its virial mass and radius (H0LiCOW X) and a halo concentration based on the results of Diemer (2018). The redshift of the group is set to the lens redshift ($z = 0.6575$), as the difference can be explained by peculiar velocity.
- (v) A model that includes both the group at $z = 0.6588$ (again set to the lens redshift) and a foreground group at $z = 0.4956$ which may have a significant effect on the lens potential based on H0LiCOW X, who estimate its flexion shift (following the definition in McCully et al. 2014). The foreground group’s centroid, mass, and scale radius are given Gaussian priors in the same way as for the group at the lens redshift. The scale radius prior from H0LiCOW X and Diemer (2018) is $r_{s,gf} = 34''.8 \pm 9''.3$.

In addition to the above models for both the SPEMD and composite models, we run one additional SPEMD model:

- (i) A model where the light profile of the lens galaxy in both bands is represented by the sum of two Chameleon profiles rather than the sum of two Sérsic profiles.

As described in Section 4.4, we combine the MCMC chains from all of these tests, weighted by the BIC (e.g. H0LiCOW IX). We calculate the relative BIC for the SPEMD models and composite models separately, then give the combined distributions equal weight in the final inference so that we are not biased by the parametrization of the mass profile.

We also run a test to verify that the microlensing time delay does not significantly impact our results. We test our fiducial SPEMD model without including the microlensing time-delay effect and compare the blinded effective time-delay distance to the model with this effect included, in Fig. 6. We find that the microlensing time delay affects the inferred $D_{\Delta t}$ at < 1 per cent, so its inclusion in our models, given our assumptions about the disc size, does not have an appreciable effect.

4.2.4 Comparison of power-law and composite models

The marginalized parameter distributions of the SPEMD model are shown in Fig. 7. We show the combined distributions of all SPEMD models where each model is given equal weight, as well as the BIC-weighted distribution. The parameter statistics for each model are given in Appendix C. There are some minor variations in the model parameters from model to model, but the $D_{\Delta t}$ distributions are generally consistent. We note that the model with Chameleon light

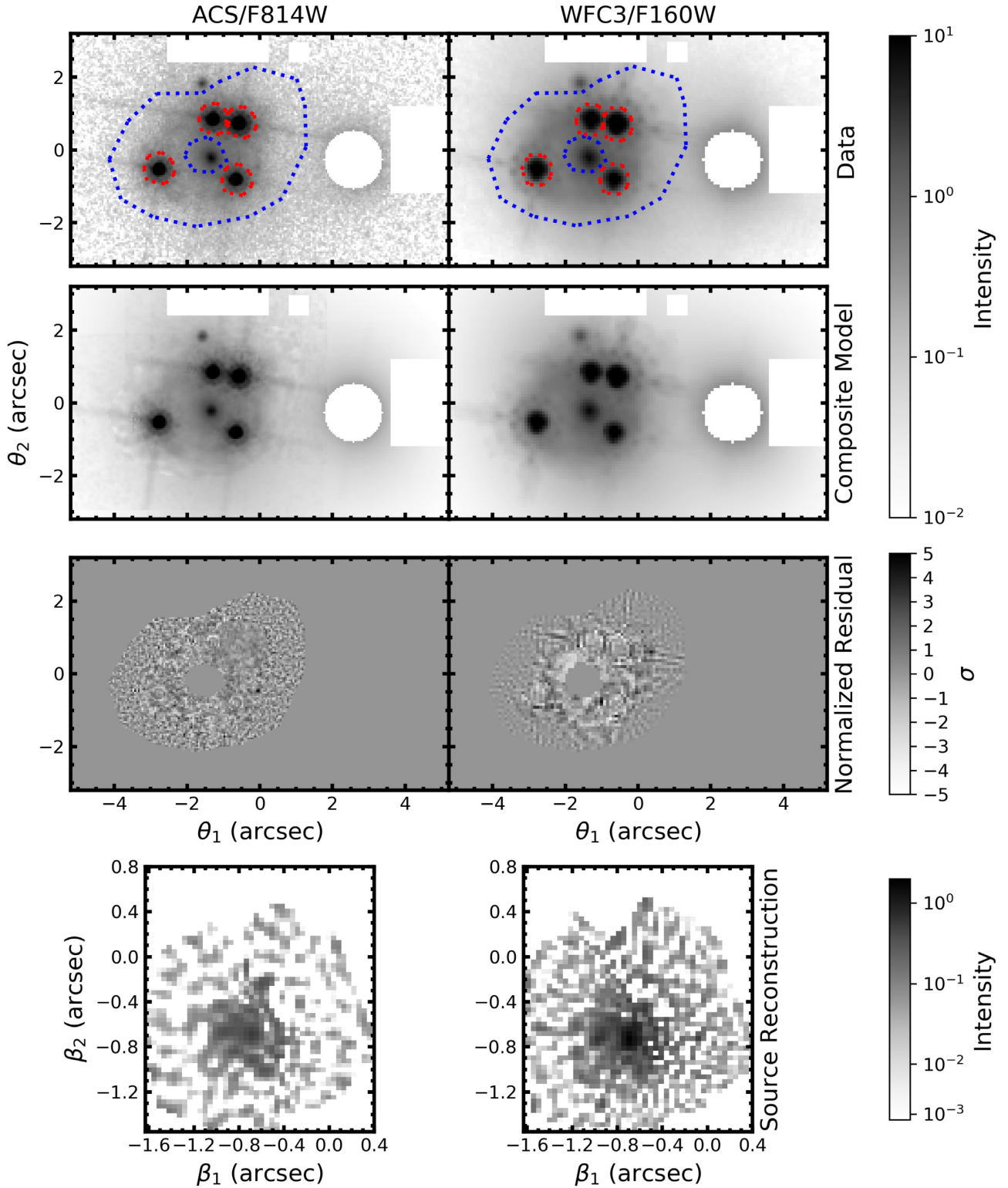


Figure 5. Same as Fig. 4, but for the fiducial composite model.

profiles for the lens galaxy is somewhat offset towards a lower $D_{\Delta l}$ (see Section 6). This model is disfavoured by our BIC weighting, so this has a minimal effect on our final results. This does not necessarily mean that the Chameleon profiles in general are a bad fit to the lens galaxy light, as the composite models (which use the Chameleon

light profile by default in the F160W band) are not similarly offset.

The multimodal distributions in some of the parameters arises primarily from differences in the posterior PDFs of different models corresponding to the various systematics tests, not from bimodality

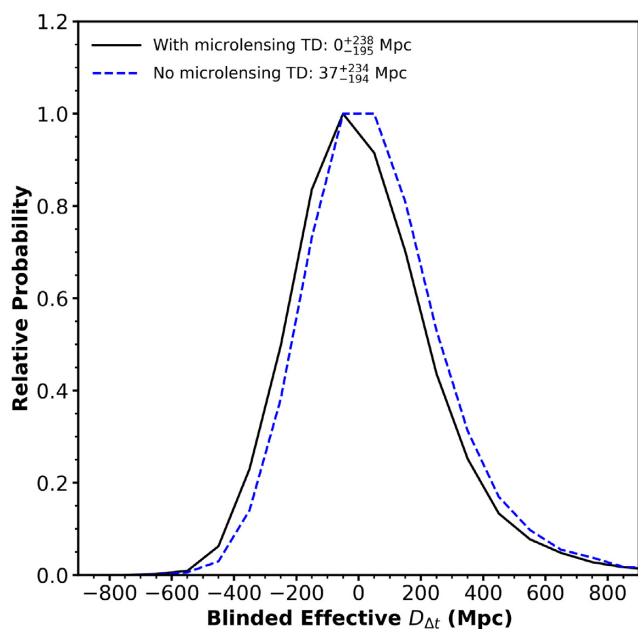


Figure 6. PDF of $D_{\Delta t}$ for the fiducial SPEMD model with (black) and without (blue) the microlensing time-delay effect. The median of the blinded effective time-delay distance PDF is insensitive to the microlensing time-delay effect to within 1 per cent.

within individual lens models. We note that despite this multimodal behaviour, the effective $D_{\Delta t}$ distribution remains stable and unimodal, suggesting that the cosmological inference is robust to the various systematics tests.

The model that includes both group haloes has the highest BIC weighting for both the SPEMD and composite models. To check that the addition of the $z = 0.4956$ group contributes meaningful information to the modelling, we run a test where the centroid of this group is given a prior located at a similar distance but rotated by 90° and 135° on the sky relative to the lens. We compare the BIC weight values of these test models to that of the model with just the group at the lens redshift and the original model with both groups. These test cases show a lower BIC weight than the original model with both groups, suggesting that the addition of the foreground group with the actual centroid prior is contributing information, although the small BIC difference is within the typical BIC variance, so it is difficult to draw a firm conclusion. The $D_{\Delta t}$ distributions remain robust within the uncertainties for each of these test cases.

The offset between the mass centroid and the light centroid in the F160W band for the SPEMD model is typically $\sim 0''.02 - 0''.03$ (roughly 150–200 pc for a flat Λ CDM cosmology with $h = 0.7$ and $\Omega_m = 0.3$) such that the mass centroid is slightly southeast of the light centroid. This might be partially explained by the influence of object X, although we note that in our SPEMD models, the mass of X is consistent with zero. The centroids of the light profiles in F814W and F160W are consistent with each other at the $\sim 0''.002$ level for both models. The SPEMD models are able to fit the quasar positions to an rms of $\sim 0''.01$, while the composite models have a larger rms of $\sim 0''.025$. Despite these differences, the SPEMD and composite models’ $D_{\Delta t}$ distributions are not drastically different, and by weighting them equally in the final inference, we are accounting in part for the astrometric uncertainty.

We show the marginalized parameter distributions of the composite model in Fig. 8. Again, we show the uniformly combined dis-

tributions as well as the BIC-weighted composite model separately, and the parameter statistics for each model are given in Appendix C. As with the SPEMD model, there are small variations in the model parameters, but the $D_{\Delta t}$ inference is consistent.

We compare the physical parameters of our BIC-weighted SPEMD model to the composite model. The results are shown in Table 1, with the parameter statistics for all composite models given in Appendix C.

4.3 Kinematics

We compute the LOS stellar velocity dispersion of the strong lens galaxy through the spherical Jeans equation (see also Treu & Koopmans 2002; Koopmans et al. 2003), similar to previous H0LiCOW analyses (e.g. Suyu et al. 2010, H0LiCOW IV). Yıldırım et al. (2019) recently showed that the assumption of spherical Jeans equation is applicable to time-delay cosmography with a single aperture-averaged lens velocity dispersion without significant bias, as in our case of WFI2033 – 4723. For a given lens model, we obtain the 3D mass profile of the lens galaxy by taking the spherical deprojection of the circularized surface mass density profile. The resulting 3D profile assumes analytical forms for both the SPEMD and the composite model. The 3D distribution of tracers is obtained by applying the same procedure to the surface brightness distribution of the lens galaxy, modelled as a Hernquist (1990) profile. We also tested a Jaffe (1983) profile, which has been shown to produce similar results (Suyu et al. 2010), and find that the results change by less than 1 per cent. We parametrize the orbital anisotropy profile as an Osipkov–Merritt model (Osipkov 1979; Merritt 1985)

$$\frac{\sigma_\theta^2}{\sigma_r^2} = 1 - \frac{r^2}{r_{\text{ani}}^2 + r^2}, \quad (9)$$

where σ_θ and σ_r are the tangential and radial velocity dispersions, respectively. Given values of the lens mass parameters in Section 4.2, the external convergence κ_{ext} in Section 5, and the anisotropy radius r_{ani} , we then calculate the LOS velocity dispersion profile by numerically integrating the solutions of the spherical Jeans equation as given by Mamon & Łokas (2005). Finally, we calculate the integral over the spectroscopic slit of the seeing-convolved brightness-weighted LOS velocity dispersion σ^{P} (equation 20 of Suyu et al. 2010) and compare to the measurements to calculate the likelihood of the kinematics data,

$$P(\sigma_{\text{LOS}} | \mathbf{v}, \boldsymbol{\pi}, \kappa_{\text{ext}}, r_{\text{ani}}) = \frac{1}{\sqrt{2\pi}\sigma_{\sigma_{\text{LOS}}}} \exp \left[-\frac{(\sigma^{\text{P}}(\mathbf{v}, \boldsymbol{\pi}, \kappa_{\text{ext}}, r_{\text{ani}}) - \sigma_{\text{LOS}})^2}{2\sigma_{\sigma_{\text{LOS}}}^2} \right], \quad (10)$$

where $\sigma_{\text{LOS}} = 250 \text{ km s}^{-1}$ and $\sigma_{\sigma_{\text{LOS}}} = 19 \text{ km s}^{-1}$ (H0LiCOW X). We adopt a uniform prior on r_{ani} in a range from 0.5 to 5 times the effective radius, r_{eff} , which we calculate to be $r_{\text{eff}} = 1''.41$ from our lens light fitting in the F160W filter. We fit to the double Sérsic light profile, as the Chameleon profile does not provide an accurate representation of the galaxy light distribution at large radii (Dutton et al. 2011). The point source contributes a very small amount to the galaxy light, but not enough to impact this calculation. We note that the choice of filter affects r_{eff} , but the impact is small and results in a negligible effect ($\lesssim 0.1$ per cent) on the final inference.

We use importance sampling (e.g. Lewis & Bridle 2002) to simultaneously combine the velocity dispersion and external convergence distributions in Section 5 with the $D_{\Delta t}^{\text{model}}$ inferred from our lens model. Specifically, for each set of lens and cosmological parameters $\{\mathbf{v}, \boldsymbol{\pi}\}$ from our lens model MCMC chain, we draw a κ_{ext} sample

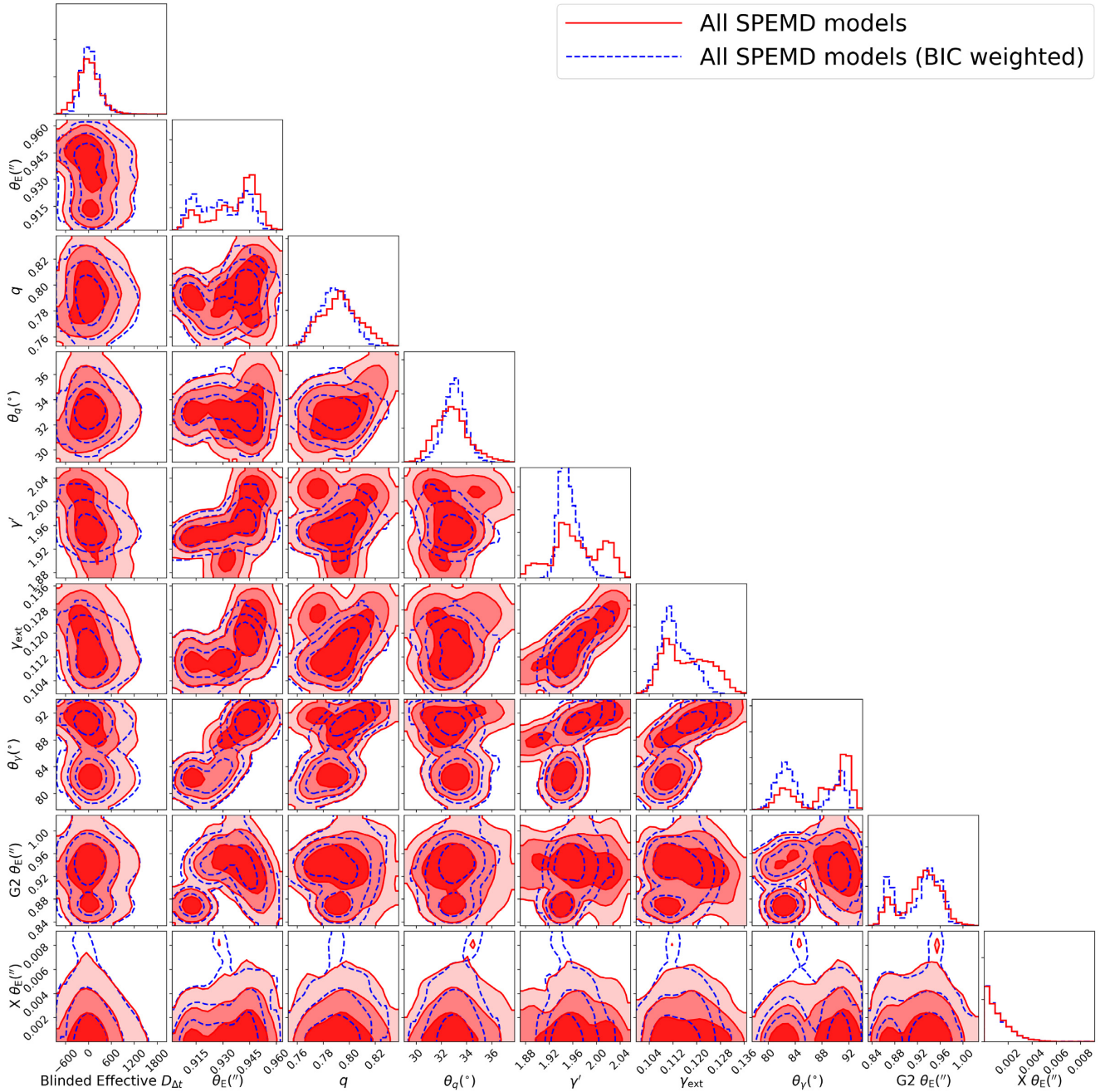


Figure 7. Marginalized parameter distributions from our SPEMD lens model results. We show the combined results from our systematics tests (shaded red contours) with each model weighted equally, as well as the BIC-weighted model results (dashed blue contours). The contours represent the 68.3 percent, 95.4 percent, and 99.7 percent quantiles.

from the distribution in Section 5 and a sample of r_{ani} from the uniform distribution $[0.5, 5]r_{\text{eff}}$. With these, we can then compute the kinematics likelihood in equation (10) for the joint sample $\{\mathbf{v}, \boldsymbol{\pi}, \kappa_{\text{ext}}, r_{\text{ani}}\}$ and use this to weight the joint sample. From the effective *model* time-delay distance computed from our multiplane lensing ($D_{\Delta t}^{\text{model}}$) and the external convergence (κ_{ext}), we can then compute the effective time-delay distance ($D_{\Delta t}$) via equation (4), keeping its absolute value blinded until we finalize our analysis. The resulting distribution of $D_{\Delta t}$ encapsulates the cosmological information from WFI2033 – 4723.

4.4 BIC weighting

We weight our models using the BIC, defined as

$$\text{BIC} = \ln(n)k - 2\ln(\hat{L}). \quad (11)$$

n is the number of data points, which is the number of pixels in the image region across both bands that are outside the fiducial AGN mask (so that we are comparing equal areas), plus eight (for the four AGN image positions), plus three (for the time delays), plus one (for the velocity dispersion). k is the number of free parameters,

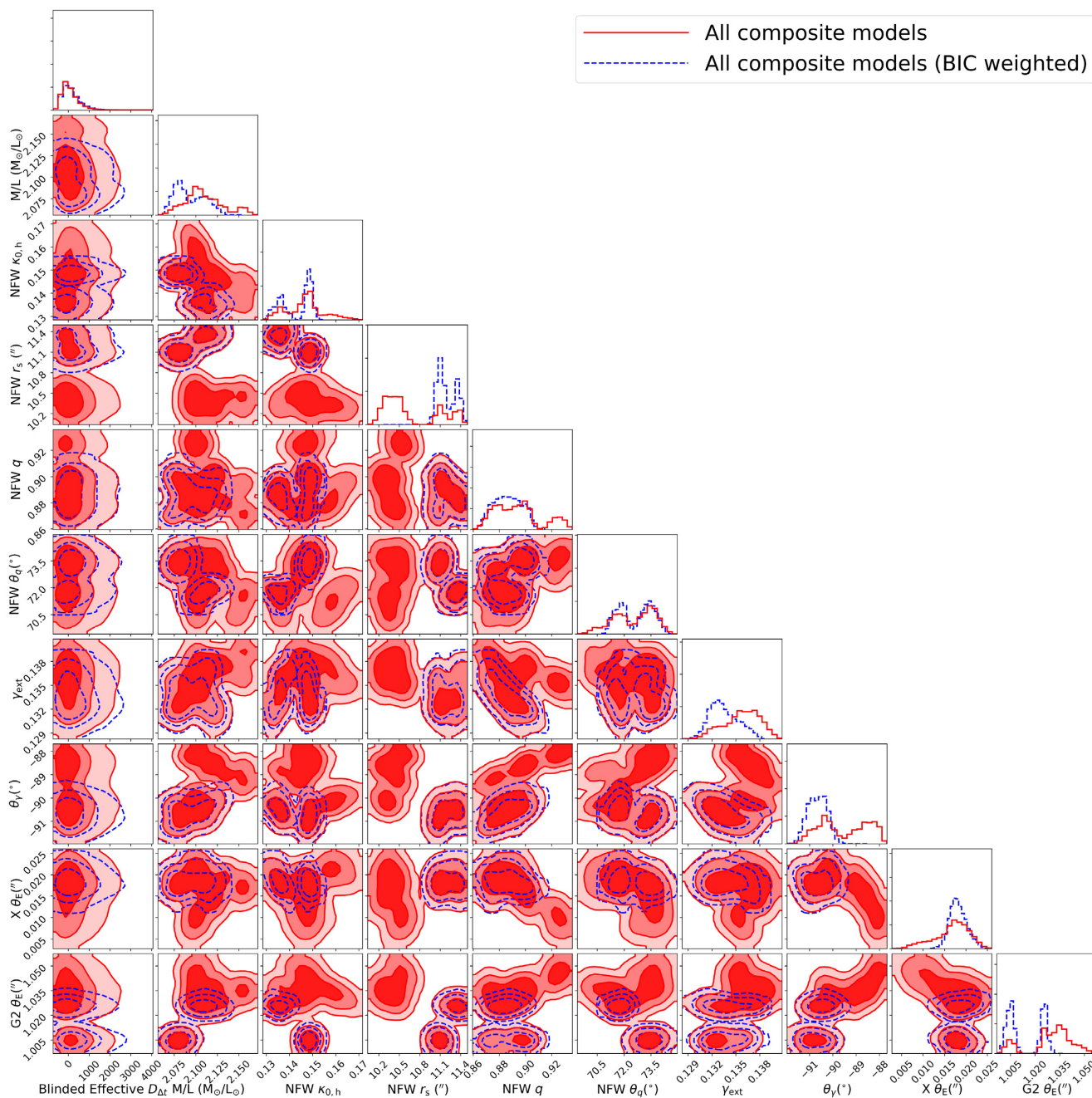


Figure 8. Marginalized parameter distributions from our composite lens model results. We show the BIC-weighted model (dashed blue contours) and the combined results from our systematics tests (shaded red contours). The contours represent the 68.3 per cent, 95.4 per cent, and 99.7 per cent quantiles.

which is the number of parameters in the lens model that are given uniform priors, plus two (for the source position), plus one (for the anisotropy radius to predict the velocity dispersion). \hat{L} is the maximum likelihood of the model, which is the product of the AGN position likelihood, the time-delay likelihood, the pixellated image plane likelihood, and the kinematic likelihood. The image plane likelihood is the Bayesian evidence of the pixellated source intensity reconstruction using the arcmask imaging data (which marginalizes over the source surface brightness pixel parameters and is thus the likelihood of the lens/cosmological parameters excluding the source

pixel parameters; see Suyu & Halkola 2010) times the likelihood of the lens model parameters within the image plane region that excludes the arcmask. We evaluate the BIC using the fiducial weight image and arcmask, as the majority of the models were optimized with these. This may penalize the model with a larger AGN mask and the 60×60 pixel source grid model with a larger arcmask, but choosing any other region would penalize the fiducial model and all the other models that used the same regions, so this choice is fair to the largest number of models. We note that our computation of the BIC described above uses all available data sets (lensing image, time

Table 1. Lens model parameters.

Parameter	BIC-weighted marginalized constraints
Singular power-law ellipsoid model	
θ_E (") ^a	$0.929^{+0.015}_{-0.016}$
q	$0.79^{+0.01}_{-0.01}$
θ_q (°)	$33.1^{+0.8}_{-0.9}$
γ'	$1.95^{+0.02}_{-0.01}$
γ_{ext}	$0.112^{+0.006}_{-0.004}$
θ_γ (°)	$83.6^{+7.1}_{-2.1}$
X θ_E (")	$0.001^{+0.001}_{-0.001}$
G2 θ_E (")	$0.932^{+0.027}_{-0.062}$
G2 q	$0.66^{+0.04}_{-0.02}$
G2 θ_q (°)	$38.5^{+4.4}_{-4.1}$
Composite model	
Stellar M/L (M_\odot/L_\odot) ^b	$2.1^{+0.2}_{-0.2}$
Chameleon1 q	$0.762^{+0.003}_{-0.003}$
Chameleon1 θ_q (°)	$23.3^{+0.2}_{-0.2}$
Chameleon2 q	$0.771^{+0.002}_{-0.003}$
Chameleon2 θ_q (°)	$26.2^{+0.3}_{-0.5}$
NFW $\kappa_{0,h}$	$0.147^{+0.002}_{-0.012}$
NFW r_s (")	$11.15^{+0.21}_{-0.09}$
NFW q	$0.89^{+0.01}_{-0.01}$
NFW θ_q (°)	$72.9^{+0.8}_{-1.3}$
γ_{ext}	$0.133^{+0.002}_{-0.001}$
θ_γ (°)	$89.4^{+0.4}_{-0.4}$
X θ_E (")	$0.018^{+0.002}_{-0.002}$
G2 θ_E (")	$1.008^{+0.019}_{-0.004}$
G2 q	$0.93^{+0.01}_{-0.01}$
G2 θ_q (°)	$39.4^{+1.6}_{-10.3}$

Notes. Reported values are medians, with errors corresponding to the 16th and 84th percentiles.

Angles are measured east of north.

^aSpherical-equivalent Einstein radius.

^bM/L within θ_E for rest-frame V band. The point source component of the lens light is assumed to be from low-level AGN emission as opposed to stellar light and is not included in the calculation. The given uncertainties are a combination of statistical effects and a systematic uncertainty equal to the difference between the calculated M/L with and without the point source contribution. The stellar mass is calculated assuming $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, but changes in the cosmology affect the M/L by a negligible amount.

delays, and lens velocity dispersion) for model comparison.⁷ The kinematics have a relatively small impact in comparison to the other terms, and does not strongly favour either the SPEMD or composite model.

We estimate the variance in the BIC, σ_{BIC}^2 , by running the fiducial model with source resolutions of [47, 48, 49, 50, 51, 52, 53, 54, 56, 58, 60] pixels on a side (the 50×50 pixel case is just the original fiducial model), keeping the arcmask the same. Changing the source resolution in this way shifts the inferred $D_{\Delta t}$ values stochastically, but there is no overall trend with resolution, and the degree of the shifts are smaller than the scatter among the different models we run. We

⁷Ranking the lensing mass models based on BIC values computed from only the lensing data would lead to insignificant changes to the final BIC-weighted $D_{\Delta t}$ distributions.

calculate the BIC for each of these models and take the variance of this set of models as the variance on the BIC, σ_{BIC}^2 . We find $\sigma_{\text{BIC}}^2 \sim 41$ for the SPEMD models and $\sigma_{\text{BIC}}^2 \sim 55$ for the composite models. In Appendix D, we show the BIC and BIC weight values for these source resolution tests.

To avoid biases due to our choice of lens model parametrization, we split the samples into the SPEMD and composite models and calculate the relative BIC and weighting for each set separately, similar to H0LiCOW IX. Specifically, we weight a model with a given BIC of x by a function $f_{\text{BIC}}(x)$, defined as the convolution

$$f_{\text{BIC}}(x) = h(x, \sigma_{\text{BIC}}) * \exp\left(-\frac{x - \text{BIC}_{\text{min}}}{2}\right), \quad (12)$$

where BIC_{min} is the smallest BIC value within a set of models (SPEMD or composite), and h is a Gaussian centred on x with a variance of σ_{BIC}^2 . We follow the analytic calculation of Yıldırım et al. (2019) in evaluating the convolution integral in equation (12). Once we have a weighted $D_{\Delta t}$ distribution for the SPEMD models and another for the composite models, we combine these two with equal weight in the final inference.

4.5 Blind analysis

We perform our analysis blindly using a similar procedure as for previous H0LiCOW analyses (Suyu et al. 2013, H0LiCOW IV, IX). In practice, this is done by subtracting the median of certain parameter PDFs from the distribution when displaying and analysing results. In particular, we blind the values of $D_{\Delta t}$ or H_0 . This blinding procedure still allows us to measure their precision and relative offsets, as well as their correlation with other lens model parameters, but without knowing their values. Blinding also eliminates confirmation bias and the tendency for experimenters to stop their analysis when they obtain a value consistent with an ‘expected’ value, and forces us to be confident in our checks of systematic errors before finalizing our result. After completing our analysis, writing this paper draft with blinded $D_{\Delta t}$ distributions, and coming to an agreement among the coauthors to unblind the results on 2019 May 7, we unblinded and did not make any further changes to the models. There is also no iteration between the lens modelling and time-delay measurements. Throughout this paper, we show blinded $D_{\Delta t}$ distributions until Section 6, where we reveal the absolute $D_{\Delta t}$ values from our inference.

5 ESTIMATING THE EXTERNAL CONVERGENCE

We estimate the external convergence κ_{ext} using the weighted number counts technique, introduced in Greene et al. (2013) and Rusu et al. (2017, hereafter H0LiCOW III) and reframed as approximate Bayesian computation in H0LiCOW IX. For details of the numerical implementation, and a justification of the applicability in cosmography, we refer the reader to H0LiCOW III. Briefly, we use the catalogue of galaxies and associated physical properties (redshifts and stellar masses) around WFI2033 – 4723 from Section 3.3 and H0LiCOW X, as well as a control catalogue of the same properties from non-contiguous regions of the sky, large enough to overcome sample variance, from the Canada–France–Hawaii Telescope Lensing Survey (CFHTLenS; Heymans et al. 2012). We compute relative galaxy number counts within the matching limiting magnitude and

Table 2. Weighted galaxy count ratios $\bar{\zeta}_q$ for WFI2033 – 4723.

Weight q	45 arcsec	45 arcsec	45 arcsec	120 arcsec	120 arcsec	120 arcsec
	fiducial	fiducial + $z = 0.49$ group	fiducial + $z = 0.49, 0.66$ groups	fiducial	fiducial + $z = 0.49$ group	fiducial + $z = 0.49, 0.66$ groups
1	$1.44^{+0.06}_{-0.08}$	$1.20^{+0.03}_{-0.10}$	$1.17^{+0.04}_{-0.07}$	$1.55^{+0.07}_{-0.11}$	$1.34^{+0.05}_{-0.08}$	$1.27^{+0.04}_{-0.08}$
z	$1.70^{+0.09}_{-0.13}$	$1.36^{+0.07}_{-0.07}$	$1.34^{+0.07}_{-0.08}$	$1.65^{+0.12}_{-0.12}$	$1.37^{+0.09}_{-0.10}$	$1.31^{+0.08}_{-0.09}$
M_*	$1.45^{+0.11}_{-0.28}$	$0.78^{+0.10}_{-0.04}$	$0.80^{+0.10}_{-0.03}$	$2.39^{+0.11}_{-0.29}$	$1.68^{+0.30}_{-0.09}$	$1.48^{+0.27}_{-0.09}$
M_*^2	$1.42^{+0.26}_{-0.49}$	$0.51^{+0.12}_{-0.06}$	$0.54^{+0.13}_{-0.05}$	$3.69^{+0.43}_{-0.78}$	$2.12^{+0.76}_{-0.17}$	$1.75^{+0.64}_{-0.14}$
M_*^3	$1.40^{+0.45}_{-0.65}$	$0.33^{+0.13}_{-0.06}$	$0.37^{+0.14}_{-0.06}$	$5.75^{+0.95}_{-1.75}$	$2.73^{+1.53}_{-0.38}$	$2.12^{+1.18}_{-0.31}$
$1/r$	$1.33^{+0.01}_{-0.08}$	$1.07^{+0.04}_{-0.06}$	$1.04^{+0.04}_{-0.05}$	$1.55^{+0.07}_{-0.11}$	$1.35^{+0.05}_{-0.09}$	$1.25^{+0.05}_{-0.08}$
z/r	$1.49^{+0.03}_{-0.14}$	$1.17^{+0.06}_{-0.06}$	$1.14^{+0.06}_{-0.05}$	$1.59^{+0.09}_{-0.10}$	$1.33^{+0.06}_{-0.08}$	$1.24^{+0.05}_{-0.07}$
M_*/r	$1.69^{+0.34}_{-0.33}$	$0.69^{+0.11}_{-0.03}$	$0.75^{+0.10}_{-0.05}$	$2.08^{+0.21}_{-0.13}$	$1.53^{+0.29}_{-0.09}$	$1.38^{+0.27}_{-0.08}$
M_*^2/r	$1.97^{+0.73}_{-0.64}$	$0.49^{+0.10}_{-0.07}$	$0.56^{+0.11}_{-0.08}$	$3.24^{+0.52}_{-0.55}$	$1.89^{+0.85}_{-0.20}$	$1.52^{+0.78}_{-0.11}$
M_*^3/r	$2.07^{+0.93}_{-1.00}$	$0.32^{+0.12}_{-0.07}$	$0.38^{+0.14}_{-0.07}$	$5.30^{+1.00}_{-1.49}$	$2.53^{+1.78}_{-0.44}$	$1.86^{+1.42}_{-0.30}$
$M_{*,\text{rms}}^2$	$1.19^{+0.11}_{-0.23}$	$0.71^{+0.08}_{-0.04}$	$0.74^{+0.08}_{-0.04}$	$1.92^{+0.11}_{-0.21}$	$1.46^{+0.24}_{-0.06}$	$1.32^{+0.23}_{-0.05}$
$M_{*,\text{rms}}^3$	$1.12^{+0.11}_{-0.21}$	$0.69^{+0.08}_{-0.04}$	$0.72^{+0.08}_{-0.04}$	$1.79^{+0.09}_{-0.20}$	$1.40^{+0.22}_{-0.07}$	$1.28^{+0.21}_{-0.06}$
$M_{*,\text{rms}}^2/r$	$1.40^{+0.24}_{-0.25}$	$0.70^{+0.07}_{-0.06}$	$0.75^{+0.07}_{-0.06}$	$1.80^{+0.14}_{-0.16}$	$1.38^{+0.28}_{-0.08}$	$1.23^{+0.29}_{-0.04}$
$M_{*,\text{rms}}^3/r$	$1.27^{+0.17}_{-0.25}$	$0.68^{+0.08}_{-0.05}$	$0.73^{+0.07}_{-0.06}$	$1.74^{+0.11}_{-0.18}$	$1.36^{+0.27}_{-0.08}$	$1.23^{+0.26}_{-0.07}$
M_*/r^3	$1.13^{+0.26}_{-0.27}$	$0.46^{+0.05}_{-0.04}$	$0.48^{+0.05}_{-0.03}$	$1.86^{+0.12}_{-0.23}$	$1.37^{+0.14}_{-0.16}$	$1.25^{+0.10}_{-0.15}$
M_*/r^2	$1.41^{+0.41}_{-0.29}$	$0.54^{+0.09}_{-0.02}$	$0.58^{+0.10}_{-0.04}$	$2.04^{+0.14}_{-0.18}$	$1.47^{+0.11}_{-0.09}$	$1.31^{+0.11}_{-0.10}$
$\sqrt{M_*}/r$	$1.44^{+0.15}_{-0.19}$	$0.81^{+0.06}_{-0.03}$	$0.83^{+0.06}_{-0.04}$	$1.78^{+0.06}_{-0.11}$	$1.40^{+0.05}_{-0.06}$	$1.29^{+0.04}_{-0.07}$
$\sqrt{M_h}/r$	$1.57^{+0.29}_{-0.39}$	$1.08^{+0.57}_{-0.22}$	$1.07^{+0.69}_{-0.21}$	$1.81^{+0.09}_{-0.28}$	$1.39^{+0.19}_{-0.06}$	$1.27^{+0.18}_{-0.05}$

Note. Medians of weighted galaxy count ratios for WFI2033 – 4723, inside two different aperture radii and down to $i \leq 22.5$ mag. Weighted counts are themselves defined in terms of medians, following the third columns in table 4 of H0LiCOW III. The errors include, in quadrature, scatter from 10 samplings of redshift and stellar mass for each galaxy in the WFI2033 – 4723 field, scatter from the four disjoint CFHTLenS fields, and also from photometric redshifts measured with two different codes, as well as detections in the i or $i + r$ bands. See H0LiCOW X for details. The weighted counts are computed after removing from counting the galaxies G2, G3, and G7 (corresponding to the fiducial lensing model from Section 4.2; see Fig. 3), and alternatively, by removing in addition to these the galaxies part of the group at the lens redshift ($z = 0.66$), as well as the ones part of the groups at both $z = 0.66$ and $z = 0.49$ (see Section 4.2.3). See the text for details of the selection of group members without spectroscopic redshifts.

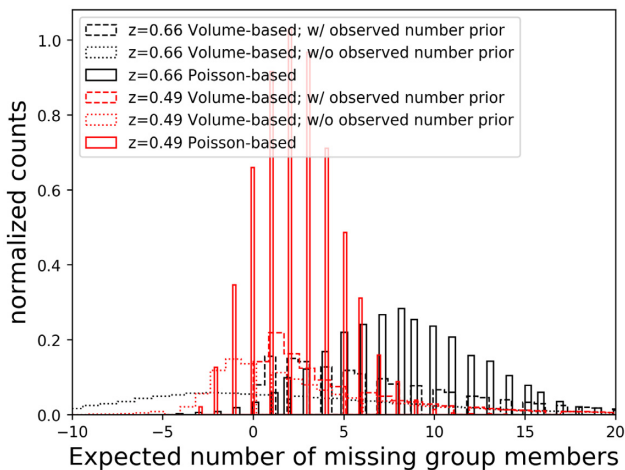


Figure 9. Estimated number of missing galaxy group members inside the ≤ 120 arcsec-radius from the lens system, due to spectroscopic incompleteness, computed with two methods, for the two galaxy groups. For the volume-based method, we plot the distribution both with and without imposing the prior knowledge of the number of galaxies which are spectroscopically confirmed to be part of the groups, which is equivalent to truncating the distributions below 0.

within the same apertures of 45 arcsec- and 120 arcsec-radii,⁸ using physically motivated weights ζ_q introduced by Greene et al. (2013) and H0LiCOW III. Here q stands for the redshift z , stellar mass M_* , the inverse of the distance r between each galaxy and the lens or the centre of the aperture, etc. A full list of the weights and of the corresponding measured relative number counts is shown in Table 2.

We perform the calculation above three times. These correspond, first, to the fiducial mass model in Section 4.2, which incorporates the nearby galaxies G2, G3, and G7 from Fig. 3. Since the effect of these galaxies has already been accounted for, we remove them from the input catalogue before running the computation. Second and third, we also remove one or both of the galaxy groups found in H0LiCOW X to impact the mass modelling beyond the tidal shear term, and therefore taken into account in the systematics tests presented in Section 4.2.3. However, since our spectroscopic completeness down to the limiting magnitude of $i < 22.5$ is only ~ 50 per cent, and

⁸In order to ensure a fair comparison of the number counts between the lens fields and the control field, in H0LiCOW X we performed detections in the same i -filter, where the images have similar seeing. To account for the coarser pixel scale in the DES data compared to CFHTLenS, we used more aggressive deblending parameters. This has no noticeable effect, with the exceptions of regions around bright stars. We ignore the negligible differences between the $ugriz$ filter curves in DES and CFHTLenS.

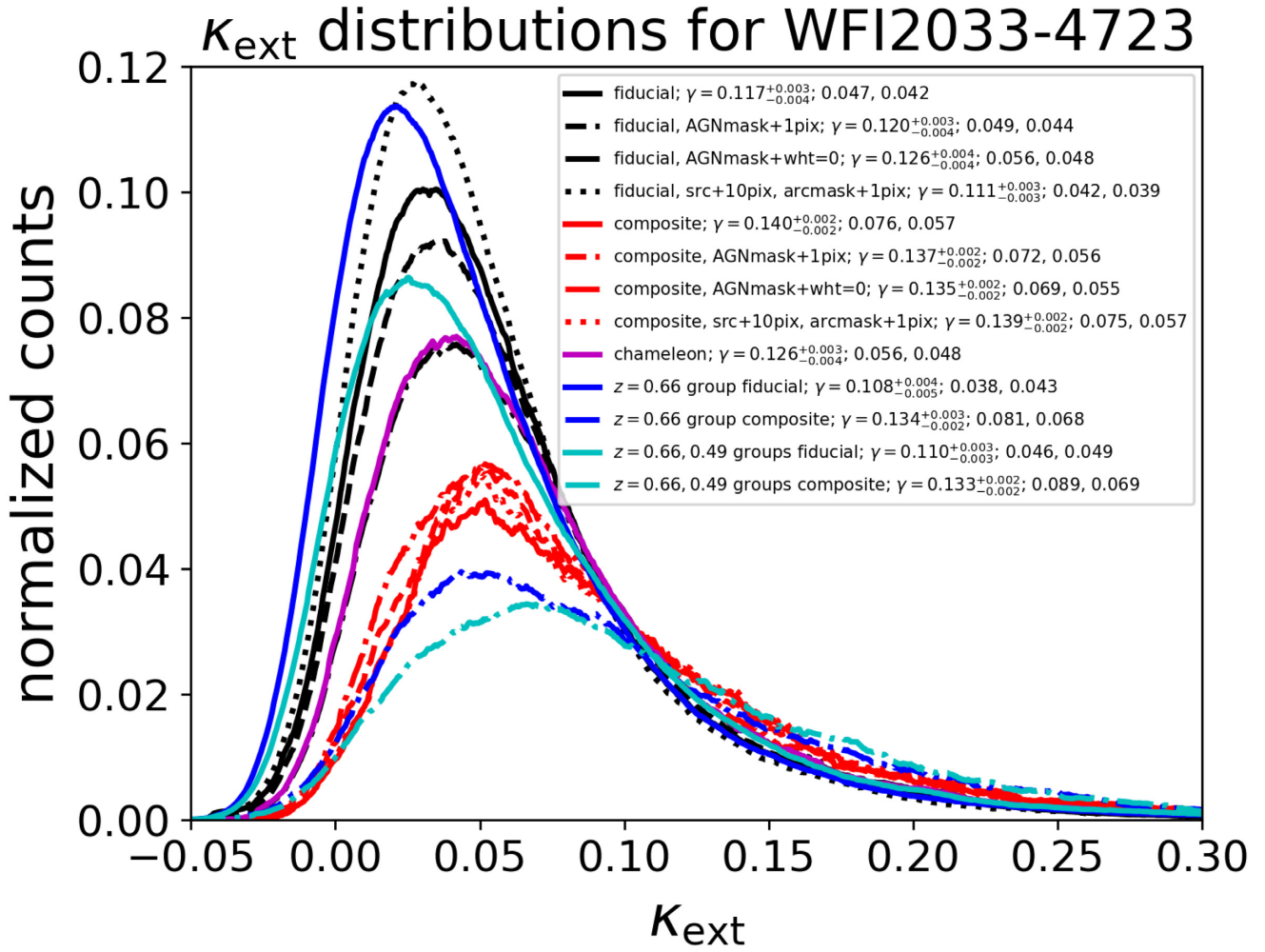


Figure 10. Distributions of $P(\kappa_{\text{ext}}|\zeta_1^{45''}, \zeta_{1/r}^{45''}, \zeta_1^{120''}, \zeta_{1/r}^{120''}, \gamma)$ for the various lensing models described in Section 4.2.3. The value of the corresponding external shear, which is a parameter of each lensing model and is used as constraint for the weighted number counts, is given in the legend. Following the shear values, the next two numbers in the legend are the pairs of (50 per centh, (84 per centh–16 per centh)/2) percentiles, which measure the median and the spread of the distributions. The size of the histogram bin is $\Delta\kappa_{\text{ext}} = 0.00055$. As the original distributions are noisy, we plot their convolution with a large smoothing window of length $50 \times \Delta\kappa_{\text{ext}}$.

also non-uniform, decreasing with radius from the lens (see Fig. 4 in H0LiCOW X), it is likely that there are other galaxies part of these groups, in addition to the ones spectroscopically confirmed. If we were to keep these galaxies in the number counts, our inferred κ_{ext} would be an overestimate, when coupled to the models from Section 4.2.3 which already include these galaxy groups. We use two different methods to account for these galaxies statistically. Briefly, in the first method we use the measured spectroscopic completeness and the total number of galaxies within the 120 arcsec-radius aperture, as well as the number of confirmed group members inside the same aperture, and we apply Poisson statistics to infer the distribution of galaxy numbers we miss due to spectroscopic incompleteness. In the second method, we use the velocity dispersions of the two groups measured in H0LiCOW X, as well as the virial radii from Wilson et al. (2016), and we calculate the expected number of galaxies inside the virial radius, from the empirical relation in Andreon & Hurn (2010). Then, based on the projected distance between the group centroid and the lens measured in H0LiCOW X, we estimate the expected number of galaxies at the intersection of the sphere of virial radius and the 120 arcsec-radius cylinder centred on the lens. Subtracting from this the number of galaxy

members spectroscopically confirmed, we arrive at a distribution of the number of missing galaxies. We show the resulting distributions from both methods in Fig. 9. For each group, the distributions from both methods overlap significantly, giving consistent results. The expected median number of galaxies missing from the group at $z = 0.66$ is 6–8, and from the group at $z = 0.49$ it is 3. Finally, we extract at random, 20 times, a number from these distributions, and remove these galaxies, picked at random from within our catalogue of galaxies around the lens, with photometric redshifts compatible with the group redshifts, before computing the weighted number counts. The resulting scatter is included in the values reported in Table 2.

Our measured relative weighted number counts, in particular for the fiducial model and inside the 120 arcsec-radius aperture, show that the field of WFI2033 – 4723 is overdense. This was also remarked in the study by Fassnacht, Koopmans & Wong (2011), where a non-weighted number count overdensity of 1.33 was obtained. While a direct comparison with this values would be biased because of the different limiting magnitude, detection filter, and the fact that Fassnacht et al. (2011) count all nearby galaxies without exception, our ~ 1.44 unweighted relative number count

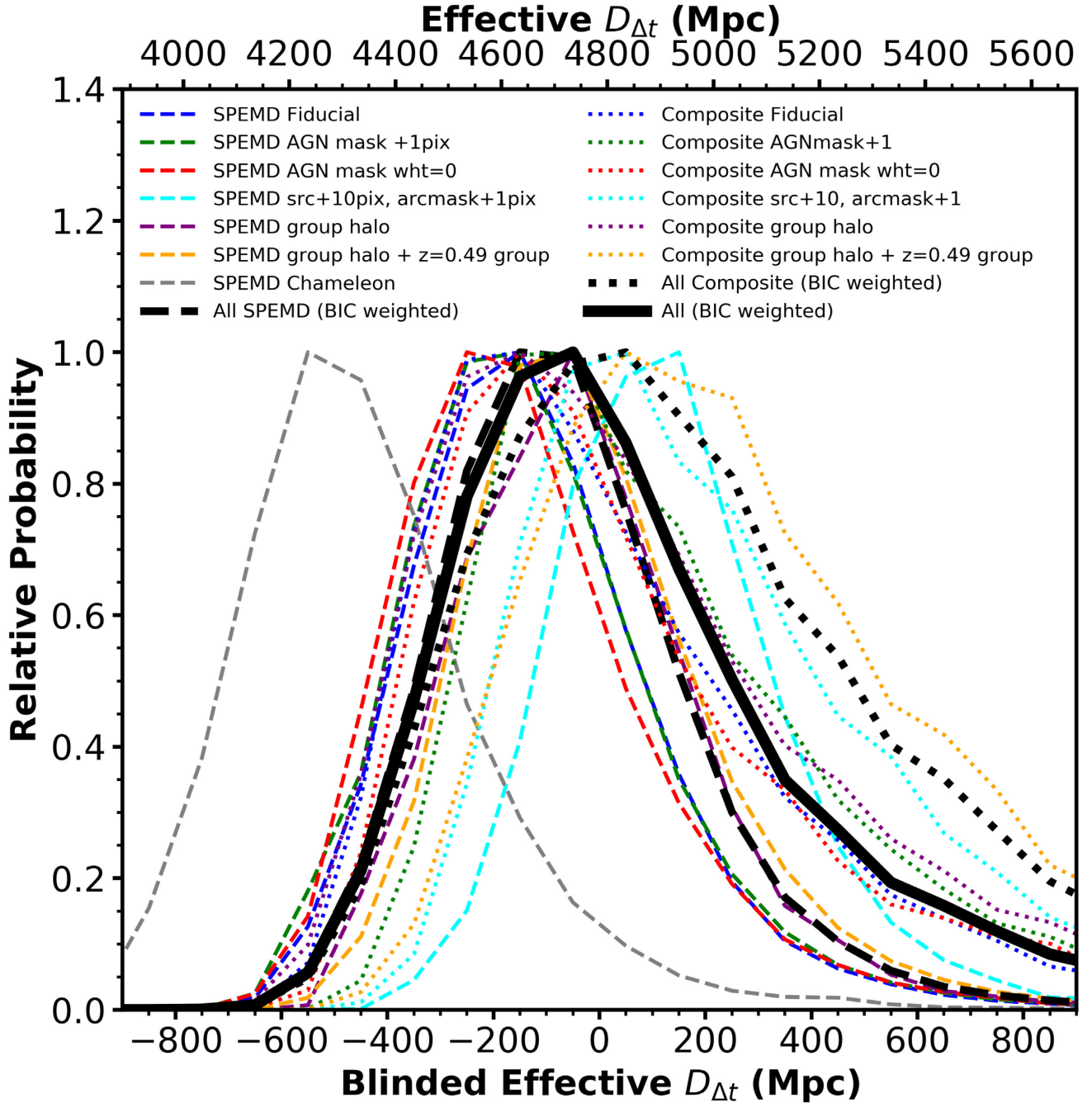


Figure 11. PDF of $D_{\Delta t}$ for WFI2033 – 4723. The SPEMD and composite models are weighted by BIC, then are each given equal weight in the final inference.

inside the same aperture of 45 arcsec-radius is consistent, within 1.5σ .

To convert the measured relative weighted number counts into a κ_{ext} distribution based on these constraints, we follow Suyu et al. (2010, 2013); Greene et al. (2013); H0LiCOW III, IX; Chen et al. (2019), and use the results of ray-tracing by Hilbert et al. (2009) through the Millennium Simulation (MS; Springel et al. 2005), in the form of a convergence and shear map (κ , γ) covering each simulated sky location. Our technique is justified by the results of Suyu et al. (2010), Hilbert et al. (2009), which showed that the distribution of κ for LOS containing strong lenses is very similar to that over all LOS. With a catalogue of galaxies painted on top of the dark matter haloes from the MS, following the semi-

analytical models in De Lucia & Blaizot (2007), and containing realistic simulated photometry, we follow a similar procedure and compute relative weighted number counts at each spatial location throughout the MS (see H0LiCOW III for details). Finally, we compute

$$\begin{aligned}
 P(\kappa_{\text{ext}}|\mathbf{d}_{\text{LOS}}) &\equiv P(\kappa_{\text{ext}}|\zeta_q, \dots) \\
 &= \int \prod_q d\zeta_q P_{\text{MS}}(\kappa_{\text{ext}}|\zeta_q^{\text{MS}} \equiv \zeta_q, \dots) P(\zeta_q, \dots|\mathbf{d}_{\text{LOS}}), \quad (13)
 \end{aligned}$$

where we combine multiple weighted number count constraints ζ_q , including from both 45 arcsec- and 120 arcsec-radius apertures. Following H0LiCOW III, we treat the external shear γ_{ext} computed

Table 3. Effective time-delay distance and BIC weighting for different lens models.

Model	$D_{\Delta t}$ (Mpc)	ΔBIC	BIC weight
SPEMD fiducial	4640^{+238}_{-195}	17	0.674
SPEMD AGN mask + 1pix	4631^{+247}_{-199}	49	0.235
SPEMD AGN mask weight = 0	4611^{+258}_{-195}	1984	0.000
SPEMD arcmask + 1pix, 60×60 source	4894^{+219}_{-190}	306	0.000
SPEMD lens group halo	4732^{+228}_{-203}	16	0.686
SPEMD lens + $z = 0.49$ group halo	4740^{+251}_{-190}	0	1.000
SPEMD chameleon light profiles	4312^{+241}_{-185}	1962	0.000
All SPEMD (BIC-weighted)	4703^{+245}_{-203}	–	–
Composite fiducial	4731^{+416}_{-246}	202	0.000
Composite AGN mask + 1pix	4836^{+418}_{-235}	206	0.000
Composite AGN mask weight = 0	4743^{+417}_{-238}	637	0.000
Composite arcmask + 1pix, 60×60 source	4944^{+437}_{-258}	559	0.000
Composite lens group halo	4779^{+475}_{-283}	16	0.770
Composite lens + $z = 0.49$ group halo	5009^{+471}_{-305}	0	1.000
All Composite (BIC-weighted)	4913^{+489}_{-320}	–	–
All	4784^{+399}_{-248}	–	–

Note. Reported values are medians, with errors corresponding to the 16th and 84th percentiles.

from the lens models in Section 4 at the location of the lens (in the case of the MS, at the centre of each aperture) analogously to the weighted number count constraints.⁹ In Section E we explore various combinations of constraints, show that our technique is free of biases, and describe several tests we ran. We settle on the combination of $P(\kappa_{\text{ext}}|\zeta_{1'}^{45''}, \zeta_{1'}^{45''}, \zeta_{1'}^{120''}, \zeta_{1'}^{120''}, \gamma)$, which employs our most robust constraints. In Fig. 10 we show the resulting distributions, corresponding to the various mass models explored in Section 4, and their associated shear values.¹⁰

6 RESULTS

After conducting the analysis described in Sections 4 and 5, and combining with the time delays from Section 3.4, we plot the final BIC-weighted $D_{\Delta t}$ distributions in Fig. 11, with the blinded values shown on the bottom axis and the unblinded values shown on the top axis. We report the median and 68 percent quantiles of $D_{\Delta t}$ for each of the models Table 3, along with the ΔBIC and associated weighting that each model receives. Our constraint on $D_{\Delta t}$ for WFI2033 – 4723 is $D_{\Delta t} = 4784^{+399}_{-248}$ Mpc, a ~ 6.6 per cent precision measurement.

From this inferred $D_{\Delta t}$, we can calculate cosmological parameters for flat ΛCDM or other cosmologies. For flat ΛCDM with

⁹Here and in Chen et al. (2019) we modify the way we implement the γ_{ext} constraint described in H0LiCOW III, in the sense that we no longer normalize by the number of LOS in each small division of the constraint range. This is because the shear values derived in Section 4 use a flat prior, and the distribution of γ in the MS maps naturally introduces a cosmological prior. This effect is negligible, except for the case of very large γ_{ext} uncertainties.

¹⁰It is unexpected that the two distributions of the composite model, which correspond to the case where one or both galaxy groups are explicitly modelled, have larger medians than the standard composite model (by $\lesssim 0.2\sigma$ or at $\lesssim 1$ per cent level), even though they are constrained by smaller values of shear and weighted counts. We attribute this to noise, and we have checked that the excess is consistent with variations between similar distributions for these models, employing a different choice of weighted count constraints.

uniform priors on H_0 (within the range $[0, 150] \text{ km s}^{-1} \text{ Mpc}^{-1}$) and Ω_m (within the range $[0.05, 0.5]$), this translates into a constraint on the Hubble constant of $H_0 = 71.6^{+3.8}_{-4.9} \text{ km s}^{-1} \text{ Mpc}^{-1}$. Within the uncertainties, our result is consistent with the previous measurements of H_0 from H0LiCOW. After B1608 + 656 (Suyu et al. 2010), this is the H0LiCOW lens producing the second smallest uncertainty on H_0 , comparable to what is expected for the lensed supernovae ‘Refsdal’ (Grillo et al. 2018). We note that weighting all of our models equally and ignoring the BIC weighting (as was done with some previous H0LiCOW lenses) gives $H_0 = 72.2^{+4.3}_{-4.8} \text{ km s}^{-1} \text{ Mpc}^{-1}$, which is a shift of < 1 per cent in the median value. Our result can be combined with the other five lenses in the H0LiCOW sample to give stronger constraints for a variety of cosmological models. The full cosmological analysis is presented in Wong et al. (2019). Our measurement is also consistent with recent ones from other techniques, not only from *Planck* and SH0ES, but also from the Carnegie Supernova Project (Burns et al. 2018), the Megamaser Cosmology project (Braatz et al. 2018), DES clustering, and weak lensing + baryon acoustic oscillations + big bang nucleosynthesis experiments (Abbott et al. 2018), the inverse distance ladder (e.g. Aubourg et al. 2015; Macaulay et al. 2019), extragalactic background light attenuation (e.g. Domínguez et al. 2019), etc.

We note that WFI2033 – 4723 has been used in the past to measure H_0 . Based on three years of monitoring and constrained by the relative quasar image positions measured from archival *HST* imaging, Vuissoz et al. (2008) infer $H_0 = 67^{+13}_{-10} \text{ km s}^{-1} \text{ Mpc}^{-1}$, using non-parametric modelling, in good agreement with our result, but with significantly larger uncertainties.

7 CONCLUSIONS

We have analysed the gravitational lens WFI2033 – 4723, performing a blind cosmographic analysis to determine the time-delay distance of this system. We use deep *HST* imaging, precise time-delay measurements, a measurement of the lens galaxy’s velocity dispersion, and deep wide-area spectroscopic and photometric data

to constrain the mass distribution along the LOS. By accurately modelling the lens and accounting for systematic uncertainties, we constrain the effective time-delay distance to be $D_{\Delta t} = 4784^{+399}_{-248}$ Mpc, a precision of 6.6 per cent. This translates to a Hubble constant of $H_0 = 71.6^{+3.8}_{-4.9}$ km s⁻¹ Mpc⁻¹ in a flat Λ CDM cosmology with uniform priors on H_0 and Ω_m , a value consistent with measurements from other techniques, as well as previous H0LiCOW lenses. A joint analysis of all six H0LiCOW lenses and our constraints on different cosmologies is presented in Wong et al. (2019).

As with all galaxy-scale lenses where time-delay cosmography is applied, we expect that our results can be improved with future, higher resolution adaptive optics imaging (Chen et al. 2016), spatially resolved kinematics (Shajib et al. 2018), and a more tailored, non-statistical treatment of the external convergence (McCully et al. 2017).

ACKNOWLEDGEMENTS

We thank Adriano Agnello, Roger Blandford, Xuheng Ding, Georges Meylan, Danka Paraficz, Chiara Spiniello, Malte Tewes, and Olga Tihonova for their contributions to the H0LiCOW project. H0LiCOW and COSMOGRAIL are made possible thanks to the continuous work of all observers and technical staff obtaining the monitoring observations, in particular at the Swiss Euler telescope at La Silla Observatory. Euler is supported by the Swiss National Science Foundation. CER and CDF are funded through the National Science Foundation (NSF) grant AST-1312329, ‘Collaborative Research: Accurate cosmology with strong gravitational lens time delays,’ and the HST grant GO-12889. This work was supported by World Premier International Research Center Initiative (WPI Initiative), Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. KCW is supported in part by an EACOA Fellowship awarded by the East Asia Core Observatories Association, which consists of the Academia Sinica Institute of Astronomy and Astrophysics, the National Astronomical Observatory of Japan, the National Astronomical Observatories of the Chinese Academy of Sciences, and the Korea Astronomy and Space Science Institute. VB and FC acknowledge support from the Swiss National Science Foundation (SNSF) and through European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (COSMICLENS: grant agreement No 787866) DS acknowledges funding support from a *Back to Belgium* grant from the Belgian Federal Science Policy (BELSPO). SHS thanks the Max Planck Society for support through the Max Planck Research Group. This research was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Research, Innovation and Science. SH acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) cluster of excellence ‘Origin and Structure of the Universe’ (www.universe-cluster.de). CDF and GC-FC acknowledge support for this work from the NSF under Grant No. AST-1715611. TT thanks the Packard Foundation for generous support through a Packard Research Fellowship, the NSF for funding through NSF grant AST-1450141, ‘Collaborative Research: Accurate cosmology with strong gravitational lens time delays’. GC-FC acknowledges support from the Ministry of Education in Taiwan via Government Scholarship to Study Abroad (GSSA). LVEK is supported in part through a Dutch Research Council NWO-VICI career grant (project number 639.043.308). PJM acknowledges support from the U.S. Department of Energy under contract number DE-AC02-76SF00515.

This paper is based on observations made with the National Aeronautics and Space Administration (NASA)/European Space Agency (ESA) *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with Programs #12889 and #9744. Support for program #12889 was provided by NASA through a grant 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.

This work is partly based on the analysis in Sluse et al. (2019), which used public archival data from the Dark Energy Survey (DES). Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l’Espai (IEEC/CSIC), the Institut de Física d’Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

This work is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l’Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. CFHTLenS data processing was made possible thanks to significant computing support from the NSERC Research Tools and Instruments grant program.

Data analysis was in part carried out on common use data analysis computer system at the Astronomy Data Center, ADC, of

the National Astronomical Observatory of Japan. This work made use of `astropy`, a community-developed core Python package for Astronomy (Astropy Collaboration 2013, 2018). Plots were produced with `matplotlib` (Hunter 2007).

REFERENCES

- Abbott T. M. C. et al. 2018, *ApJS*, 239, 18
- Andreon S., Hurn M. A., 2010, *MNRAS*, 404, 1922
- Appenzeller I. et al., 1998, *Messenger*, 94, 1
- Astropy Collaboration et al., 2013, *A&A*, 558, A33
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Aubourg É. et al., 2015, *Phys. Rev. D*, 92, 123516
- Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, *ApJ*, 724, 511
- Bacon R. et al., 2010, in Proc. SPIE Conf. Ser. Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. SPIE, Bellingham, p. 773508
- Barkana R., 1998, *ApJ*, 502, 531
- Birrer S., Amara A., 2018, *Phys. Dark Universe*, 22, 189
- Birrer S., Treu T., 2019, *MNRAS*, 489, 2097
- Birrer S., Amara A., Refregier A., 2015, *ApJ*, 813, 102
- Birrer S. et al., 2019, *MNRAS*, 484, 4726
- Blandford R., Narayan R., 1986, *ApJ*, 310, 568
- Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Moustakas L. A., 2006, *ApJ*, 638, 703
- Bonvin V., Tewes M., Courbin F., Kuntzer T., Sluse D., Meylan G., 2016, *A&A*, 585, A88
- Bonvin V. et al., 2018, *A&A*, 616, A183
- Bonvin V., Tihhonova O., Millon M., Chan J. H.-H., Savary E., Huber S., Courbin F., 2019a, *A&A*, 621, A55
- Bonvin V. et al., 2019b, *A&A*, 629, A97
- Braatz J. et al., 2018, in Tarchi A., Reid M. J., Castangia P., eds, Proc. IAU Symp. 336, Astrophysical Masers: Unlocking the Mysteries of the Universe. Cambridge University Press, Cambridge, p. 86
- Burns C. R. et al., 2018, *ApJ*, 869, 56
- Cackett E. M., Horne K., Winkler H., 2007, *MNRAS*, 380, 669
- Cantale N., Courbin F., Tewes M., Jablonka P., Meylan G., 2016, *A&A*, 589, A81
- Chen G. C.-F. et al., 2016, *MNRAS*, 462, 3457
- Chen G. C.-F. et al., 2018a, *MNRAS*, 481, 1115
- Chen G. C.-F. et al., 2019, *MNRAS*, 490, 1743
- Chen H.-Y., Fishbach M., Holz D. E., 2018b, *Nature*, 562, 545
- Collett T. E., Auger M. W., 2014, *MNRAS*, 443, 969
- Collett T. E., Cunningham S. D., 2016, *MNRAS*, 462, 3255
- Collett T. E. et al., 2013, *MNRAS*, 432, 679
- Courbin F., Eigenbrod A., Vuissoz C., Meylan G., Magain P., 2005, in Mellier Y., Meylan G., eds, Proc. IAU Symp. 225, Gravitational Lensing Impact on Cosmology. Cambridge University Press, Cambridge, p. 297
- Courbin F. et al., 2011, *A&A*, 536, A53
- De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
- de Vaucouleurs G., 1948, *Ann. Astrophys.*, 11, 247
- Di Valentino E., Linder E. V., Melchiorri A., 2018, *Phys. Rev. D*, 97, 043528
- Diemer B., 2018, *ApJS*, 239, 35
- Ding X. et al., 2018, preprint (arXiv:1801.01506)
- Domínguez A. et al., 2019, preprint (arXiv:1903.12097)
- Dutton A. A., Treu T., 2014, *MNRAS*, 438, 3594
- Dutton A. A. et al., 2011, *MNRAS*, 417, 1621
- Eigenbrod A., Courbin F., Vuissoz C., Meylan G., Saha P., Dye S., 2005, *A&A*, 436, 25
- Eigenbrod A., Courbin F., Meylan G., Vuissoz C., Magain P., 2006, *A&A*, 451, 759
- Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, *ApJ*, 289, L1
- Fassnacht C. D., Pearson T. J., Readhead A. C. S., Browne I. W. A., Koopmans L. V. E., Myers S. T., Wilkinson P. N., 1999, *ApJ*, 527, 498
- Fassnacht C. D., Xanthopoulos E., Koopmans L. V. E., Rusin D., 2002, *ApJ*, 581, 823
- Fassnacht C. D., Gal R. R., Lubin L. M., McKean J. P., Squires G. K., Readhead A. C. S., 2006, *ApJ*, 642, 30
- Fassnacht C. D., Koopmans L. V. E., Wong K. C., 2011, *MNRAS*, 410, 2167
- Fazio G. G. et al., 2004, *ApJS*, 154, 10
- Feeney S. M., Peiris H. V., Williamson A. R., Nissanke S. M., Mortlock D. J., Alsing J., Scolnic D., 2019, *Phys. Rev. Lett.*, 122, 061105
- Flaugher B. et al., 2015, *AJ*, 150, 150
- Gao F. et al., 2016, *ApJ*, 817, 128
- Gavazzi R., Treu T., Rhodes J. D., Koopmans L. V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, *ApJ*, 667, 176
- Gavazzi R., Treu T., Koopmans L. V. E., Bolton A. S., Moustakas L. A., Burles S., Marshall P. J., 2008, *ApJ*, 677, 1046
- Golse G., Kneib J.-P., 2002, *A&A*, 390, 821
- Gorenstein M. V., Shapiro I. I., Falco E. E., 1988, *ApJ*, 327, 693
- Greene Z. S. et al., 2013, *ApJ*, 768, 39
- Grillo C. et al., 2018, *ApJ*, 860, 94
- Hernquist L., 1990, *ApJ*, 356, 359
- Heymans C. et al., 2012, *MNRAS*, 427, 146
- Hilbert S., Hartlap J., White S. D. M., Schneider P., 2009, *A&A*, 499, 31
- Hook I. M., Jørgensen I., Allington-Smith J. R., Davies R. L., Metcalfe N., Murowinski R. G., Crampton D., 2004, *PASP*, 116, 425
- Hu W., 2005, in Wolff S. C., Lauer T. R., eds, ASP Conf. Ser. Vol. 339, Observing Dark Energy. Astron. Soc. Pac., San Francisco, p. 215
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jaffe W., 1983, *MNRAS*, 202, 995
- Kassiola A., Kovner I., 1993, *ApJ*, 417, 450
- Keeton C. R., 2003, *ApJ*, 584, 664
- Keeton C. R., Kochanek C. S., 1997, *ApJ*, 487, 42
- Kissler-Patig M. et al., 2008, *A&A*, 491, 941
- Kochanek C. S., 2002, *ApJ*, 578, 25
- Kochanek C. S., Apostolakis J., 1988, *MNRAS*, 235, 1073
- Kochanek C. S., Morgan N. D., Falco E. E., McLeod B. A., Winn J. N., Dembicky J., Ketzbeck B., 2006, *ApJ*, 640, 47
- Koopmans L. V. E., 2004, preprint (arXiv:astro-ph/0412596)
- Koopmans L. V. E., Treu T., Fassnacht C. D., Blandford R. D., Surpi G., 2003, *ApJ*, 599, 70
- Kovner I., 1987, *ApJ*, 316, 52
- Kreisich C. D., Cyr-Racine F.-Y., Doré O., 2019, preprint (arXiv:1902.00534)
- Lemon C. A., Auger M. W., McMahon R. G., 2019, *MNRAS*, 483, 4242
- Lewis A., Bridle S., 2002, *Phys. Rev. D*, 66, 103511
- Liao K. et al., 2015, *ApJ*, 800, 11
- Macaulay E. et al., 2019, *MNRAS*, 486, 2184
- Magain P., Courbin F., Sohy S., 1998, *ApJ*, 494, 472
- Mamon G. A., Łokas E. L., 2005, *MNRAS*, 363, 705
- McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2014, *MNRAS*, 443, 3631
- McCully C., Keeton C. R., Wong K. C., Zabludoff A. I., 2017, *ApJ*, 836, 141
- Merritt D., 1985, *AJ*, 90, 1027
- Moffat A. F. J., 1969, *A&A*, 3, 455
- Momcheva I., Williams K., Keeton C., Zabludoff A., 2006, *ApJ*, 641, 169
- Momcheva I. G., Williams K. A., Cool R. J., Keeton C. R., Zabludoff A. I., 2015, *ApJS*, 219, 29
- Morgan N. D., Caldwell J. A. R., Schechter P. L., Dressler A., Egami E., Rix H.-W., 2004, *AJ*, 127, 2617
- Morgan C. W., Hyer G. E., Bonvin V., Mosquera A. M., Cornachione M., Courbin F., Kochanek C. S., Falco E. E., 2018, *ApJ*, 869, 106
- Motta V., Mediavilla E., Rojas K., Falco E. E., Jiménez-Vicente J., Muñoz J. A., 2017, *ApJ*, 835, 132
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Oguri M., 2007, *ApJ*, 660, 1
- Osipkov L. P., 1979, *Pis ma Astron. Zhurnal*, 5, 77
- Petters A. O., Levine H., Wambsgans J., 2001, *Singularity Theory and Gravitational Lensing*. Birkhauser, Basel

- Pirard J.-F. et al., 2004, in Moorwood A. F. M., Iye M., eds, Proc. SPIE Conf. Ser. Vol. 5492, Ground-based Instrumentation for Astronomy. SPIE, Bellingham, p. 1763. Available at: <http://adsabs.harvard.edu/abs/2004SPIE.5492.1763P>
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Planck Collaboration VI, 2018, preprint ([arXiv:1807.06209](https://arxiv.org/abs/1807.06209))
- Poulin V., Smith T. L., Karwal T., Kamionkowski M., 2019, *Phys. Rev. Lett.*, 122, 221301
- Refsdal S., 1964, *MNRAS*, 128, 307
- Richard J. R. B. J. V., 2017, MUSE Pipeline User Manual, VLT-MAN-ESO-261650, Issue 7.0. ESO, 7.0 edn. Available at: https://www.eso.org/sci/activities/vltsv/muse/ESO-261650_7_MUSE_User_Manual.pdf
- Riess A. G. et al., 2016, *ApJ*, 826, 56
- Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *ApJ*, 876, 85
- Rusu C. E. et al., 2017, *MNRAS*, 467, 4220
- Saha P., 2000, *AJ*, 120, 1654
- Schechter P. L. et al., 1997, *ApJ*, 475, L85
- Schneider P., 2014, *A&A*, 568, L2
- Schneider P., Sluse D., 2013, *A&A*, 559, A37
- Schneider P., Sluse D., 2014, *A&A*, 564, A103
- Schneider P., Ehlers J., Falco E. E., 1992, *Gravitational Lenses*. Springer, Switzerland
- Seljak U., 1994, *ApJ*, 436, 509
- Shajib A. J., Treu T., Agnello A., 2018, *MNRAS*, 473, 210
- Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
- Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, *A&A*, 544, A62
- Sluse D. et al., 2019, *MNRAS*, 490, 613
- Springel V. et al., 2005, *Nature*, 435, 629
- Starkey D. A., Horne K., Villforth C., 2016, *MNRAS*, 456, 1960
- 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., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, *ApJ*, 711, 201
- Suyu S. H. et al., 2012a, preprint ([arXiv:1202.4459](https://arxiv.org/abs/1202.4459))
- Suyu S. H. et al., 2012b, *ApJ*, 750, 10
- Suyu S. H. et al., 2013, *ApJ*, 766, 70
- Suyu S. H. et al., 2014, *ApJ*, 788, L35
- Suyu S. H. et al., 2017, *MNRAS*, 468, 2590
- Tewes M., Courbin F., Meylan G., 2013, *A&A*, 553, A120
- Tie S. S., Kochanek C. S., 2018, *MNRAS*, 473, 80
- Tihhonova O. et al., 2018, *MNRAS*, 477, 5657
- Treu T., Koopmans L. V. E., 2002, *ApJ*, 575, 87
- Treu T., Marshall P. J., 2016, *A&A Rev.*, 24, 11
- Unruh S., Schneider P., Sluse D., 2017, *A&A*, 601, A77
- Vanderriest C., Schneider J., Herpe G., Chevretton M., Moles M., Wlerick G., 1989, *A&A*, 215, 1
- Vattis K., Koushiappas S. M., Loeb A., 2019, *PhysRevD*, 99, 121302
- Vuissoz C. et al., 2008, *A&A*, 488, 481
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87
- Williams K. A., Momcheva I., Keeton C. R., Zabludoff A. I., Lehár J., 2006, *ApJ*, 646, 85
- Wilson M. L., Zabludoff A. I., Ammons S. M., Momcheva I. G., Williams K. A., Keeton C. R., 2016, *ApJ*, 833, 194
- Wong K. C., Keeton C. R., Williams K. A., Momcheva I. G., Zabludoff A. I., 2011, *ApJ*, 726, 84
- Wong K. C. et al., 2017, *MNRAS*, 465, 4895
- Wong K. C. et al., 2019, preprint ([arXiv:1907.04869](https://arxiv.org/abs/1907.04869))
- Wucknitz O., 2002, *MNRAS*, 332, 951
- Yıldırım A., Suyu S. H., Halkola A., 2019, preprint ([arXiv:1904.07237](https://arxiv.org/abs/1904.07237))

APPENDIX A: RESIDUALS WITHOUT POWER-LAW WEIGHTING

For completeness, we show the normalized residuals for the fiducial SPEMD model (Fig. A1) and fiducial composite model (Fig. A2) using the weight images without the power-law weighting in the region near the AGN images. We see that there are strong residuals due to the AGN images, which motivates our downweighting of these regions.

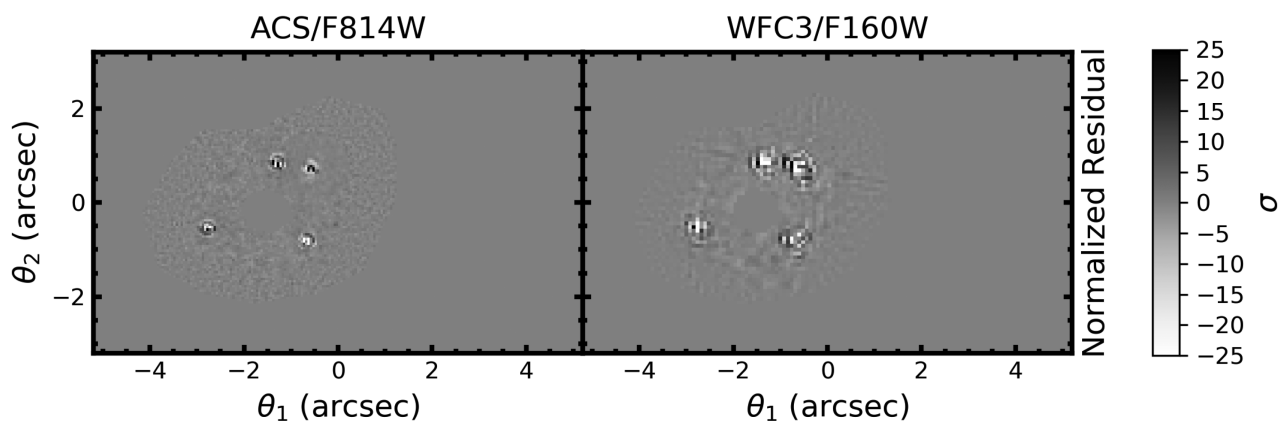


Figure A1. Normalized residual for the fiducial SPEMD model without power-law weighting.

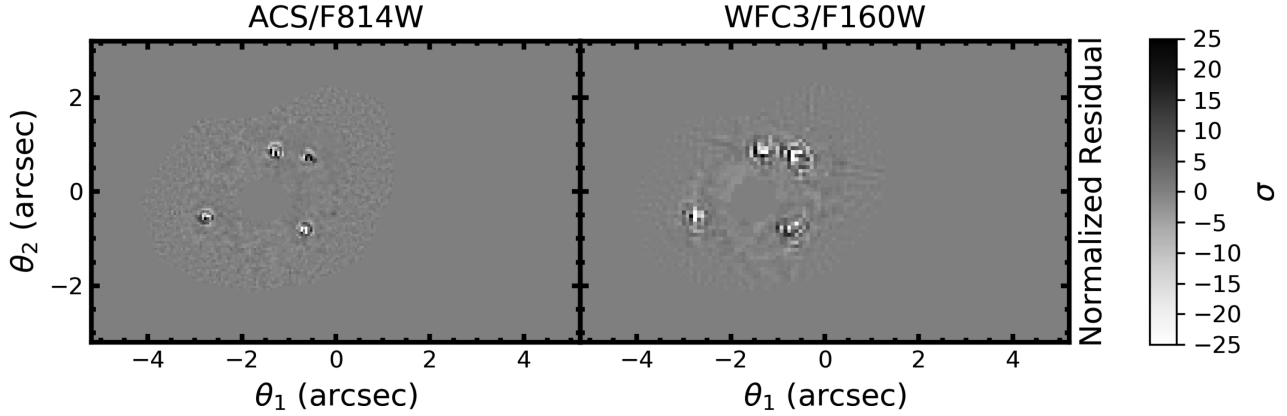


Figure A2. Same as Fig. A1 but for the fiducial composite model.

APPENDIX B: IMPACT OF DIFFERENT COSMOLOGIES

In multilens plane modelling, we need to sample the cosmological parameters in order to carry out the ray tracing. For computational reasons, we directly vary H_0 but keep other cosmological parameters fixed ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$). $D_{\Delta t}$ has a weak dependence on these other parameters in principle, but we found that varying these parameters affected the posterior $D_{\Delta t}$ distribution by < 1 per cent for HE 0435 – 1223 (H0LiCOW IV). We perform a similar check for WFI2033 – 4723 in which we run the fiducial SPEMD model while allowing either Ω_m to vary, or allowing both Ω_m and w to vary. The resulting effective $D_{\Delta t}$ distributions are shown in Fig. B1. The peaks of the distribution are consistent to within 1 per cent of the absolute value, which shows that the results are insensitive to these extra cosmological parameters at the level of accuracy that we are currently working at, similar to HE 0435 – 1223.

In the case of multiple lens planes (e.g. Blandford & Narayan 1986; Kovner 1987; Kochanek & Apostolakis 1988; Schneider et al. 1992; Petters et al. 2001; Collett & Auger 2014; McCully et al. 2014; Schneider 2014), there is not a unique time-delay distance for the system, but instead multiple time-delay distances between planes i and j ,

$$D_{\Delta t}^{ij} \equiv (1 + z_i) \frac{D_i D_j}{D_{ij}}, \quad (\text{B1})$$

with z_i being the redshift of plane i . The multiplane time delay is given by

$$t = \sum_{i=1}^{s-1} \frac{D_{\Delta t}^{i,i+1}}{c} \left[\frac{(\theta_{i+1} - \theta_i)^2}{2} - \beta_{i,i+1} \psi_i(\theta_i) \right], \quad (\text{B2})$$

where s is the index of the source plane (starting with $i = 1$ as the lowest redshift lens plane and counting up towards the source plane). ψ_i is the lens potential related to the scaled deflection angle via $\nabla \psi_i = \alpha_i$, and β is a ratio of angular diameter distances among deflector planes and the source,

$$\beta_{ij} = \frac{D_{ij} D_s}{D_j D_{is}}. \quad (\text{B3})$$

From equation (B2), we see that the time delay depends on the multiple time-delay distances and β terms. In general, it is difficult to constrain all of these quantities independently, so we adopt specific cosmological models to compute the distances for the ray tracing, then compare the time-delay distance measurements from these different background cosmologies. For the case of WFI2033 – 4723,

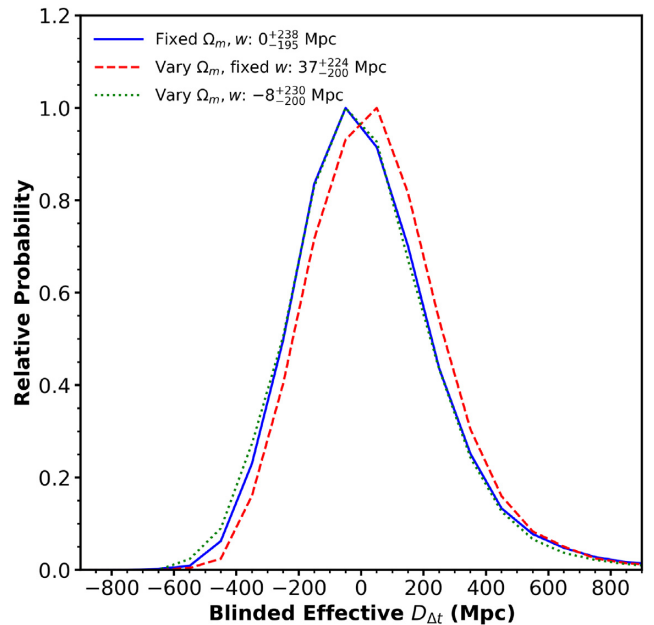


Figure B1. PDF of $D_{\Delta t}$ for the various cosmologies. We compare the fiducial SPEMD model to one in which Ω_m is allowed to vary (with $\Omega_m + \Omega_\Lambda = 1$), and one in which w is also allowed to vary. The distributions are blinded by subtracting the median of the fiducial SPEMD model PDF. The different cosmology tests are indicated by the legend, and the median and 68 per cent quantiles of the $D_{\Delta t}$ distributions are given. The median of the blinded effective time-delay distance PDF is insensitive to the extra cosmological parameters to within 1 per cent.

where G2 is at a different redshift from the main lens plane and is not strongly lensing the background source but merely perturbing it, the effect on the time delays is weak. The lack of sensitivity to Ω_m and w seen in Fig. B1 suggests that WFI2033 – 4723 is not sensitive to the extra β terms at an interesting level to probe it directly in the same way as a double source plane lens (e.g. Gavazzi et al. 2008; Collett & Auger 2014). Since the time delays are mostly set by the main lens plane, we can measure the ‘effective’ $D_{\Delta t}$ (which is $D_{\Delta t}^{is}$ with i as the main lens plane and s as the source plane) that is independent of assumptions on the background cosmology. This robust distance determination then permits us to constrain any reasonable cosmological model via the distance-redshift relation.

APPENDIX C: MODEL PARAMETERS

We show the marginalized parameter constraints for each of the SPEMD models in Table C1 and for each of the composite models in Table C2.

Table C1. SPEMD model parameters.

Parameter	Marginalized constraints						
	Fiducial	AGNmask + 1	AGNwht = 0	Arc + 1,60src	Group	Group + $z = 0.49$ group	Chameleon
θ_E (") ^a	0.944 ^{+0.004} _{-0.004}	0.943 ^{+0.004} _{-0.004}	0.950 ^{+0.004} _{-0.005}	0.933 ^{+0.004} _{-0.004}	0.912 ^{+0.003} _{-0.003}	0.927 ^{+0.005} _{-0.006}	0.946 ^{+0.003} _{-0.003}
q	0.80 ^{+0.01} _{-0.01}	0.80 ^{+0.01} _{-0.01}	0.81 ^{+0.01} _{-0.01}	0.79 ^{+0.01} _{-0.01}	0.79 ^{+0.01} _{-0.01}	0.78 ^{+0.01} _{-0.01}	0.78 ^{+0.01} _{-0.01}
θ_q (°)	32.7 ^{+0.8} _{-0.9}	32.8 ^{+0.8} _{-0.9}	34.7 ^{+1.1} _{-1.0}	31.9 ^{+0.8} _{-0.7}	32.9 ^{+0.6} _{-0.7}	33.5 ^{+0.8} _{-0.7}	31.5 ^{+0.7} _{-0.7}
γ'	1.96 ^{+0.02} _{-0.02}	1.98 ^{+0.02} _{-0.02}	2.01 ^{+0.01} _{-0.02}	1.90 ^{+0.02} _{-0.01}	1.94 ^{+0.01} _{-0.01}	1.95 ^{+0.01} _{-0.01}	2.02 ^{+0.01} _{-0.01}
γ_{ext}	0.117 ^{+0.004} _{-0.004}	0.120 ^{+0.003} _{-0.004}	0.125 ^{+0.004} _{-0.004}	0.109 ^{+0.003} _{-0.003}	0.110 ^{+0.003} _{-0.004}	0.110 ^{+0.003} _{-0.003}	0.126 ^{+0.003} _{-0.003}
θ_γ (°)	-89.6 ^{+0.9} _{-1.3}	-89.1 ^{+0.8} _{-1.3}	-87.7 ^{+0.5} _{-0.6}	88.0 ^{+0.7} _{-0.6}	82.4 ^{+0.9} _{-1.0}	82.6 ^{+1.8} _{-1.9}	-88.4 ^{+0.5} _{-0.3}
$X \theta_E$ (")	0.001 ^{+0.001} _{-0.001}	0.001 ^{+0.001} _{-0.001}	0.001 ^{+0.001} _{-0.001}	0.001 ^{+0.001} _{-0.001}	0.001 ^{+0.001} _{-0.001}	0.001 ^{+0.001} _{-0.001}	0.001 ^{+0.001} _{-0.000}
G2 θ_E (")	0.926 ^{+0.029} _{-0.027}	0.939 ^{+0.023} _{-0.025}	0.930 ^{+0.036} _{-0.036}	0.945 ^{+0.021} _{-0.027}	0.868 ^{+0.008} _{-0.008}	0.947 ^{+0.017} _{-0.014}	0.929 ^{+0.015} _{-0.014}
G2 q	0.66 ^{+0.02} _{-0.02}	0.67 ^{+0.02} _{-0.02}	0.67 ^{+0.02} _{-0.03}	0.68 ^{+0.02} _{-0.01}	0.70 ^{+0.01} _{-0.02}	0.65 ^{+0.01} _{-0.02}	0.69 ^{+0.01} _{-0.02}
G2 θ_q (°)	41.5 ^{+4.0} _{-4.7}	39.8 ^{+4.1} _{-3.4}	38.2 ^{+6.0} _{-5.3}	35.2 ^{+2.9} _{-3.4}	40.4 ^{+2.6} _{-1.9}	35.4 ^{+2.4} _{-2.1}	46.2 ^{+3.7} _{-4.0}

Notes. Reported values are medians, with errors corresponding to the 16th and 84th percentiles.

Angles are measured east of north.

^aSpherical-equivalent Einstein radius.

Table C2. Composite model parameters.

Parameter	Marginalized constraints					
	Fiducial	AGNmask + 1	AGNwht = 0	Arc + 1,60src	Group	Group + $z = 0.49$ group
Stellar M/L (M_\odot/L_\odot) ^a	2.2 ^{+0.2} _{-0.2}	2.1 ^{+0.2} _{-0.2}	2.1 ^{+0.2} _{-0.2}	2.1 ^{+0.2} _{-0.2}	2.1 ^{+0.2} _{-0.2}	2.1 ^{+0.2} _{-0.2}
Chameleon1 q	0.759 ^{+0.003} _{-0.001}	0.761 ^{+0.001} _{-0.001}	0.761 ^{+0.001} _{-0.001}	0.758 ^{+0.001} _{-0.001}	0.763 ^{+0.001} _{-0.001}	0.765 ^{+0.000} _{-0.000}
Chameleon1 θ_q (°)	23.1 ^{+0.2} _{-0.2}	23.1 ^{+0.2} _{-0.2}	23.3 ^{+0.2} _{-0.1}	23.3 ^{+0.1} _{-0.2}	23.3 ^{+0.1} _{-0.1}	23.4 ^{+0.1} _{-0.1}
Chameleon2 q	0.771 ^{+0.001} _{-0.001}	0.772 ^{+0.001} _{-0.001}	0.774 ^{+0.001} _{-0.001}	0.770 ^{+0.001} _{-0.001}	0.770 ^{+0.001} _{-0.001}	0.767 ^{+0.001} _{-0.001}
Chameleon2 θ_q (°)	26.1 ^{+0.3} _{-0.3}	26.4 ^{+0.3} _{-0.3}	25.5 ^{+0.3} _{-0.3}	25.8 ^{+0.3} _{-0.3}	26.4 ^{+0.2} _{-0.2}	26.3 ^{+0.2} _{-0.2}
NFW $\kappa_{0,h}$	0.143 ^{+0.003} _{-0.007}	0.148 ^{+0.003} _{-0.003}	0.147 ^{+0.003} _{-0.004}	0.159 ^{+0.005} _{-0.004}	0.136 ^{+0.002} _{-0.003}	0.149 ^{+0.001} _{-0.001}
NFW r_s (")	10.46 ^{+0.07} _{-0.14}	10.31 ^{+0.12} _{-0.08}	10.53 ^{+0.08} _{-0.08}	10.32 ^{+0.07} _{-0.08}	11.35 ^{+0.05} _{-0.10}	11.10 ^{+0.05} _{-0.05}
NFW q	0.88 ^{+0.02} _{-0.01}	0.90 ^{+0.00} _{-0.00}	0.92 ^{+0.01} _{-0.01}	0.88 ^{+0.01} _{-0.01}	0.88 ^{+0.01} _{-0.01}	0.89 ^{+0.01} _{-0.02}
NFW θ_q (°)	71.6 ^{+1.1} _{-1.0}	73.5 ^{+0.4} _{-0.4}	73.6 ^{+0.5} _{-0.7}	71.3 ^{+0.7} _{-0.8}	71.7 ^{+0.3} _{-0.4}	73.4 ^{+0.4} _{-0.4}
γ_{ext}	0.138 ^{+0.001} _{-0.002}	0.137 ^{+0.001} _{-0.001}	0.135 ^{+0.001} _{-0.001}	0.137 ^{+0.002} _{-0.001}	0.134 ^{+0.002} _{-0.002}	0.133 ^{+0.002} _{-0.001}
θ_γ (°)	-89.1 ^{+0.3} _{-0.3}	-88.6 ^{+0.1} _{-0.2}	-88.1 ^{+0.1} _{-0.1}	89.9 ^{+0.3} _{-0.2}	89.5 ^{+0.3} _{-0.4}	89.2 ^{+0.5} _{-0.4}
$X \theta_E$ (")	0.018 ^{+0.002} _{-0.002}	0.015 ^{+0.002} _{-0.002}	0.010 ^{+0.002} _{-0.002}	0.021 ^{+0.002} _{-0.002}	0.018 ^{+0.003} _{-0.002}	0.018 ^{+0.002} _{-0.002}
G2 θ_E (")	1.034 ^{+0.004} _{-0.005}	1.036 ^{+0.004} _{-0.003}	1.046 ^{+0.005} _{-0.005}	1.034 ^{+0.004} _{-0.004}	1.026 ^{+0.002} _{-0.002}	1.005 ^{+0.002} _{-0.003}
G2 q	0.94 ^{+0.01} _{-0.01}	0.91 ^{+0.01} _{-0.01}	0.90 ^{+0.01} _{-0.01}	0.95 ^{+0.01} _{-0.01}	0.93 ^{+0.01} _{-0.01}	0.93 ^{+0.00} _{-0.00}
G2 θ_q (°)	39.7 ^{+1.3} _{-2.1}	40.5 ^{+0.6} _{-0.7}	40.3 ^{+0.8} _{-0.8}	40.5 ^{+1.0} _{-0.8}	29.4 ^{+1.0} _{-0.8}	40.5 ^{+0.8} _{-1.0}

Notes. Reported values are medians, with errors corresponding to the 16th and 84th percentiles.

Angles are measured east of north.

^aM/L within θ_E for rest-frame V band. The point source component of the lens light is assumed to be from low-level AGN emission as opposed to stellar light and is not included in the calculation. The given uncertainties are a combination of statistical effects and a systematic uncertainty equal to the difference between the calculated M/L with and without the point source contribution. The stellar mass is calculated assuming $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, but changes in the cosmology affect the M/L by a negligible amount.

APPENDIX D: SOURCE RESOLUTION CHANGES

In Table D1, we show the BIC and BIC weight values for the source resolution tests described in Section 4.4. All models are the fiducial models run with different source resolutions. The BIC weights have been renormalized.

APPENDIX E: FURTHER TESTS ON THE SELECTION AND IMPLEMENTATION OF THE WEIGHTED NUMBER COUNT CONSTRAINTS**E1 The use of γ_{ext} constraints**

In Section 4 we have shown that a large value of $\gamma_{\text{ext}} \sim 0.10\text{--}0.14$ is required to model the lens system, even after the nearby perturbers

Table D1. Effective time-delay distance and BIC weighting for different source resolutions.

Model	$D_{\Delta t}$ (Mpc)	ΔBIC	Relative BIC weight
SPEMD fiducial, 47×47 source	4711^{+232}_{-184}	92	0.316
SPEMD fiducial, 48×48 source	4706^{+227}_{-180}	76	0.470
SPEMD fiducial, 49×49 source	4685^{+241}_{-210}	139	0.049
SPEMD fiducial, 50×50 source	4640^{+238}_{-195}	86	0.365
SPEMD fiducial, 51×51 source	4631^{+229}_{-184}	62	0.605
SPEMD fiducial, 52×52 source	4712^{+244}_{-193}	94	0.295
SPEMD fiducial, 53×53 source	4691^{+243}_{-202}	59	0.641
SPEMD fiducial, 54×54 source	4645^{+230}_{-172}	32	0.864
SPEMD fiducial, 56×56 source	4686^{+231}_{-186}	0	1.000
SPEMD fiducial, 58×58 source	4724^{+238}_{-188}	14	0.956
SPEMD fiducial, 60×60 source	4844^{+243}_{-183}	12	0.964
Composite fiducial, 47×47 source	4933^{+401}_{-253}	0	1.000
Composite fiducial, 48×48 source	4753^{+418}_{-251}	10	0.507
Composite fiducial, 49×49 source	5011^{+420}_{-259}	114	0.000
Composite fiducial, 50×50 source	4732^{+417}_{-247}	26	0.175
Composite fiducial, 51×51 source	4810^{+417}_{-254}	118	0.000
Composite fiducial, 52×52 source	4956^{+422}_{-261}	60	0.013
Composite fiducial, 53×53 source	4893^{+431}_{-258}	69	0.006
Composite fiducial, 54×54 source	4819^{+422}_{-258}	114	0.000
Composite fiducial, 56×56 source	4886^{+423}_{-251}	112	0.000
Composite fiducial, 58×58 source	4885^{+424}_{-250}	129	0.000
Composite fiducial, 60×60 source	4843^{+415}_{-248}	189	0.000

Note. Reported values are medians, with errors corresponding to the 16th and 84th percentiles.

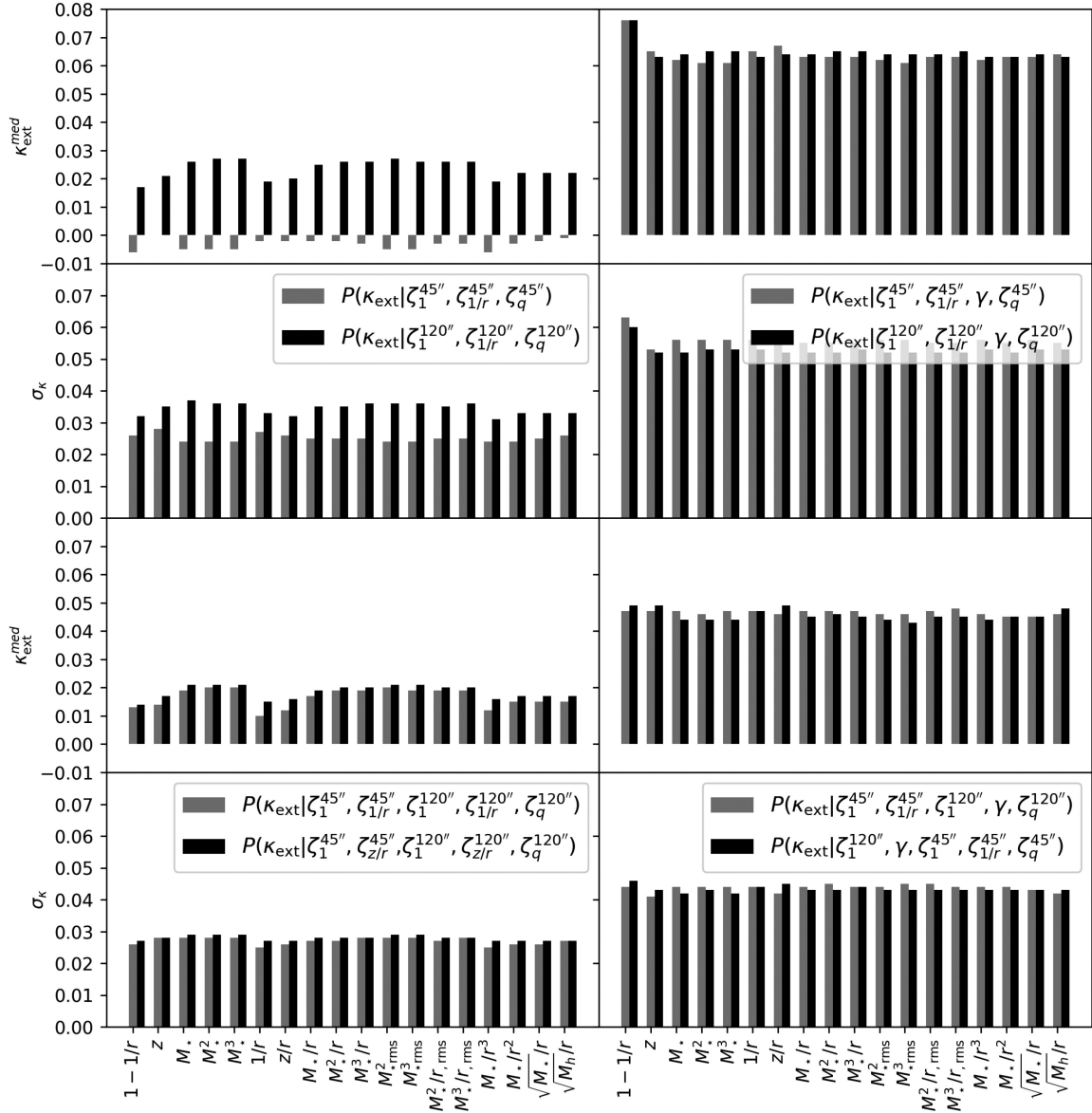


Figure E1. Statistics of $P(\kappa_{\text{ext}})$ for various combinations of weighted counts with or without shear, in the case of the fiducial lensing model. ‘1 – 1/r’ means that the constraint from ‘ $\zeta_{1/r}$ ’ is not used.

are being accounted for. In addition, the different inner mass profiles of the lens, explored in Section 4, all require a large external shear. This means that the large-scale environment and/or LOS structures must be responsible for this effect.

The specific large-scale structures responsible for this effect have not been identified, since the mass models incorporating the two large galaxy groups identified in H0LiCOW X still require a large γ_{ext} . This was also pointed out in previous studies, based on more limited ancillary data and less detailed modelling, both superseded in our work. For example, Wong et al. (2011) compute an expected $\gamma_{\text{ext}} = 0.08 \pm 0.03$ based on a spectroscopic galaxy catalogue, where galaxies are treated as SIS, and galaxy groups as NFW haloes. They note that the orientation of this shear caused by the environment does not match the one obtained from the mass models of the lens. Vuissoz et al. (2008), based on mass models constrained only by the positions of the quasar images (and in some models by the measured time delays as well), find γ_{ext} as large as 0.3, but as small as 0.06 if they

incorporate the galaxy group which includes the lens. However, our more complete spectroscopic catalogue shows that the group centroid is more distant from the lens (~ 30 arcsec compared to ~ 10 arcsec), leading to a larger amount of external shear necessary to model the system, after accounting for the group.

Based on the above, while we concede that we cannot find a mass model which explains most of the shear, we are confident that the external shear values measured for the various models in Section 4 are robust, and that we are justified to attribute the shear to the lens environment and/or LOS structures which are not captured in our mass models, but are captured in our statistical approach to computing κ_{ext} . This justifies our use of the γ_{ext} constraints to infer κ_{ext} in Fig. 10. As we will show in Appendix E2, this constraint has a dominant effect on our inference, which is expected from results of ray-tracing through the MS (e.g. see fig. 8 in Collett & Cunnington 2016), where $|\kappa|$ and $|\gamma|$ are found to correlate. We note that the lens RXJ1131 – 1231, which has also been modelled as part of

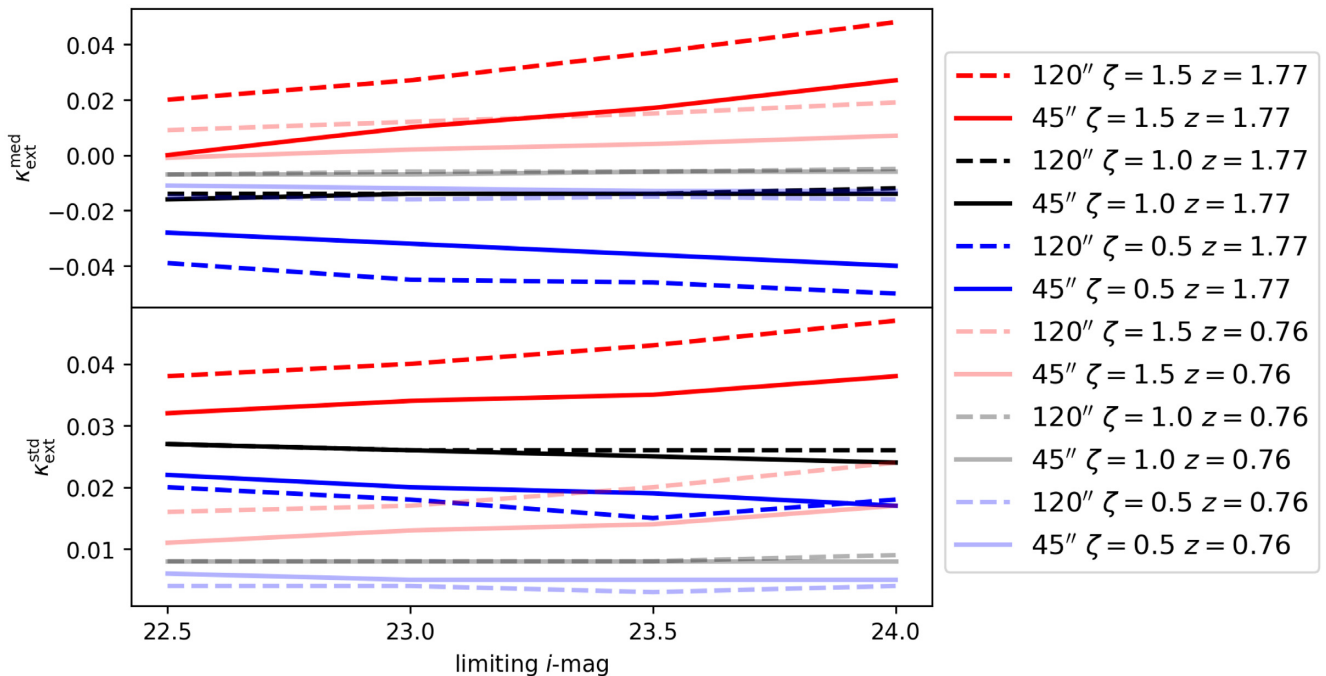


Figure E2. *Upper plot:* Behaviour of the median of κ_{ext} as a function of the limiting magnitude $i = 22.5, 23.0, 23.5,$ and 24.0 mag, for two choices of the aperture radii and three choices of the relative (unweighted) number counts. For completeness, we show the behaviour for two source redshifts in the MS, $z = 1.77$ (solid symbols) and $z = 0.76$ (transparent symbols). *Lower plot:* Behaviour of the semidifference between the 16th and 84th percentiles for the same distributions.

H0LiCOW (Suyu et al. 2013; Chen et al. 2019), also has a fairly large measured shear of ~ 0.08 , which was used to constrain its κ_{ext} . Independently, its κ_{ext} was measured using a different methodology by McCully et al. (2017), which has found κ_{ext} to be offset to smaller, but none-the-less consistent values.

After unblinding our analysis, we checked what the impact of using κ_{ext} inferred without the shear constraint would have been on our analysis. We obtain $H_0 = 76_{-3.0}^{+2.9}$ km s $^{-1}$ Mpc $^{-1}$, a value 6 percent larger than the one in Section 6, and with significantly increased statistical precision due to the tighter κ_{ext} distributions (see Fig. E1).

In order to avoid biases, we must ensure that we construct $P(\kappa_{\text{ext}}|\gamma, \dots)$ such that it is consistent with the discussion above. In particular, we must ensure that when we select LOS with large shear from the MS, the shear is not due to galaxies very close to the LOS, or to galaxy groups/clusters so massive that we would incorporate them in our mass models, in the real data. We address each of these in the following.

E1.1 Galaxies close to the LOS

In order to ensure that the weighted number counts are not dominated by the galaxies very close to the LOS, in Greene et al. (2013); H0LiCOW III, IX we have used a 5 arcsec-radius inner mask, and set an upper limit to the weights incorporating $1/r$ of $1/10$ arcsec for each galaxy. While the usage of a mask when computing weighted number counts is agnostic to the actual existence of galaxies inside of it, and therefore their contribution to κ_{ext} at that particular spatial location, here and in Chen et al. (2019) we only select LOS from the MS which have no galaxies inside the 5 arcsec radius. Such galaxies, if above the magnitude threshold, would be modelled explicitly in the real data. This radius corresponds to the inner $\gtrsim 4$

pixels of the (κ, γ) map from Hilbert et al. (2009). The addition of this constraint has the effect of lowering κ_{ext} by approximately $0.1\kappa_{\text{ext}}$.

E1.2 Massive large-scale structures

For the FOV around WFI2033 – 4723, we used our spectroscopic catalogue to identify large galaxy groups, which we incorporated in our mass models, in cases where they exceeded the flexion shift threshold. We must take this fact into account when we infer κ_{ext} by selecting LOS from the MS. The easiest way to account for this in the MS is to use a complete catalogue of galaxy groups, and to test at each location of the κ map whether their flexion shifts exceed the threshold. If so, those LOS are removed. In this way, we only use the κ_{ext} distribution free of the contribution of massive structures, which is what is needed.

In practice, we use a catalogue of galaxies from the MS, which identifies their parent haloes and the masses of those haloes. We impose the observational constraints from WFI2033 – 4723, that the groups (in this case the parent haloes) we model have at least five galaxies within our magnitude threshold. For these haloes, we convert their masses into velocity dispersions (assuming the SIS mass profile, for simplicity), and compute the radius of the circle around the halo centroid inside which the flexion shift for each of these structures exceeds the threshold. We then remove the LOS inside those circles from the κ_{ext} computation. We have compared the κ_{ext} distributions with and without removing these groups, and have found them to be indistinguishable, given the rarity of these groups. We therefore safely ignore the effect of massive but rare large-scale structures on our analysis.

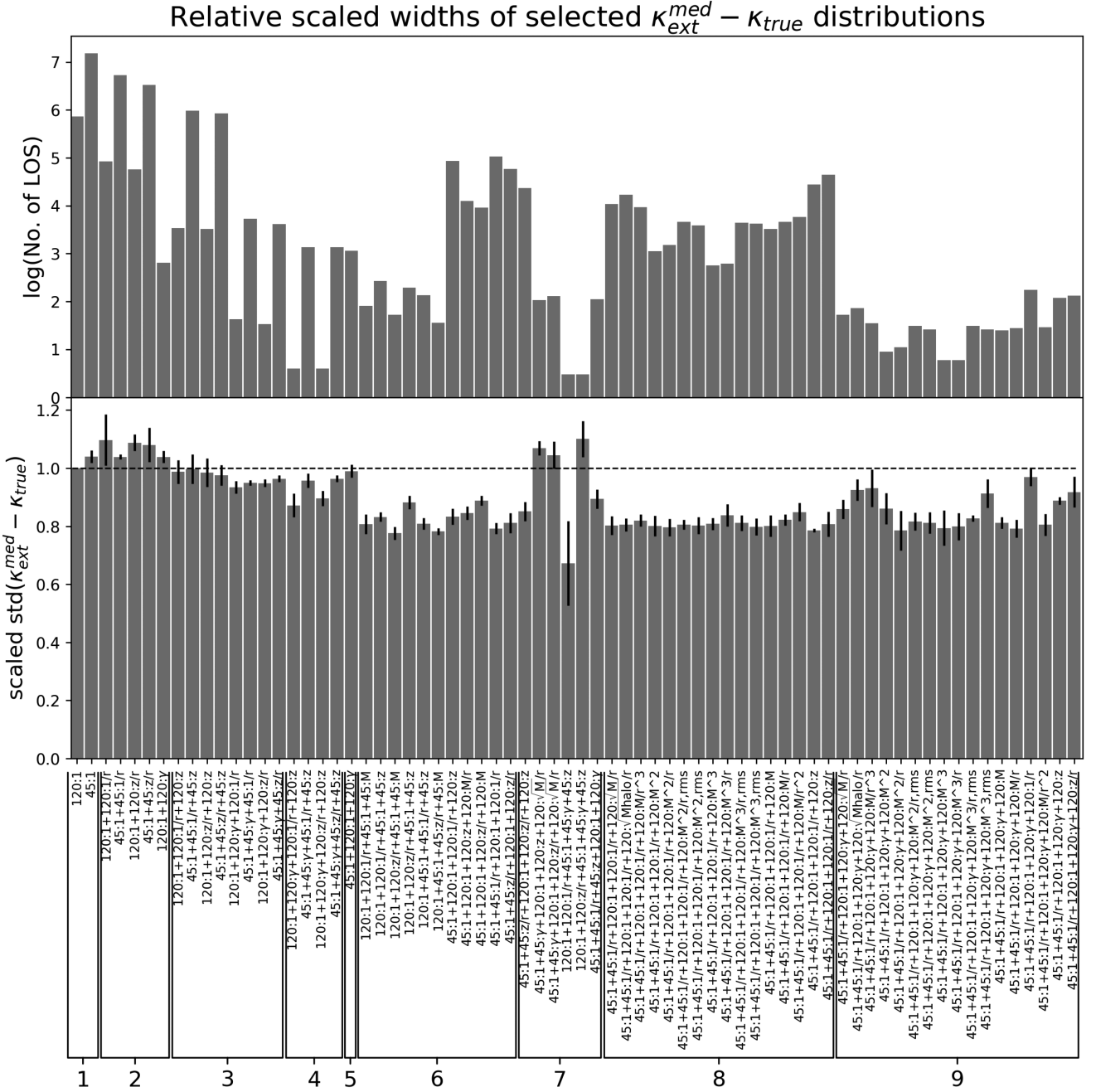


Figure E3. Lower plot: Standard deviations of 70 representative $\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{true}}$ distributions, relative to the standard deviation of $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{true}} | \zeta_1^{120''})$, and scaled with respect to $\kappa_{\text{ext}}^{\text{med}}$. See the text for details. Error bars represent the scatter resulting from running the simulations three times. Eight groups of distributions are identified: (1) single apertures, single constraint, (2) single apertures, two joint constraints, (3) single apertures, three joint constraints, (4) single apertures, four joint constraints, (5) two apertures, three joint constraints, (6) two apertures, four joint constraints, (7) two apertures, five joint constraints, (8) distributions of the type $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{true}} | \zeta_1^{45''}, \zeta_{1/r}^{45''}, \zeta_1^{120''}, \zeta_{1/r}^{120''}, \zeta_q^{120''})$, where q stands for various constraints; (9) distributions of the type $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{true}} | \zeta_1^{45''}, \zeta_{1/r}^{45''}, \zeta_1^{120''}, \zeta_q^{120''}, \gamma)$. Upper plot: median number of LOS which survived the given constraints and were used to compute each distribution. The more LOS, the more reliable the results of this simulations are.

E2 The choice of conjoined constraints

Greene et al. (2013) and H0LiCOW III have explored different combinations of weighted number count ratio constraints, most of which incorporate unweighted number counts ζ_1 , as well as $\zeta_{1/r}$. These correspond to the most robust constraints which can be determined from imaging data, usually with the tightest uncertainties. In Fig. E1 we explore the median and standard deviation

of the κ_{ext} distributions for various combinations of constraints, typically incorporating the two constraints above. We explore combinations of constraints measured inside the same aperture, as well as measured inside both the 45 arcsec- and 120 arcsec-radii apertures. The latter is because the weighted number count ratios we measure in Table 2 appear larger inside the 120 arcsec-radii aperture, suggesting that WFI2033 – 4723 is positioned

on the outskirts of a galaxy overdensity, and this observation might include statistical information useful for tightening the κ_{ext} distribution.

Comparing the distributions constrained by $(\zeta_1^{45''}, \zeta_{1/r}^{45''}, \zeta_q^{45''})$ and $(\zeta_1^{120''}, \zeta_{1/r}^{120''}, \zeta_q^{120''})$, we see that the former produce lower $\kappa_{\text{ext}}^{\text{med}}$. This is partly because of the somewhat lower weighted number counts inside the smaller aperture, but mostly due to the fact that the same overdensity over a larger aperture implies a larger structure, and therefore larger convergence. This is demonstrated in Fig. E2. In the case of constraints from both apertures, such as $(\zeta_1^{45''}, \zeta_{1/r}^{45''}, \zeta_1^{120''}, \zeta_{1/r}^{120''}, \zeta_q^{120''})$, $\kappa_{\text{ext}}^{\text{med}}$ is brought closer to the average value obtained from the two individual apertures. Amongst the various ζ_q , $\kappa_{\text{ext}}^{\text{med}}$ varies by ~ 0.01 , or at the 1 per cent level.

Once the large γ_{ext} is used as a constraint, $\kappa_{\text{ext}}^{\text{med}}$ reaches much larger values, as expected. All distributions including the $\zeta_{1/r}$ constraint are in agreement to within the 1 per cent level, even if we compare constraints from different aperture radii. Interestingly, once the constraints from different apertures are combined, the distributions for different ζ_q remain in agreement, but $\kappa_{\text{ext}}^{\text{med}}$ decreases by almost ~ 2 per cent, or $\sim 0.3\sigma$. We cannot fully explain this result, but we note that we have explored the $\kappa_{\text{ext}} - \gamma_{\text{ext}}$ plane constrained by either, as well as both apertures. It appears that this behaviour is a result of both the very large γ_{ext} value and of the large weighted count constraints, and disappears for smaller values of either.

We do not use more than five conjoined constraints, because the number of MS LOS surviving the cut becomes too small, and the distributions are noisy. Our conclusions regarding the convergence distributions from different constraints are similar to those from H0LiCOW III, wherever a direct comparison is possible.

E3 Tests for bias and scatter

In H0LiCOW III, we ran two types of simulations using the MS data, to check whether or not our combination of constraints biases $P(\kappa_{\text{ext}})$. We ran these for the case of WFI2033 – 4723 as well, and found that $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{ext}}^{\text{true}} | \zeta_q, \dots)$ is centred on zero for all ζ_q, \dots combinations in Figs E1 and E3. All our $P(\kappa_{\text{ext}})$ distributions are, therefore, unbiased.

We also wish to determine which combinations of constraints produces the tightest $P(\kappa_{\text{ext}})$. However, from Greene et al. (2013) as well as Fig. E1 we see that the medians and standard deviations of $P(\kappa_{\text{ext}})$ are always correlated, such that smaller $\kappa_{\text{ext}}^{\text{med}}$ implies smaller σ_{κ} . We therefore follow a different approach, where we use the scatter in $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{ext}}^{\text{true}} | \zeta_q, \dots)$. We normalize the scatter in $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{ext}}^{\text{true}} | \zeta_1^{45''})$ ¹¹ to the unit value, and we fit the standard

deviation of $P(\kappa_{\text{ext}} | \zeta_1^{45''})$ as a function of its median with a linear function, while we vary $\zeta_1^{45''}$. Finally, we divide the standard deviation of $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{ext}}^{\text{true}} | \zeta_q, \dots)$ by the value of this linear function at the corresponding $\kappa_{\text{ext}}^{\text{med}}$, and we show the results, for selected combinations of constraints, in Fig. E3. We can see that combining the results from the two different aperture radii typically results in scaled standard deviations reduced by up to ~ 20 per cent, whether the γ_{ext} constraint is used or not. While the figure shows a spike in value for the scaled standard deviation corresponding to the combination of weights we choose to consider as fiducial in this work, $(\zeta_1^{45''}, \zeta_{1/r}^{45''}, \zeta_1^{120''}, \zeta_{1/r}^{120''}, \gamma)$, most of the surrounding similar distributions (with only one of the ζ_q constraints replaced) show small values, and we know from Fig. E1 that the resulting $P(\kappa_{\text{ext}})$ for these constraints are consistent with each other.

¹National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

²Subaru Telescope, National Astronomical Observatory of Japan, 650 N Aohoku Pl, Hilo, HI 96720, USA

³Department of Physics, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA

⁴Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, Chiba 277-8583, Japan

⁵Laboratoire d'Astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland

⁶STAR Institute, Quartier Agora - Allée du six Août, 19c B-4000 Liège, Belgium

⁷Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany

⁸Physik-Department, Technische Universität München, James-Frank-Straße 1, D-85748 Garching, Germany

⁹Institute of Astronomy and Astrophysics, Academia Sinica, 11F of ASMA, No.1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

¹⁰Exzellenzcluster Universe, Boltzmannstr. 2, D-85748 Garching, Germany

¹¹Ludwig-Maximilians-Universität, Universitäts-Sternwarte, Scheinerstr. 1, D-81679 München, Germany

¹²Institute of Astronomy, University of Cambridge, Madingley Rd, Cambridge CB3 0HA, UK

¹³Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands

¹⁴Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA

¹⁵Pyörrekuja 5A, FI-04300 Tuusula, Finland

¹⁶Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands

¹⁷Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA

¹¹Using instead $P(\kappa_{\text{ext}}^{\text{med}} - \kappa_{\text{ext}}^{\text{true}} | \zeta_1^{120''})$ produces consistent results.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.