

The COSMOS-Web Lens Survey(COWLS) III: forecasts versus data

Natalie B. Hogg¹, James W. Nightingale^{2,★}, Quihan He³, Jacqueline McCleary⁴,
Guillaume Mahler^{3,5,6}, Aristeidis Amvrosiadis³, Ghassem Gozaliasl^{7,8}, Edward Berman⁴,
Richard J. Massey⁶, Diana Scognamiglio⁹, Maximilien Franco¹⁰, Daizhong Liu¹¹,
Marko Shuntov^{12,13}, Louise Paquereau¹⁴, Olivier Ilbert¹⁵, Natalie Allen^{12,13}, Sune Toft^{12,13},
Hollis B. Akins¹⁰, Caitlin M. Casey^{10,16,12}, Jeyhan S. Kartaltepe¹⁷, Anton M. Koekemoer¹⁸,
Henry Joy McCracken¹⁴, Jason D. Rhodes⁹, Brant E. Robertson¹⁹, Nicole E. Drakos²⁰,
Andreas L. Faist²¹, Hossein Hatamnia²² and Sophie L. Newman²³

Affiliations are listed at the end of the paper

Accepted 2025 September 22. Received 2025 September 22; in original form 2025 March 18

ABSTRACT

We compare forecasts for the abundance and properties of strong gravitational lenses in the COSMOS-Web survey, a 0.54 deg^2 survey of the COSMOS field using the NIRCam and MIRI instruments aboard *James Webb Space Telescope (JWST)*, with the first catalogue of strong lens candidates identified in the observed NIRCam data, COWLS. We modify the `lenspop` package to produce a forecast for strong lensing in COSMOS-Web. We add a new mock galaxy catalogue to use as the source population, as well as the COSMOS-Web survey specifications, including the transmission data for the four NIRCam filters used. We forecast 107 strong lenses can be detected in COSMOS-Web across all bands, assuming complete subtraction of the lens galaxy light. The majority of the lenses are forecast to have small Einstein radii ($\theta_E < 1 \text{ arcsec}$) and lie at redshifts in the range $0 < z < 2$, while the source redshift distribution peaks at $z \sim 3$ and has a long tail extending up to $z \sim 11$, unambiguously showing that strong lensing in *JWST* can probe the entirety of the epoch of reionization. We compare our forecast with the distributions of Einstein radii, lens photometric redshifts, and lens and source magnitudes in the observed lenses, finding that while the forecast and observed Einstein radii distributions match, the redshifts and magnitudes do not. The observed lens redshift distribution peaks at a slightly lower redshift than the forecast one, while the lens magnitudes are systematically brighter in the observed data than in the forecast.

Key words: gravitational lensing: strong – methods: statistical – cosmology: observations.

1 INTRODUCTION

Strong gravitational lensing – the deflection of light by a massive object that produces multiple images of the source – provides a wealth of information about our Universe. The magnification of distant sources has facilitated the study of high-redshift galaxies (Y. D. Hezaveh et al. 2013; A. M. Swinbank et al. 2015), while forward modelling of the lensed images can provide insights into the mass distribution of the inner part of lens galaxies (A. J. Shajib et al. 2021; C. Y. Tan et al. 2024) and the potential presence of low-mass dark sub-halos within the main lens or along the line of sight (S. Vegetti et al. 2012; C. M. O’Riordan et al. 2023).

A lensed transient source allows the time delay between separate images to be determined, which provides a measurement of H_0 , the expansion rate of the Universe at $z = 0$ (S. Refsdal 1964; S. H. Suyu & others 2014; K. C. Wong & others 2020). Bayesian hierarchical modelling enables constraints to be placed on other

cosmological parameters, such as the equation of state of dark energy, w (N. B. Hogg 2024; T. Li et al. 2024). A further cosmological application of strong lensing is the proposed measurement of the weak lensing distortions to strong lensing images (S. Birrer et al. 2017; S. Birrer, A. Refregier & A. Amara 2018; P. Fleury, J. Larena & J.-P. Uzan 2021; N. B. Hogg et al. 2023; T. Duboscq et al. 2024; N. B. Hogg et al. 2025), which may provide an independent and complementary constraint on the large-scale distribution of dark matter in the Universe to that provided by traditional cosmic shear measurements from galaxy surveys, e.g. T. M. C. Abbott & others (2018), H. Aihara & others (2018), and C. Heymans & others (2021).

These scientific applications of strong lensing are naturally limited by the number of strong lenses available to analyse, which is currently around 1000 (A. Sonnenfeld et al. 2020), but the very large numbers of lenses forecast to be observed in ongoing or forthcoming wide-area surveys such as *Euclid* and the Large Synoptic Survey Telescope (LSST) at the Vera C. Rubin Observatory should yield huge developments in these fields of study (A. Sonnenfeld & M. Cautun 2021). For example, *Euclid* alone is expected to find $\mathcal{O}(10^5)$ strong lenses (T. E. Collett 2015; P. Holloway et al. 2023;

* E-mail: James.Nightingale@newcastle.ac.uk

G. Ferrami & J. S. B. Wyithe (2024), bringing strong lensing into the era of precision astrophysics and cosmology, with a large number of strong lenses already identified in the Early Release Observations (J. A. Acevedo Barroso & others 2025; R. Pearce-Casey & others 2025; B. C. Nagam & others 2025; C. M. O’Riordan 2025) and the Quick Data Release, or Q1, observations (P. Holloway et al. 2025; T. Li et al. 2025; N. E. P. Lines et al. 2025; K. Rojas et al. 2025; M. Walmsley et al. 2025).

Understanding the strong lensing science return that can be expected from a given survey relies on an accurate forecast of the number of strong lenses that a survey will observe. However, accurately forecasting the abundance and properties of strong gravitational lenses can be difficult. First, the number of strong lenses that exist in the observable Universe, or some fraction of it, must be calculated. To do this, knowledge of or assumptions about both the source and deflector populations must be obtained or made. Secondly, once the idealized strong lens population is characterized, lenses must be simulated and ‘observed’ according to the specifications of the target survey. This then determines the number of lenses that will not only exist in a given survey area, but will be detected by the instrument used to observe it.

Furthermore, it is not straightforward to characterize the number of detectable lenses that will actually end up being identified and catalogued as such. This depends on the strategy used to inspect the data – for example, expert human visual inspectors versus machine-learning-led searches (P. J. Marshall et al. 2009; F. Lanasse et al. 2018; A. More et al. 2024) or citizen science approaches (P. J. Marshall & others 2016; A. More & others 2016) – and the reliability of that strategy; see K. Rojas & others (2023) for an investigation into the reliability of visual inspection. In turn, the success of an inspection campaign can only be estimated if an accurate forecast for the number of detectable lenses is known.

This work thus addresses the following question: how well do strong lensing forecasts match what is seen in real data? To answer this question, we make a forecast for the number of detectable lenses in the COSMOS-Web survey and compare their abundance and predicted properties with the first COSMOS-Web catalogue of detected strong lenses, COWLS (COSMOS-Web Lens Survey; G. Mahler & others 2025; J. Nightingale & others 2025). COSMOS-Web is a survey of 0.54 deg^2 of the COSMOS field, imaged using the NIRCcam and MIRI instruments onboard the *James Webb Space Telescope* (JWST; C. M. Casey & others 2023). The COWLS catalogue was constructed following expert visual inspection of NIRCcam data, the MIRI observations being omitted at this stage. Accordingly, our forecast is made for the NIRCcam observations of COSMOS-Web alone.

The forecasting tool we use is the Python package `lenspop`, which was used to forecast strong lens abundances and guide observing strategies for surveys such as *Euclid* and LSST (T. E. Collett 2015; Y. Mellier & others 2025). The package was shown to correctly predict the number of strong lenses found in the Strong Lensing Legacy Survey (SL2S; R. A. Cabanac et al. 2007), along with the distribution of their Einstein radii, validating the forecasting procedure. The package has since been used to forecast the strong lensing abundance for other forthcoming surveys, such as the *Nancy Grace Roman Space Telescope* (C. Weiner, S. Serjeant & C. Sedgwick 2020) and the *China Space Station Telescope* (X. Cao & others 2024), as well as the abundance of strongly lensed supernovae in LSST observations (A. Sainz de Murieta et al. 2024).

Given the demonstrated accuracy of the `lenspop` model, we can use the results of the forecast to both interpret current data and guide future strong lensing searches in COSMOS-Web. We further compare

our results with the COSMOS-Web strong lensing forecasts made by P. Holloway et al. (2023) and G. Ferrami & J. S. B. Wyithe (2024), which followed different methodologies, as we will later discuss.

This paper is organized as follows: in Section 2 we describe the search for strong lenses in COSMOS-Web and the first catalogue of detections. In Section 3 we describe our forecasting methodology. In Section 4 we present the results of our forecast, while in Section 5 we compare the results of the forecast with the observed data and other forecasts; we conclude with Section 6.

2 COSMOS-WEB AND THE SEARCH FOR STRONG LENSES

In this section, we first provide an overview of the COSMOS-Web survey, followed by a description of the search for strong lenses in COSMOS-Web.

2.1 COSMOS-Web

COSMOS-Web is a survey of a 0.54 deg^2 area of the COSMOS field in the $F115W$, $F150W$, $F277W$, and $F444W$ filters of NIRCcam, with a 0.19 deg^2 sub-area also imaged with the MIRI instrument in the $F770W$ filter (C. M. Casey et al. 2023). The survey design was motivated by previous efforts with the *Hubble Space Telescope* (HST) to image certain areas of sky to as great a depth as possible (R. E. Williams et al. 1996; S. V. W. Beckwith et al. 2006; A. M. Koekemoer et al. 2007), allowing systematic study of galaxies at high redshift; and also by the Cosmic Evolution Survey (COSMOS; N. Scoville & others 2007), which mapped a contiguous 2 deg^2 sky area, far larger than all previous deep field studies, allowing galaxy formation and evolution to be connected to their environment and the large-scale structure over a vast redshift range, $0 < z < 6$.

COSMOS-Web emulates that philosophy, with a sky area sufficiently sizeable to capture structures on the order of 100 Mpc, and imaging reaching 5σ point-source depths of up to magnitude 28.2, giving it the ability to detect and observe thousands of galaxies up to $z \sim 11$ (C. M. Casey et al. 2024; M. Franco et al. 2024) as well as probing both their immediate and wider surroundings. The primary science goals of the survey are the discovery of galaxies in the epoch of reionization, i.e. $6 \lesssim z \lesssim 11$, as well as mapping the distribution of reionization and investigating what drives it; the identification of rare quiescent galaxies at $z > 4$ that will allow constraints to be placed on the formation of very massive galaxies; and the measurement of the evolution of the stellar-to-halo mass relation in galaxies via weak gravitational lensing.

Furthermore, previous studies of the COSMOS field have yielded the discovery of over 100 strong gravitational lens candidates (C. Faure & others 2008; N. Jackson 2008; A. More et al. 2012; S. Jin et al. 2018; M. Pourrahmani, H. Nayyeri & A. Cooray 2018; E. O. Garvin et al. 2022; J. Pearson & others 2023; S. Jin et al. 2024; W. Mercier et al. 2024). Due to COSMOS-Web, the known strong lens candidates in the field now have high-resolution imaging in the four NIRCcam filters in addition to the previous photometric and spectroscopic observations. At the survey’s inception, it was estimated that 90 new lenses would be discovered, due to the ability of JWST to resolve small Einstein radii lenses and the unprecedented depth of the survey in the near-infrared (C. M. Casey et al. 2023). This work refines this estimation, and furthermore, provides predictions for the distributions of lens and source properties, which we can then compare with those of the newly discovered candidate lenses.

2.2 The COSMOS-Web Lens Survey

The initial search for strong lenses in COSMOS-Web was conducted by members of the collaboration's lensing working group, who individually visually inspected images of the $\sim 43\,000$ most massive galaxies in the COSMOS-Web data as identified as those brighter than AB magnitude 23 in the $F277W$ band, ranking each image as either 'yes' (the image contains a strong lens), 'maybe' (the image maybe contains a strong lens), or 'no' (the image does not contain a strong lens). Visual inspectors were presented with two colour composite images of each candidate, the same images with the lens light subtracted using a Sérsic light profile, and the Sérsic profiles used for the subtraction. There were 17 candidates that received 'yes' ratings by more than half the inspectors at this stage, and are considered to be the most spectacular strong lenses in the COSMOS-Web data. A detailed description of these 17 objects is presented in G. Mahler et al. (2025).

Over 1000 images were ranked as 'maybe', and of these candidates, all that were ranked as 'yes' or 'maybe' by at least half the visual inspection team, plus a significant number of edge cases identified by the lead inspector were selected for detailed follow-up. The PyAutoLens [graphic](#) software (J. W. Nightingale et al. 2021) was used to try to fit a lens model to each of the resulting 419 candidates. The images were then re-examined by six visual inspectors using these lens models, along with a more detailed ranking system. Candidates were ranked as either of the following :

- (i) A: high confidence this is a strong lens.
- (ii) B: likely a strong lens, but there is ambiguity.
- (iii) U: unlikely to be a strong lens, but not impossible.
- (iv) S: a singly imaged strong lens feature/arc (i.e. without an observed counter image).
- (v) X: not a lens.
- (vi) I (Optional and independent): the candidate is interesting and warrants follow-up investigation.

Once each candidate had been graded, it was assigned a score, with 2 points for an A grade and 1 point for either B or S. The highest possible score is therefore 12, for a lens ranked as A by all six inspectors. This procedure led to a discovery of 42 strong lenses that scored at least 8/12; there were 99 strong lenses that scored at least 6/12. The lens modelling performed by J. Nightingale et al. (2025) means that all of the candidates have an accompanying lens model; full details of the strong lens search and the resulting catalogue of lenses can be found in that paper.

To facilitate the comparison with the forecast performed in this work, we label the 59 spectacular plus highly ranked candidates (17 + 42) as 'high confidence' strong lenses, and the 116 spectacular plus highly and averagely ranked candidates (17 + 99) as 'high and medium confidence' strong lenses. For the purposes of this paper, we do not take into account any of the COWLS lens candidates with scores of less than 6/12. We will discuss this, and other selection effects further in Section 5.3.

3 FORECASTING METHODOLOGY

We use the Python package `lenspop`¹ to perform our forecast. The package works by computing the idealized number of strong lenses in the Universe, modelling all deflectors as singular isothermal ellipsoids (SIEs) and using galaxies drawn from a given mock

catalogue as sources (`lenspop` originally contained one such catalogue, but we have modified it to supply our own). The idealized lenses are then 'observed' according to the chosen survey, yielding not only a prediction for the number of lenses, but also a catalogue of their properties, such as Einstein radii and redshifts. In this section we provide an overview of the `lenspop` model, and refer readers to T. E. Collett (2015) for more details. We also describe our modifications to `lenspop`, which allow us to use it to forecast for COSMOS-Web.²

3.1 The deflector population in `lenspop`

The `lenspop` package simulates a population of deflectors by assuming that the strong lensing cross-section is dominated by elliptical galaxies, meaning that all deflectors are modelled by the SIE profile (R. Kormann, P. Schneider & M. Bartelmann 1994). The SIE convergence is given by

$$\kappa(x, y) = \frac{1}{2} \left(\frac{\theta_E}{\sqrt{qx^2 + y^2/q}} \right)^2, \quad (1)$$

where θ_E is the circularized Einstein radius of the lens, q is the axis ratio of the ellipse, and the x, y coordinates are defined in a coordinate system aligned with the major and minor axes of the lens.

The number density of deflectors is computed by integrating a modified Schechter function expressed in terms of the velocity dispersion, σ_v , derived by Y.-Y. Choi, C. Park & M. S. Vogeley (2007) from the Sloan Digital Sky Survey (SDSS) DR5 data (D. G. York & others 2000; J. K. Adelman-McCarthy & others 2007),

$$dN = \phi_* \left(\frac{\sigma_v}{\sigma_*} \right)^\alpha \exp \left[- \left(\frac{\sigma_v}{\sigma_*} \right)^\beta \right] \frac{\beta}{\Gamma(\alpha/\beta)} \frac{1}{\sigma_v} d\sigma_v, \quad (2)$$

where $\phi_* = 8 \times 10^{-3} h^3 \text{Mpc}^{-3}$, $\sigma_* = 161 \text{ kms}^{-1}$, $\alpha = 2.32$, and $\beta = 2.67$. Here, α describes the low-velocity power-law index, and β describes the high velocity exponential cut-off index. The shape and normalization of this function do not evolve with redshift.

Velocity dispersions for each deflector are drawn from equation (2), while redshifts for each deflector are drawn from the differential comoving volume defined by the maximum deflector redshift. In the original `lenspop` model this was set at $z = 2$, but in this work we increase the maximum redshift to $z = 3$, since we expect that higher redshift lenses will be present in COSMOS-Web due to *JWST*'s sensitivity to faint and red objects. For example, the first strong lens detected in the survey, the COSMOS-Web Ring, is at $z = 2.02$, with a source at $z = 5.10$ (W. Mercier et al. 2024; M. Shuntov et al. 2025).

Since all deflectors are modelled as SIEs,³ their Einstein radii are given by

$$\theta_E = 4\pi \frac{\sigma_v^2}{c^2} \frac{D_{ds}}{D_{os}}, \quad (3)$$

where D_{ds} and D_{os} are the angular diameter distances between deflector and source and observer and source, respectively. These are computed from the redshift for each deflector and source (once the source population has been drawn, see Section 3.2), in a spatially flat Lambda cold dark matter cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$,

²The modified version of `lenspop` is available on Github: <https://github.com/nataliehogg/lenspop>.

³Above $z = 3$, we may expect that a larger proportion of the deflector population will be disc, spiral, or irregular galaxies, necessitating a modification of the deflector model to account for this. We leave the quantification of a potential $z \geq 3$ deflector population to a future work.

¹<https://github.com/tcollett/LensPop>.

and $h = 0.7$. The axis ratio for each deflector is drawn from a Rayleigh distribution, truncated at $q = 0.2$ to avoid highly elliptical mass profiles.

The deflector light profiles are modelled by an elliptical de Vaucouleurs profile (G. Vaucouleurs 1948, equivalent to the Sérsic profile with the Sérsic index $n = 4$) whose centre and ellipticity are the same as those of the mass profile. Given a velocity dispersion, the fundamental plane relation for elliptical galaxies yields the effective radius and absolute magnitude for the deflector's light. The `lenspop` model employs the fundamental plane obtained by J. B. Hyde & M. Bernardi (2009a, b) from SDSS DR4 and DR6 data.

The quantities that describe each deflector are therefore its redshift z_{lens} , velocity dispersion σ_v , Einstein radius θ_E , axis ratio q , effective radius R_{eff} , and absolute magnitude M in each band. Weak lensing by line-of-sight objects is not included in the model. We do not modify any part of the above computation for our work.

3.2 The source population in `lenspop`

The source population parameters are the redshifts, light profiles, and the absolute magnitudes in each observation band. The redshifts and magnitudes are drawn from a given mock galaxy catalogue. In the original `lenspop` model, the mock galaxy catalogue used was produced for LSST by adding baryonic effects such as gas cooling, star formation, supernovae, and active galactic nuclei to the dark-matter-only Millennium simulation (G. De Lucia et al. 2006; A. J. Connolly et al. 2010). The catalogue covers a $4.5 \times 4.5 \text{ deg}^2$ area of the sky, with the highest redshift objects being around $z = 6$ and the smallest halo mass being $2.5 \times 10^9 M_\odot$.

The source light profiles are all assumed to be elliptical exponential profiles, with the ellipticity drawn from a Rayleigh distribution truncated at $q = 0.2$, in line with the ellipticity distribution of the deflector profiles. The effective radii of the source light profiles are given by

$$\log_{10} \left(\frac{R_{\text{eff}}}{\text{kpc}} \right) = \left(\frac{M}{-19.5} \right)^{-0.22} \left(\frac{1+z}{5} \right)^{-1.2} + \mathcal{N}, \quad (4)$$

where M is the absolute magnitude and $\mathcal{N}$ is a normally distributed random variable with a scatter of 0.3.

To improve the suitability of `lenspop` as a forecasting tool for COSMOS-Web, we must modify the source population used. Since *JWST* is a near-infrared telescope, it is sensitive to fainter and redder sources than those in the surveys that `lenspop` was originally designed to forecast for, such as LSST. To ensure that these faint sources are present in our forecast, we added a new mock galaxy catalogue to `lenspop` to be used as the source population.

We chose to use the JADES extraGalactic Ultra-deep Artificial Realization (JAGUAR) mock catalogue (C. C. Williams et al. 2018), which was originally created during the survey design of the *JWST* Advanced Deep Extragalactic Survey (JADES), a Guaranteed Time Observation program to observe the Hubble Deep and Ultra Deep fields using all three *JWST* instruments: NIRCcam, NIRSpec, and MIRI (D. J. Eisenstein & others 2023).

The final JAGUAR mock catalogue comprises 10 realizations of an $11 \times 11 \text{ arcmin}^2$ area, with each realization containing roughly 300 000 objects up to $z \sim 15$, with the smallest halo mass being $1 \times 10^6 M_\odot$ (C. C. Williams et al. 2018). The catalogue contains both star-forming and quiescent galaxies. A spectral energy distribution (SED) is assigned to each mock galaxy using the BEAGLE tool (J. Chevallard & S. Charlot 2016), with the complete catalogue providing galaxy properties such as redshift, half-light radius, axis ratio, and fluxes corresponding to observations in NIRCcam filters.

To ensure consistency in the properties of our source population, we replace the calculations of the effective radius and axis ratio for each source described above with their values provided in the catalogue. We use only a single realization of the mock. We note that the JAGUAR mock catalogue covers a relatively small area of the sky; the main effect this will have on our results is a potential underestimation of the abundance of the most massive source galaxies in the Universe.

3.3 Observing and detecting lenses

After the creation of the deflector and source populations following the procedure described above, the number of strong lenses potentially observable in the sky can be computed. First, equation (2) is integrated with respect to redshift, which provides the number of galaxies that can act as deflectors. Whether a given deflector lenses a source is computed in the following manner: a square box around each deflector is constructed such that it contains, on average, one source galaxy. This is done using the source area number density, d , such that the length of the side of the box in arcseconds is given by $l = \sqrt{1/d}$. For the source catalogue used in this work, $l \approx 1.2''$. The deflector is located at the centre of the box, while the source is placed within the box at random. The source–deflector pair thus creates a strong lensing system if and only if $z_{\text{source}} > z_{\text{lens}}$ and $\theta_E^2 > x_s^2 + y_s^2$, i.e. the source is located within the Einstein radius of the deflector in the image plane. We note that if (large) external shear induced by line-of-sight objects were to be included in the lens model the latter criterion would no longer necessarily hold, as the astroid caustics characteristic of shear can extend outside the radial caustic associated with the SIE potential, altering the strong lensing cross-section (C. S. Kochanek 2004; D. Huterer, C. R. Keeton & C.-P. Ma 2005; D.-W. Lee & S.-J. Kim 2014).

In the original `lenspop` model, this calculation led to an ideal population of 1.1×10^7 lenses; with our modifications, we find 2.8×10^8 . This order of magnitude increase comes from the greater depth of the new source population compared to the old, as well as its increased redshift range. We also simulate deflectors up to $z = 3$, rather than $z = 2$ in the original work of T. E. Collett (2015).

At this point, the so-called transfer function is applied to determine which strong lenses from the idealized population could be observed by a given survey. To do this, the idealized lenses are evaluated on a pixel grid, with a pixel scale given by the pixel scale of the chosen survey, and convolved with a Gaussian point spread function (PSF). Poisson noise is added, and further specifications of the survey are taken into account, including sky area, seeing (for ground-based telescopes), exposure time, zero-points, and filters. For the complete survey specifications used in the original `lenspop` model, see table 1 of T. E. Collett (2015).

For this work, we added the filter transmission data (C. Rodrigo, E. Solano & A. Bayo 2012) for the four NIRCcam filters used in COSMOS-Web to `lenspop`, along with the survey specifications for COSMOS-Web; see Table 1 for the specifications used. Note that we apply the same pixel scale of 0.031 arcsec to all four bands, whereas in reality the *F277W* and *F444W* bands have a pixel scale of 0.063 arcsec. This is because `lenspop` currently only allows for a single value to be passed as the pixel scale.

Furthermore, we do not modify the PSF computation, with each image still being convolved with a Gaussian PSF. We do not expect this to significantly impact the forecast number of detectable lenses, as the dominant PSF-related selection effect for extended sources is that of its size [full width at half-maximum (FWHM)] rather than its shape (A. Sonnenfeld et al. 2023). We change the default PSF

Table 1. Properties of the COSMOS-Web survey used in our `lenspop` forecast.

Property	Value(s)			
Sky area (deg ²)	0.54			
Exposure time (s)	257			
Pixel scale (")	0.031			
Filters	<i>F115W</i>	<i>F150W</i>	<i>F277W</i>	<i>F444W</i>
PSF FWHM (")	0.04	0.05	0.092	0.145
Sky brightness (mag arcsec ⁻²)	30.96	29.96	28.96	28.15

FWHM used in `lenspop` to those of the NIRCcam filters; the values are listed in Table 1.

Lastly, for a lens to be deemed detectable, it must pass a number of selection cuts. The criteria specified by T. E. Collett (2015), which we also follow, are as follows:

- (i) multiple images of the source must be produced, $\theta_E^2 > x_s^2 + y_s^2$;
- (ii) the image and counter-image must be resolved, $\theta_E^2 > r_s^2 + (s/2)^2$, where r_s is the angular unlensed source size and s is the seeing (PSF FWHM);
- (iii) tangential shearing of the arcs must be detectable, $\mu_{\text{tot}} r_s > s$, $\mu_{\text{tot}} > 3$, where μ_{tot} is the total magnification of the source;
- (iv) the signal-to-noise ratio must be large enough that the previous criteria are obviously satisfied to a human visual inspector, $\text{SNR}_{\text{tot}} > 20$, where SNR_{tot} is the total signal-to-noise ratio.

4 RESULTS

In this section, we describe our results. We note that the forecast presented in this work was produced by NBH concurrently with the first round of visual inspection of the data, and there was no attempt to keep the forecasted results blinded from the inspection team while the visual inspection was ongoing. NBH also contributed to both rounds of visual inspection described in G. Mahler et al. (2025) and J. Nightingale et al. (2025).

4.1 The forecast abundance of strong lenses

With the above-described modifications to `lenspop`, we predict that 107 strong lenses (i.e. objects that meet the four criteria outlined in Section 3.3) will be detected in COSMOS-Web across all bands, assuming lens light subtraction. To estimate a lower bound on the number of detectable lenses, we can examine so-called difference images, produced by subtracting the data in a given band from another at a shorter wavelength (R. Gavazzi et al. 2014). We compute the *F115W*–*F150W*, *F115W*–*F277W*, and *F115W*–*F444W* differences, finding that 14 lenses would be discovered in each.

Our result compares favourably with the number of strong lenses found in COSMOS-Web after two rounds of visual inspection. After the first round of visual inspection, 17 strong lenses were discovered (G. Mahler et al. 2025), which already exceeds the lower bound on the abundance predicted in our difference images. The second round of visual inspection targeted the 419 most promising remaining candidates; based on their scores, we identify 59 as high confidence strong lenses, or 116 as high and medium confidence strong lenses.

Considering just the number of high confidence strong lenses, and assuming the `lenspop` model is perfectly accurate, this implies that potentially as many as 48 strong lenses are present in the COSMOS-Web field that were not identified as such by the G. Mahler et al. (2025) and J. Nightingale et al. (2025) visual inspections, or that some of the weaker candidates from those inspections should have

their scores reassessed and increased. These may only be identified after spectroscopic follow-up to obtain lens and source redshifts. On the other hand, if instead the visual inspection is considered to be perfectly accurate and that all true strong lenses in COSMOS-Web are already included in the high confidence sample, this implies that the `lenspop` model is significantly overpredicting the abundance of strong lenses.

With the small number statistics of COSMOS-Web, this difference between prediction and reality may be considered tolerable, but for larger surveys such as *Euclid*, a detection of up to 50 percent fewer high confidence candidate lenses than predicted by T. E. Collett (2015) may be dissatisfying. It is also worth noting that the first systematic search for strong lenses in *Euclid* Early Release Observations, which covered 0.7 deg² of the sky, led to the discovery of three ‘A’ grade and 13 ‘B’ grade candidates,⁴ of which five were successfully fit with a lens model (J. A. Acevedo Barroso et al. 2025). Extrapolating to the full *Euclid* area, this implies $100\,000^{+70\,000}_{-30\,000}$ strong lenses. The 170 000 strong lenses predicted by T. E. Collett (2015) with `lenspop` sits on the upper limit of this extrapolation.

4.2 Simulated images

The lens properties in our forecast catalogue can be used to simulate images of COSMOS-Web-like strong lenses. Four illustrative examples are shown in Fig. 1, produced using the `lenstronomy` graphic package (S. Birrer & A. Amara 2018; S. Birrer et al. 2021). Each panel shows a type of strong lens that could be seen in COSMOS-Web, with different lens and source colours clearly visible. The colours are representative and are produced by combining the data simulated in three different NIRCcam bands, *F150W*, *F277W*, and *F444W*. The images follow the same scale in magnitude and are ordered from left to right in terms of decreasing magnitude of the lens light in the *F444W* band.

As is typical with strong lensing, the lensed source light in each case is faint compared to the light of the lens galaxy, and the small Einstein radii mean that some lensed features are partially obscured by or blended with the lens light. Since the `lenspop` model assumes all deflectors are SIEs, the arc features in the simulated images are regular and easily identifiable as the product of strong lensing. This is not the case in real data, where deflectors can have more complex gravitational potentials, which, combined with complex lens and source light morphology, can lead to less obvious strong lensing features. All of these problems were motivators for the second stage of visual inspection by J. Nightingale et al. (2025), in which lens models were fit to the candidate lenses and inspected along with the images as an additional discriminatory factor.

4.3 Properties of the forecast strong lenses

Along with the calculation of the number of strong lenses in the survey, we can also examine the properties of the forecast systems. In Fig. 2, we show the distributions of Einstein radii, velocity dispersions, and lens and source redshifts for our forecast catalogues. The median Einstein radius is 0.55 arcsec, with 88 percent of the lenses having Einstein radii smaller than 1 arcsec. The PSF in the *F115W* band provides a hard lower limit on the Einstein radius of a detectable lens, 0.04 arcsec. Only 0.2 percent of the lenses have

⁴The *Euclid* strong lens gradings are slightly more confident than those defined in this paper, with ‘A’ grade candidates considered to be ‘sure lenses’ and ‘B’ grade candidates considered ‘probable lenses’.



Figure 1. Multiband images of four simulated COSMOS-Web-like lenses using properties of the forecast population. From left to right, there is a near-complete Einstein ring, an extended arc, a single image (the counter-image being obscured by the lens light), and a distinctive blue lensed source. The colours are representative and are produced by combining the data simulated in three different NIRC2 bands ($F150W$, $F277W$, and $F444W$).

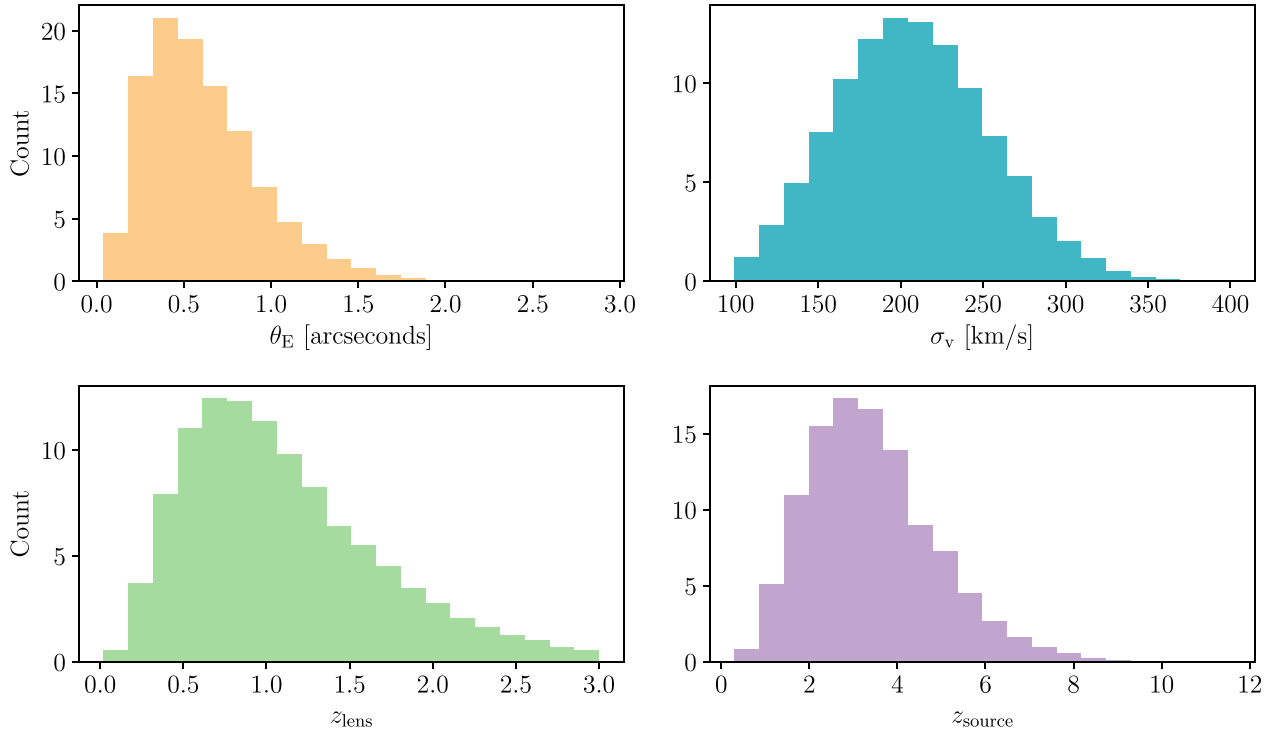


Figure 2. The distribution of Einstein radii (top left), velocity dispersions (top right), lens redshifts (bottom left), and source redshifts (bottom right) for our forecast lenses. Each histogram is weighted so the total count sums to 107, the number of lenses we forecast for COSMOS-Web.

Einstein radii larger than 2 arcsec. In conjunction with the distribution of velocity dispersions, which peaks at 200 km s^{-1} , this implies that the majority of strong lenses detected in COSMOS-Web (and by extrapolation, in other small area *JWST* surveys) will be low-mass galaxies.

The bottom left panel of Fig. 2 shows that the lens redshift distribution peaks at $z = 1$, with a long tail extending to our maximum allowed lens redshift, $z = 3$. The source redshift distribution, shown in the bottom right panel, peaks at $z = 3.2$. Despite the source population containing galaxies up to $z \sim 15$, the maximum source redshift in our catalogue of forecast lenses is $z = 11.6$, with 5.3 percent of the sources lying between this value and $z = 6$. This explicitly demonstrates the potential that COSMOS-Web and other *JWST* surveys have to find strong lens source galaxies that lie in the epoch of reionization. The difference between the distributions of the forecast source redshifts and that of the JAGUAR catalogue from which the source population is drawn is shown in Fig. 3.

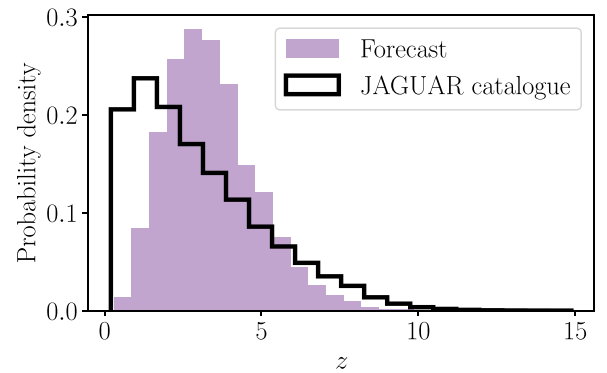


Figure 3. Distribution of the source redshifts in our COSMOS-Web-like forecast systems (solid purple) and the redshifts of galaxies in the JAGUAR mock catalogue from which the source population was drawn (unfilled black).

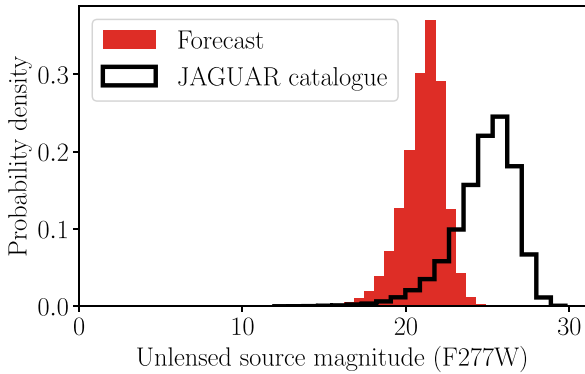


Figure 4. Distribution of the forecast unlensed source magnitude in the $F277W$ band (solid red) and the unlensed source magnitude in the JAGUAR catalogue from which the forecast population was drawn (unfilled black).

The absence of the high-redshift tail in the forecast source redshift distribution implies that lensing of extremely high-redshift sources is rare (as expected due to the comparative rarity of galaxies at high redshifts, meaning there are fewer sources to be lensed), and that even if an extremely high-redshift source is lensed, the magnification is likely insufficient to lead to a detection in COSMOS-Web. However, a discovery of systems with source redshifts $6 \lesssim z \lesssim 11$ in COSMOS-Web data – possible if our forecast is correct – would represent the highest redshift lensed sources ever discovered.

In Fig. 4 we show the distribution of the forecast unlensed source AB magnitude in the $F277W$ band, along with the distribution of the equivalent magnitude in the JAGUAR catalogue from which the sources were drawn. We choose the $F277W$ band as a representative example of the distribution of source magnitude in each of the four bands. From this figure we can see that the majority of the sources in the forecast are drawn from the brightest sources in the JAGUAR catalogue, meaning that the forecast population is not underestimated due to lack of depth in the JAGUAR catalogue. We can also compare our forecast to the distribution of source magnitudes in the LSST catalogue used for the source population by T. E. Collett (2015). For example, in the z -band of the LSST catalogue, 2.0 per cent of sources have magnitudes fainter than 28, whereas in our forecast for the $F115W$ filter of NIRC*am*,⁵ 6.8 per cent of sources are fainter than magnitude 28. This highlights the importance of our use of the JAGUAR catalogue instead of the LSST catalogue of the original *lenspop* forecasts, as without it we would have possibly missed the faintest sources discoverable in COSMOS-Web.

The comparative lack of lenses with large Einstein radii in our forecast is correlated with the features of the other distributions described above. The lenses expected to be detected in COSMOS-Web are low-mass galaxies, with 8.3 per cent occurring above $z = 2$. The Einstein radius of an isothermal lens is proportional to $D_{\text{ds}}/D_{\text{os}}$ (see equation 3), meaning that for a given source redshift, high-redshift lenses have smaller Einstein radii. Given that large Einstein radii lenses are typically easy to identify via visual inspection, they are unlikely to have been missed in the COSMOS-Web data despite their expected rarity. As we will discuss further in Section 5.1, this expectation is supported by the distribution of the observed Einstein radii of the 17 spectacular lenses, which is skewed towards larger values.

⁵We select the LSST z -band for direct comparison to our forecast as it is the closest in wavelength to the $F115W$ filter of NIRC*am*.

5 DISCUSSION

5.1 Comparison with properties of the observed data

Besides the comparison of the abundance of strong lensing in COSMOS-Web with our forecast, the lens modelling performed by J. Nightingale et al. (2025) allows us to compare properties of our forecast lenses with properties of the observed strong lenses.

5.1.1 Einstein radii

In Fig. 5 we show the forecast distribution of Einstein radii (yellow histograms) compared to the distribution of this parameter obtained from modelling the observed strong lenses (black unfilled histograms). Our forecast yields a single value for this parameter, computed using equation (3); the lens modelling of J. Nightingale et al. (2025) was performed on a single band per candidate lens, using whichever band was judged to have the brightest source light. The Einstein radius was computed as the area within the critical curve for the inferred SIE mass model for the given band.

From Fig. 5, we can see that, as expected, the 17 spectacular lenses do not contain any very small Einstein radius systems, and larger values of this quantity are overrepresented compared to the forecast distribution. In contrast, once the additional lenses identified in the second round of visual inspection are added, the distributions of the Einstein radii become more similar to the forecast distribution, although the number of very small Einstein radius systems is still less than predicted by our forecast.

Assuming the *lenspop* forecast to be perfectly accurate, this result motivates re-inspection of the COSMOS-Web data, with a focus on lower mass galaxies, such as those observed by Y. Asada et al. (2023), V. Gelli et al. (2023), and X. Lin et al. (2023), which will in turn have smaller Einstein radii. A less stringent magnitude cut in the preparation of the data to be inspected may result in more low-mass galaxies and therefore more lenses with small Einstein radii being identified; though, since many of these Einstein radii are likely to be close in angular size to the PSF FWHM of NIRC*am*, with the very smallest being unresolved in the longer wavelength bands and approaching the PSF FWHM in the shorter wavelength bands, they will be challenging for visual inspectors to correctly identify (K. Rojas et al. 2023). However, they may be an ideal target for machine-learning-led searches. The overrepresentation of larger Einstein radius systems in the observed data compared to the forecast can be understood from the fact that the observed data contains a number of group-scale systems (strong lensing by a group of galaxies), which is not taken into account in the *lenspop* model.

We perform the two-sample Kolmogorov–Smirnov (KS) test (P. Sprent 1998) to determine the similarity of the forecast and observed distributions. The two-sample KS test measures the maximum vertical distance between cumulative distribution functions associated with two samples of data. Our null hypothesis is that the observed Einstein radii are drawn from the same distribution as that of our forecast sample. We run the KS test on all three observed data sets: the 17 spectacular lenses, the 59 high confidence lenses, and the 116 high and medium confidence lenses (both of the latter including the 17 spectacular lenses).

We find that the null hypothesis is rejected with 95 per cent confidence in the case of the spectacular lenses, with a p -value of 0.006, but is not rejected for the larger samples, with p -values of 0.56 and 0.78, respectively. The KS statistic decreases from 0.10 for the

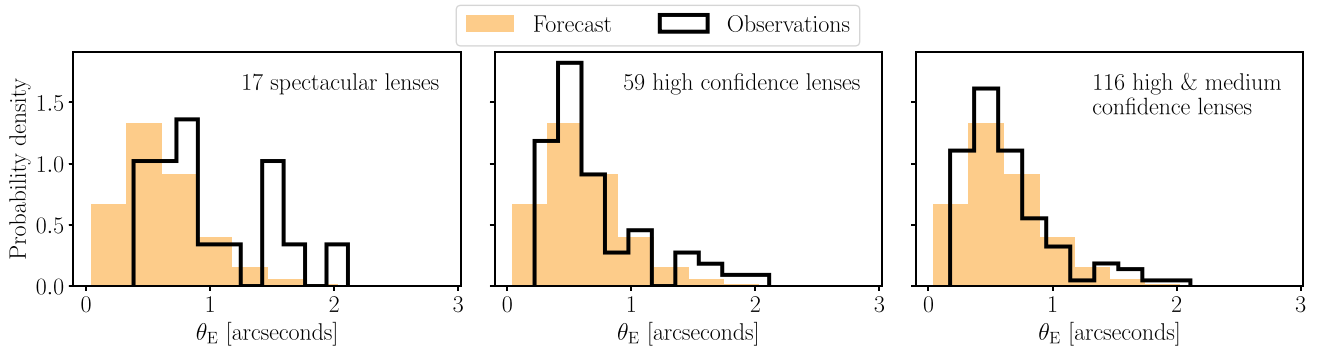


Figure 5. Our forecast distribution of Einstein radii (solid yellow) compared to the observed lenses (unfilled black). *Left:* the 17 spectacular lenses. *Middle:* the 59 high confidence lenses. *Right:* the 116 high and medium confidence lenses.

high confidence lenses to 0.06 for the high plus medium confidence lenses, implying that the distribution of Einstein radii in the 116 high plus medium confidence lenses is most similar to that of the forecast distribution.

A drawback of the KS test is that it is most sensitive to differences in the centre of distributions rather than in the tails (N. M. Razali & Y. B. Wah 2011). Since both our forecast and observed distributions of Einstein radius are clearly positively skewed, we can instead use the Anderson–Darling (AD) test (T. W. Anderson & D. A. Darling 1952), which is more sensitive to differences in the tails, to compare the distributions. For the distribution of the Einstein radii of the spectacular lenses, the null hypothesis that this sample is drawn from the same underlying distribution as the forecast sample is rejected at 95 per cent confidence by the AD test, with a p -value of 0.001. However, as with the KS test, the null hypothesis is *not* rejected at this confidence level for the distributions of the Einstein radius in the high confidence and high and medium confidence lenses, with p -values of 0.16 and 0.25, respectively. The AD test statistics for the latter two cases are 0.75 and 0.036, implying that, under the AD test, the 116 high confidence lenses are also the most similar to the forecast distribution. We summarize all the results of our statistical tests in tables in Appendix A.

It must be borne in mind that the overall similarity between the forecast and observed Einstein radius distributions is driven in part by how the Einstein radii are computed; in the forecast, all lenses are assumed to have SIE mass profiles, and all the candidate lenses in the observed data were fit with SIE profiles, meaning that we are comparing like with like. However, future modelling of the candidates may find that more complex mass profiles are preferred for some of the lenses, making comparison with the `lenspop` forecast less robust.

5.1.2 Lens redshifts

In Fig. 6 we show the distributions of the lens redshifts in our forecast (green histograms) and observed (unfilled black histograms) populations. We use the estimated photometric redshifts of the strong lenses for our comparison; these redshifts have been checked against spectroscopic measurements where available, and are in good agreement with those, albeit with some scatter above $z = 1$. We choose to use the photometric redshifts as they extend to $z \sim 4$, while the spectroscopic redshifts only reach $z \sim 2$, below the redshift limit imposed in our forecast. More details on the redshifts of the observed strong lenses can be found in J. Nightingale et al. (2025);

importantly, there is no correlation between lens candidate ranking and lens redshift.

From Fig. 6, we can see that, in general, the predicted distribution does not fully capture any of the observed populations at low redshifts. The observed lens redshift distribution in all cases peaks below $z = 1$, and slightly below the peak of the forecast distribution in each case. For the 116 high and medium confidence lenses, it is clear that the high-redshift tail of the distribution is more extended in the observed data compared to the forecast. In particular, there is a single lens, COSJ100151+022347, whose photometric redshift is estimated at $z = 3.58$, well above the limit we imposed in our forecast of $z = 3$. This lens received a grade of 6/12, based on two ‘A’ grades, two ‘B’ grades, one ‘U’ grade, and one ‘X’ grade, and does not have a spectroscopic redshift. If the photometric redshift of this and the other five lenses above $z = 2$ were to be confirmed with spectroscopy, this would be indicative of the need to rerun the forecast with an even higher upper limit for the lens redshift distribution.

To explore the question of high-redshift lenses further, in the top row of Fig. 7 we show the five systems with the highest photometric lens redshifts amongst the 116 high and medium confidence lenses, excluding two of the spectacular lenses described in G. Mahler et al. (2025). One of these two lenses is the COSMOS-Web ring, which has been described in detail in W. Mercier et al. (2024), and at the time of discovery was the highest known lens redshift at $z = 2.02$. The lenses and their associated photometric redshifts are listed in Table 2. On the bottom row of the same figure, we show a selection of five simulated systems with the highest lens redshifts in our forecast.

From this figure, we can see that, for the simulated lenses, the colours of the lens and source are readily distinguishable, and clear arc features are apparent in most of the systems despite their high redshift and consequent faintness. This lends weight to the idea that more strong lenses at $z > 2$ are identifiable in *JWST* imaging. More variability is seen in the colours and morphology of the observed candidate lenses than in the simulated ones; this is due to the simplicity of the lens and source models employed in our `lenspop` forecast.

Furthermore, we can use the AD test to quantify the differences between the forecast and observed distributions. In this case, the null hypothesis of the two samples being drawn from the same underlying distribution is not rejected in the case of the 17 spectacular lenses, but is rejected for both the 59 high confidence and 116 high and medium confidence lenses, at 95 per cent confidence. The full test results are shown in Appendix A.

The computation of the forecast lens redshifts relies purely on cosmology, with no astrophysics involved. This implies that the

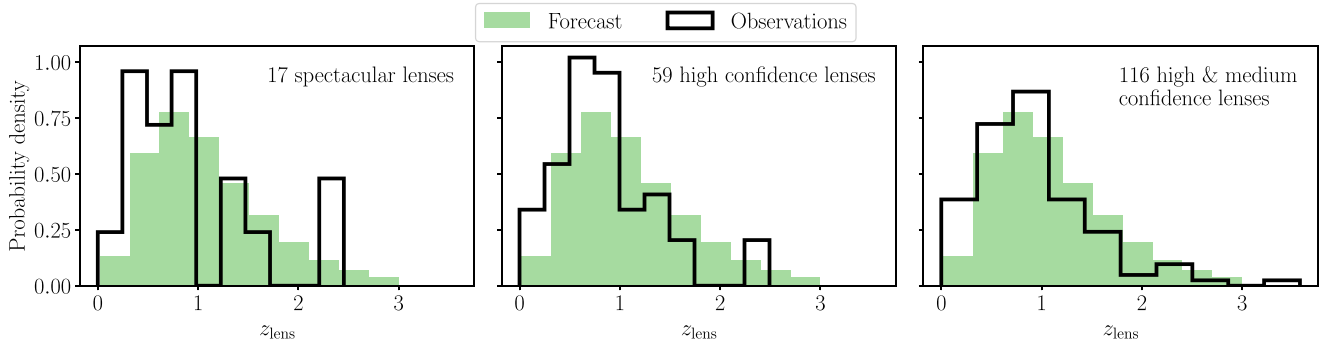


Figure 6. Our forecast distribution of lens redshifts (solid green) compared to the observed lenses (unfilled black). *Left:* the 17 spectacular lenses. *Middle:* the 59 high confidence lenses. *Right:* the 116 high and medium confidence lenses. For the observed lenses we show the photometric redshifts.

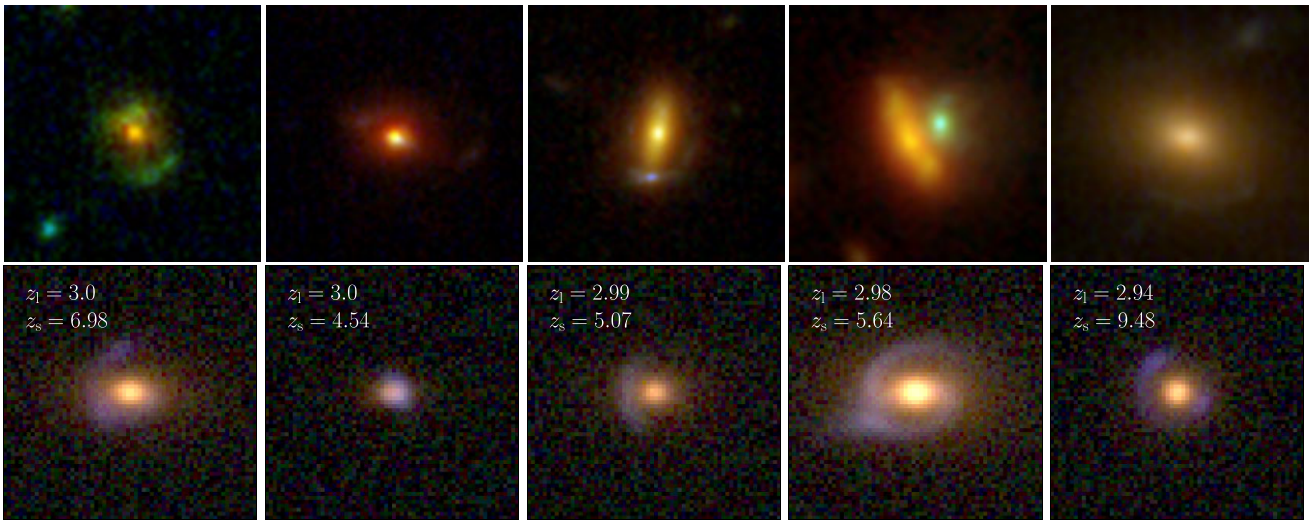


Figure 7. Top: RGB images of the five systems with the highest photometric lens redshifts ($z_{\text{lens}} > 1.8$) from the 116 high and medium confidence lenses, excluding two of the spectacular lenses presented in G. Mahler et al. (2025). Bottom: a selection of five systems with the highest redshifts in our forecast. The lens and source redshifts are denoted in the top left of each panel. Colours are representative and are produced by combining the data simulated in three different NIRCcam bands ($F150W$, $F277W$, and $F444W$).

Table 2. Details of the five lenses with the highest photometric redshifts in the COWLS catalogue, excluding two spectacular lenses presented in G. Mahler et al. (2025).

Lens	Score	θ_E	z_{lens}
COSJ100151+022347	6/12	0.33 arcsec	3.58
COSJ100120+022435	6/12	0.86 arcsec	2.54
COSJ100122+022132	9/12	0.52 arcsec	2.49
COSJ100120+015834	7/12	0.62 arcsec	2.44
COSJ100134+022037	7/12	0.52 arcsec	1.82

differences between the forecast and observed distributions stems from a selection effect in the observed data, rather than a fundamental physical difference in the creation of the two populations. For example, lenses at lower redshifts are likely to be brighter than those at higher redshifts, and therefore easier to identify during visual inspection. This may be a contributing factor in the discovery of more lenses at $z < 1$ than expected from the forecast. Confirmed spectroscopic redshifts are needed for all lenses before definitive conclusions can be drawn.

5.1.3 Lens magnitudes

Next, we can compare the lens magnitudes in our forecast and observed populations of lenses. In Fig. 8, we show the distribution of the AB magnitude of the lens galaxies in each NIRCcam band in our forecast (red histograms) and in the 59 high confidence lenses (unfilled black histograms). For brevity we choose to focus on the 59 high confidence lenses, but the overall behaviour of the distributions in each band is the same for both the 17 spectacular lenses and the 116 high and medium confidence lenses.

From Fig. 8, we can see that the distributions of the observed lens magnitudes tend to peak at brighter magnitudes than those of the forecast. Furthermore, the difference between the observed and forecast distributions increases with band wavelength. The AD test rejects the hypothesis that the two distributions were drawn from the same underlying distribution at 95 per cent confidence for all of the bands.

There is also an indication of bimodality in the distribution of the observed lens magnitudes in each band. To investigate this further, we performed a kernel density estimation to obtain the smoothed probability density function from the binned samples, which allowed

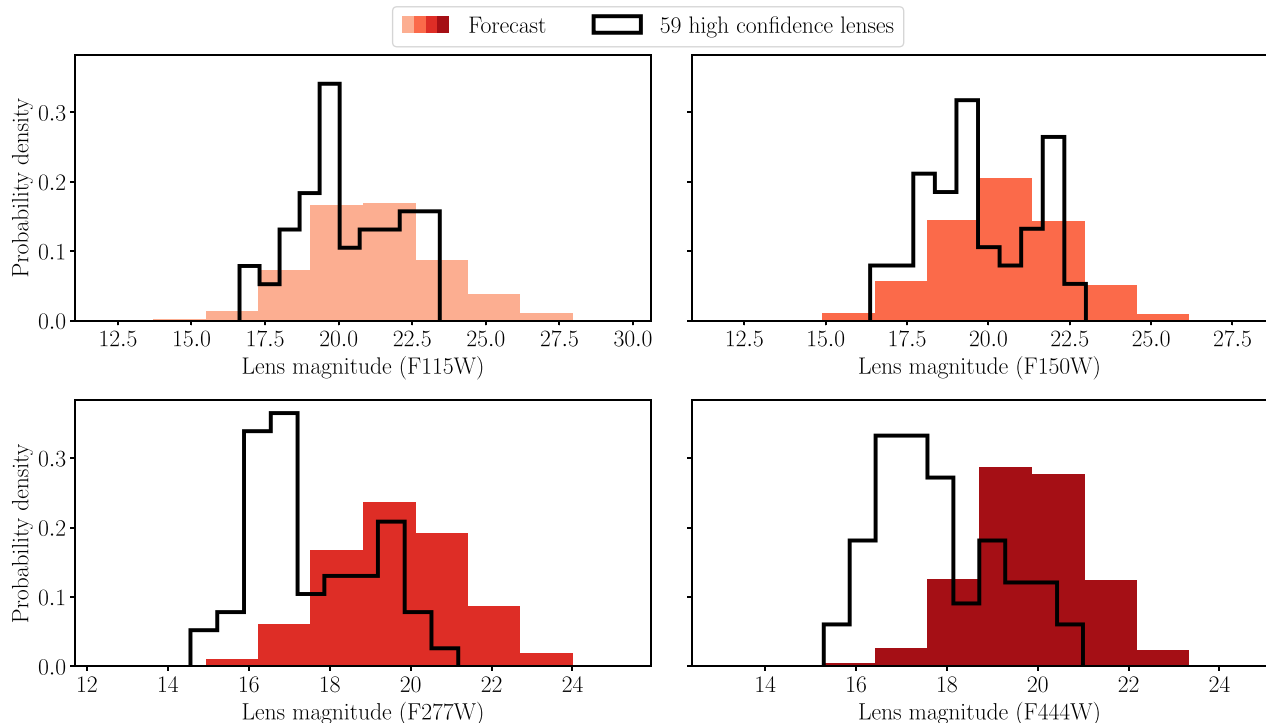


Figure 8. Our forecast distribution of lens magnitudes in each band (solid red) compared to the 59 high confidence observed lenses (unfilled black).

us to identify that there is only a single mode in the distribution in each band (B. W. Silverman 1981). We also carried out the dip test for deviation from a unimodal (uniform) distribution (J. A. Hartigan & P. M. Hartigan 1985), finding no significant difference from a unimodal distribution for any of the bands. From this we conclude that any visual indication of bimodality is an artefact of the chosen binning scheme for the histograms and does not have an underlying physical cause.

In general, we may conclude that the observed lenses are brighter in the *F277W* and *F444W* bands than would be expected from the *lenspop* forecast. This finding has three possible underlying causes. First, the estimation of the lens magnitudes by J. Nightingale et al. (2025) may have picked up a systematic bias during the lens modelling. However, if this were the case, we would expect all the bands to be affected equally, rather than what we actually see, which is that the effect is more pronounced in the long-wavelength bands.

Secondly, if the *lenspop* model is considered to be perfectly accurate, our finding could indicate that there remains an undiscovered population of faint lenses in the COSMOS-Web data. This is possible, since a magnitude cut was imposed before the beginning of the visual inspection, with only objects brighter than magnitude 23 in *F277W* being considered (which is clearly apparent in the distributions of the observed lens magnitudes in Fig. 8). This possibility could easily be addressed by carrying out another visual inspection of the data with a more optimistic magnitude cut, for example examining all objects brighter than magnitude 26.

Thirdly, we can consider the alternative possibility that the visual inspection is perfectly accurate, and that all lenses that exist in the COSMOS-Web data have been identified as such (this is already glossing over the truth, as we do not present or discuss any lens candidate that scored less than 6/12 in this work, an arbitrary but potentially consequential cut-off). In that case, it is the accuracy of the *lenspop* model that must be examined.

Lens magnitudes in the *lenspop* model are computed using the fundamental plane from J. B. Hyde & M. Bernardi (2009a, b) derived from SDSS data. The fundamental plane is an empirically derived relation between galaxy properties (S. Djorgovski & M. Davis 1987; A. Dressler et al. 1987), and can be expressed as $R \propto \sigma_v^\alpha / I^\beta$, where R is the effective radius of the light profile of the galaxy, σ_v is the velocity dispersion, I is the surface brightness, and α and β are parameters needing to be fit using a sample of observed galaxies.

Again, considering that the visual inspection is perfectly accurate, our result suggests that magnitudes of lens galaxies observed in COSMOS-Web may not follow the same fundamental plane relation as that used in *lenspop*, from galaxies catalogued in SDSS. This is in line with the fact that the spectroscopic redshifts from SDSS used to compute the fundamental plane do not extend above $z = 0.35$ (J. B. Hyde & M. Bernardi 2009a), in stark contrast to our lens galaxy photometric redshift distribution, which peaks at around $z = 1$ and extends to above $z = 3$.

While we cannot investigate this hypothesis in detail without spectroscopic redshifts and measurements of the velocity dispersions for the entire COWLS catalogue, we can make a related comparison, between a stellar mass distribution computed from our forecast velocity dispersions, and the stellar masses of the observed lenses (obtained via photometry, see J. Nightingale et al. 2025).

J. B. Hyde & M. Bernardi (2009a) provide the best-fitting coefficients for linear and quadratic pairwise scaling relations between the various galaxy properties that are linked via the fundamental plane (table 1 of that work). In particular, they provide linear and quadratic fits for the velocity dispersion as a function of stellar mass. We invert the quadratic relation to obtain stellar mass as a function of velocity dispersion. This allows us to compute a distribution of stellar masses from our forecast velocity dispersions, which can then be compared with the observed data. We show this comparison in Fig. 9.

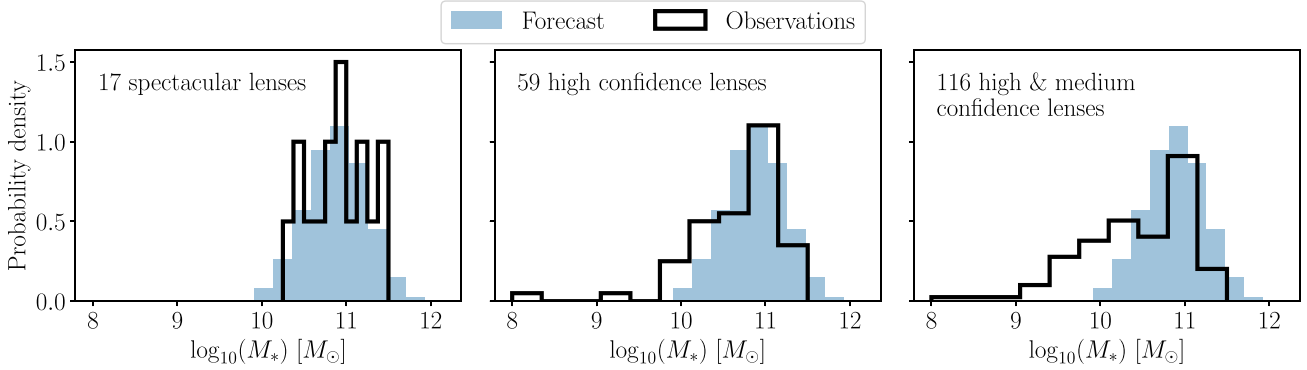


Figure 9. Our forecast distribution of lens stellar mass (solid blue) compared to the observed lenses (unfilled black). Left: the 17 spectacular lenses. Middle: the 59 high confidence lenses. Right: the 116 high and medium confidence lenses.

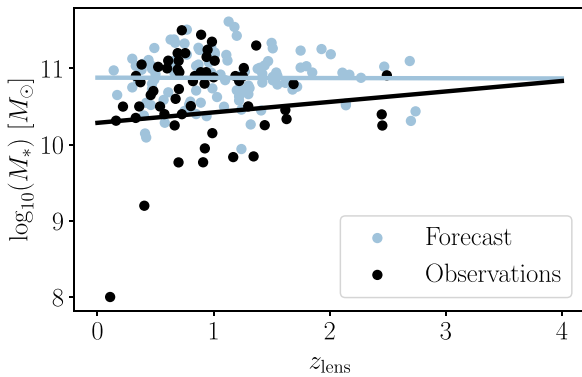


Figure 10. A scatter plot of lens redshift versus lens stellar mass for our forecast population (solid blue) and 59 high confidence observed lenses (unfilled black).

From this figure, we can see that the forecast and observed distributions of stellar mass match well for the 17 spectacular lenses, both peaking at around $M_* = 10^{11} M_\odot$. As we increase the sample size, we see that lenses with stellar masses well below that predicted by our forecast begin to appear, with a very large low-mass tail in the distribution for the high and medium confidence lenses.

One interpretation of this finding is that the velocity dispersion function produced by Y.-Y. Choi et al. (2007) cannot correctly predict lens stellar mass at high redshifts, resulting in our forecast predicting too many massive galaxies at high redshifts, $z > 2$. A simple check of this hypothesis is to plot our forecast and observed lens redshifts against stellar mass, to see if the correlations are similar across the two populations.

In Fig. 10, we show a scatter plot of lens redshift against lens stellar mass for our forecast and observed populations, showing the 59 high confidence lenses as a representative example. We also plot the lines of best fit to these data computed via linear regression. From this figure, we can see that, at $z < 2$, i.e. at *low* redshifts, our forecast contains more massive galaxies than seen in the data. Conversely, as redshift increases, the two populations appear to be in better agreement. However, since there are so few observed lenses at $z > 2$, it is not appropriate to draw strong conclusions about the true agreement of the two populations.

It was also pointed out in COWLS Paper I (J. Nightingale et al. 2025) that the lens candidate score and lens stellar mass are

correlated, with higher scoring candidates tending to be the most massive lenses. If we *do* trust the extrapolation of the Y.-Y. Choi et al. (2007) velocity dispersion function to $z > 2$, our finding may hint that the low-scoring low-mass candidates are not true strong lenses, as they are not present in the forecast. We reiterate that this comparison itself relies on the fundamental plane relations fit by J. B. Hyde & M. Bernardi (2009a). A more robust test of how well various velocity dispersion functions match the data will only be possible with spectroscopy for our lens candidates.

5.1.4 Unlensed source magnitudes

Lastly, we can compare the unlensed source magnitudes in our forecast and observed populations of lenses, i.e. the magnitudes of the sources in the source planes. In Fig. 11, we show the distributions of this quantity in each of the NIRCcam bands for the forecast lenses (red histograms) and the 59 high confidence observed lenses (black unfilled histograms). As with the lens magnitudes, we choose this set of lenses as a representative example. From this figure, we can see that we have relatively good agreement in the overall shape and peak location of the source magnitude distribution in each band. However, the AD test rejects the null hypothesis of similarity between the forecast and observed distributions. In particular, we can see that, as the wavelength increases, the observed lenses contain an overabundance of faint sources and a lack of bright sources compared to the forecast.

However, the forecast source magnitudes are not computed by *lenspop*, but are instead taken directly from the JAGUAR source catalogue. The explanation of the difference between the observed and forecast distributions therefore again has more than one possible explanation. First, the magnitudes in the JAGUAR catalogue may be inaccurate, but this is unlikely as the mock galaxies in the catalogue were produced from an empirical model based solely on observational data (C. C. Williams et al. 2018). One potentially relevant factor in the production of the JAGUAR mock is the use of the G. Bruzual & S. Charlot (2003) stellar population synthesis model via BEAGLE; differences in such models have been shown to propagate to the galaxy SEDs (S. L. Newman & others 2025). Nevertheless, unless the COWLS strong lens sources are a particularly unusual sample of galaxies, we can expect their properties to match well with the galaxies in the JAGUAR mock catalogue.

Secondly, the computation of the unlensed source magnitudes for the observed lenses may carry some inaccuracies. The observed lensed sources were reconstructed by J. Nightingale et al. (2025)

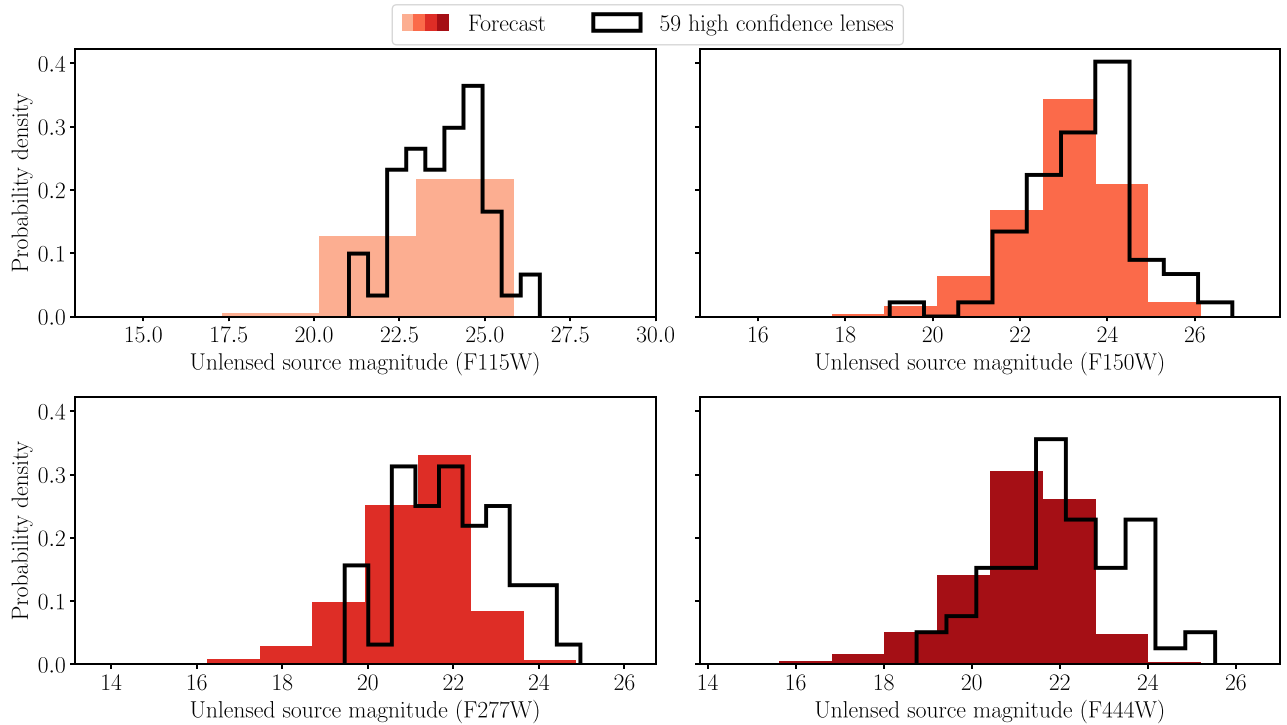


Figure 11. Our forecast distribution of unlensed source magnitudes in each band (solid red) compared to the 59 high confidence observed lenses (unfilled black).

using a multi-Gaussian expansion and Voronoi mesh approach (Q. He & others 2024). The magnitudes are computed from the resulting reconstructed source, and will naturally depend on the inferred lens model and final source mesh configuration. How much the modelling choices affect the source reconstruction cannot be examined in detail until different lens and source models have been applied to the COWLS catalogue.

A final contribution to the differences between the forecast and observed distributions may be a selection effect, but we would generally expect the visual inspection to select for preferentially brighter sources rather than fainter ones, as bright images are easier for human visual inspectors to identify and examine than faint ones. Since we see almost the opposite effect here, this may lead us to conclude that some bright sources have been missed in the visual inspection, while some of the (likely weaker) candidate lenses with faint sources may be false positives.

5.2 Comparison with other forecasts for COSMOS-Web

5.2.1 Holloway et al. (2023)

P. Holloway et al. (2023) performed a forecast for a number of *JWST* surveys, including COSMOS-Web. Instead of relying on simulated lenses, they used mock galaxies from the JAGUAR catalogue as both sources and lenses. Furthermore, they modified the effective radii of the deflectors to better match the Sérsic indices in the catalogue, and assigned halo masses from the DREaM catalogue (N. E. Drakos & others 2022) to the JAGUAR galaxies using the subhalo abundance matching technique. This work predicted ~ 65 strong lenses across all bands in COSMOS-Web, compared to the 107 we forecast using *lenspop* and the 116 high and medium confidence lenses in the COWLS catalogue from the observed COSMOS-Web data.

In order to investigate the source of the difference between the forecasts, we first reran our *lenspop* analysis using the modified JAGUAR catalogue produced by P. Holloway et al. (2023) as the source population. This resulted in the same number of strong lenses predicted as well as distributions for the lens system properties consistent with our original findings, meaning that the difference between the results of this work and that of P. Holloway et al. (2023) is not driven by the differences in the source catalogues. However, the P. Holloway et al. (2023) prediction of 3.1×10^9 strong lenses in the Universe is larger than our finding of 2.8×10^8 lenses. Combined with the smaller number of strong lenses predicted by P. Holloway et al. (2023), this implies that the difference in the number of detectable lenses in COSMOS-Web is likely the result of differences in how the survey specifications were applied to the idealized lens population.

Examining the properties of the forecast lens populations, we find that < 1 per cent of our forecast lenses are fainter than the magnitude cut applied by P. Holloway et al. (2023) in the *F115W* band, with no lenses in any of the other three bands being fainter than their respective cuts. The smallest Einstein radius of a lens in the P. Holloway et al. (2023) forecast lens catalogue is 0.07 arcsec, while our forecast predicts 0.01 per cent of lenses having Einstein radii between 0.04 and 0.07 arcsec. The source and lens redshift distributions predicted by P. Holloway et al. (2023) are very similar to those produced by *lenspop*. The discrepancy between the predictions therefore cannot be explained by *lenspop* predicting more fainter or smaller lenses in the population than the method of P. Holloway et al. (2023).

Lastly, we note that P. Holloway et al. (2023) attempted to validate their forecast by comparing their predictions for lensing abundance in the COSMOS field with previous observations of the field made using *HST*. In particular, C. Faure et al. (2008) found 67 candidate

strong lenses, compared to the prediction by P. Holloway et al. (2023) of 34. N. Jackson (2008) identified two certain, one highly probable and 112 further possible but unlikely lens systems, leading to 160 candidates in total; applying the same search criteria, P. Holloway et al. (2023) found 80. Taken together, these results could indicate a potential systematic underestimation of strong lensing abundance in their forecasts. This may be supported by the discovery of 116 high and medium confidence strong lenses in COSMOS-Web, but if only the 59 high confidence strong lenses are considered, the forecast of P. Holloway et al. (2023) is in excellent agreement with the data.

5.2.2 Ferrami & Wyithe (2024)

G. Ferrami & J. S. B. Wyithe (2024) built a model for strong lensing statistics which is more akin to the `lenspopt` method, using the assumption of all deflectors being SIEs to calculate the strong lensing optical depth. However, instead of relying on a mock galaxy catalogue to simulate the source population, they computed the number density of sources lensed by a given deflector analytically.

Their work tested the effect of different velocity dispersion functions on the number of lenses detectable by *Euclid*, with their fiducial model being that of C. A. Mason et al. (2015), which is redshift-dependent. They compared properties of their forecast sample with one produced using the Y.-Y. Choi et al. (2007) velocity dispersion function employed in this work (equation 2). They found that the fiducial C. A. Mason et al. (2015) model results in deflector and source redshift distributions, which peak at a slightly lower redshift compared to the Y.-Y. Choi et al. (2007) model, and a distribution of Einstein radii, which peaks at a higher value compared to the Y.-Y. Choi et al. (2007) model.

G. Ferrami & J. S. B. Wyithe (2024) did not produce a forecast for COSMOS-Web using the Y.-Y. Choi et al. (2007) velocity dispersion function, but using their fiducial C. A. Mason et al. (2015) model they predicted 46 lenses detectable in the *F115W* filter, 74 in the *F150W* filter, and 121 in the *F277W* filter, all assuming lens light subtraction. These numbers are broadly consistent with our prediction of 107 lenses across all bands. They did not produce a forecast for the *F444W* filter.

However, the forecast distributions of deflector and source redshifts, velocity dispersion, and Einstein radii in each band do not match those predicted with our `lenspopt` model. In particular, the source redshift distribution is consistent with zero at $z \sim 5$, unlike our finding of source redshifts up to $z \sim 11$, likely due to our use of the JAGUAR catalogue, which contains sources up to $z \sim 15$.

Furthermore, the tail of our distribution of Einstein radii reaches about 2 arcsec, compared to around 3 arcsec in G. Ferrami & J. S. B. Wyithe (2024). As discussed in Section 4.3, these differences are correlated, as Einstein radii are larger for lenses at lower redshifts for a given source redshift. Given the similarity in the forecasting methods, the cause of these differences is likely the different velocity dispersion function used by G. Ferrami & J. S. B. Wyithe (2024), which results in their forecast containing a larger number of massive galaxies than ours.

5.3 Selection effects

All of the comparisons we have made so far are founded on the assumption that the observed and forecast lenses were selected in exactly the same way. While the criteria of T. E. Collett (2015), which we also employ (see Section 3.3), are designed to select for lenses in the same way that visual inspectors decide whether an object exhibits lensing features, the correspondence is not one to one.

For example, the first criterion that a candidate must meet to be deemed a lens in our forecast is that multiple images of the source must be produced. This criterion is baked in to the calculation of the strong lensing optical depth. But in the second round of visual inspection described in J. Nightingale et al. (2025), inspectors had the option to grade a lens as ‘S’, meaning that only a singly imaged lensing feature was visible.⁶ An ‘S’ grade contributed one point towards the total score of a candidate; 47 candidates in the complete COWLS catalogue received at least one ‘S’ grade (11.2 per cent).

A full forward modelling of the selection function to understand the bias of our lenses with respect to the parent galaxy population goes beyond the scope of this paper. Such efforts have been made for other lens populations, see e.g. A. Sonnenfeld (2022, 2024). We here describe the various contributing factors to the selection of the forecast and observed lenses and provide a brief discussion of how they may have affected the results described in this work.

5.3.1 Selection of the forecast lenses

As a reminder, the criteria we use to call a forecast lens detectable are that multiple images are produced, the image and counter-image must be resolved, and the total source magnification and signal-to-noise ratio must exceed a given threshold. To investigate the effects of these choices on the detectable lens population, we can modify the chosen thresholds.

Keeping the other criteria fixed, we first increased the signal-to-noise ratio required to call a lens detectable from 20 to 40; consequently, the number of detectable lenses decreased to 55, almost exactly half the original number of 107. The distribution of the Einstein radii of these lenses does not contain the very smallest Einstein radii lenses of our original forecast, as expected from the fact that lenses with smaller Einstein radii will generally have lower signal-to-noise ratios. Furthermore, the distribution of velocity dispersions does not contain the lowest-mass galaxies, and the source redshift distribution misses the vast majority of sources above $z = 6$; the unlensed source magnitude distributions miss the faintest sources. In contrast, the distribution of lens redshifts and lens magnitudes remain consistent with our original result.

Increasing the threshold for the minimum source magnification from 3 to 4 results in a forecast of 68 detectable lenses, with the distributions of all the parameters remaining close to the original results. Increasing the resolution threshold at which a lens is determined to be detectable, from $\theta_E^2 > r_s^2 + (s/2)^2$ to $\theta_E^2 > r_s^2 + s^2$, results in 104 lenses, with distributions again consistent with the original results. This indicates that the signal-to-noise ratio threshold is by far the most important criterion for determining strong lensing detectability, excluding the requirement that multiple images of the source must be produced.

5.3.2 Selection of the observed lenses

The first and most obvious explicit selection the COWLS catalogue was subject to is the cut in magnitude that was imposed to choose objects for visual inspection. Only those objects brighter than AB magnitude 23 in the *F277W* filter were inspected; while our forecast predicts that 1.5 per cent of the COWLS lenses should lie above this magnitude, by construction we find none. Furthermore, as we have seen, our forecast predicts significantly more fainter lenses in general

⁶We note that in many cases, the presence of the counter-image was revealed by the lens modelling conducted in that work.

than what is found in COWLS: 41 per cent of the forecast lenses are fainter than magnitude 20 in the *F227W* in the forecast, compared to just 5 per cent in COWLS.

Another important selection made in the preparation of this work was the choice to only consider those COWLS candidates with scores of 6/12 or more. The forecast, by construction, is completely comprised of true lenses which should all score 12/12 in an inspection; limiting our comparison with the forecast to only those COWLS candidates that scored 12/12 (all 17 spectacular lenses from G. Mahler et al. (2025) plus two lenses from J. Nightingale et al. (2025)) would not have been justified from a statistical standpoint. A way to quantify the selection effect of the grading scheme would be to have visual inspectors examine simulated images of the forecast lenses, like those shown in Figs 1 and 7, to see how many ostensibly 12/12 lenses are actually graded as such. This would still not be a perfect test, as the SIE lens model in the forecast tends to result in regular, unambiguous lensing features.

The final, and implicit, selection effect is that of human fallibility. The first half of the first round of visual inspection in COWLS was conducted by four inspectors, and the second half by five; the second round was conducted by six inspectors. When analysing the ability of a visual inspection team to correctly identify lenses, K. Rojas et al. (2023) found that the minimum number of inspectors needed to mitigate noise is six. Future visual inspections of the COSMOS-Web field for strong lenses could be aided by increasing the number of inspectors or by using machine-learning search methods.

6 CONCLUSIONS

In this work we made a forecast for the abundance and properties of strong gravitational lenses detectable in the COSMOS-Web survey, a survey of the 0.54 deg² COSMOS field using the NIRCам and MIRI instruments aboard *JWST*. We used the `lenspop` model of T. E. Collett (2015) to make our forecast, after implementing several modifications to improve its suitability as a forecasting tool for COSMOS-Web. In particular, we changed the mock galaxy catalogue from which the source population is drawn to the JAGUAR catalogue, and added the NIRCам filter data and COSMOS-Web survey specifications. This led to a prediction of 107 lenses detectable across all bands, or a minimum of 14 detectable in the *F115W–F150W*, *F115W–F277W*, and *F115W–F444W* difference images.

Our forecast predicted the strong lenses in COSMOS-Web are dominated by small Einstein radius systems ($\theta_E < 1$ arcsec) with a distribution of velocity dispersions that peaks at $\sigma_v = 200$ km s^{−1}. The distribution of lens redshifts is predicted to peak at $z = 1$, with a significant tail up to $z = 3$. The distribution of source redshifts peaks at $z = 3.2$, with a non-negligible fraction of the sources lying at $z > 6$, directly probing the entire epoch of reionization. The highest source redshift in our forecast population is $z = 11.6$.

We then compared the properties of the forecast strong lenses with that of the observed data from the COWLS catalogue. Strong lenses were identified in COSMOS-Web by two rounds of visual inspection; the first round found 17 strong lenses and the second round, which included modelling of 419 additional candidates, found 99 lens candidates that scored at least 6/12; there were 42 lens candidates that scored at least 8/12. This yielded a total catalogue of 59 high confidence strong lenses, or 116 high and medium confidence strong lenses.

The distribution of Einstein radii in the 17 spectacular lenses does not match that of our `lenspop` forecast, with both the KS and AD tests rejecting the null hypothesis that these samples were drawn from the same underlying distribution at 95 per cent confidence. This

is driven by a lack of lenses with small Einstein radii in the observed sample compared to the forecast. However, the null hypothesis is not rejected by these tests when comparing the distribution of Einstein radii for the 59 high confidence lenses or the 116 high and medium confidence lenses.

In contrast, the distributions of the lens photometric redshifts, lens magnitudes, and unlensed source magnitudes are not consistent with the forecast distributions for these quantities. In particular, there are fewer lenses above $z = 1$ in the observed sample than in the forecast, and the lens magnitude distributions peak at brighter magnitudes than expected from the forecast. This may be due to the computation relying on a fundamental plane relation fit with SDSS data in which the galaxies are at much lower redshifts than the typical COWLS lens galaxy.

The unlensed source magnitudes appear to be slightly fainter in the observed sample than would be expected from the forecast, particularly in the long-wavelength bands; the use of the JAGUAR mock catalogue means that the forecast should be a good representation of galaxies observable in COSMOS-Web, even up to high redshifts. Modelling the COWLS sample with a variety of lens and source models may yield better agreement in the source magnitudes than is currently seen. A more in-depth comparison of the lens and source galaxies with the `lenspop` forecast will only be possible once velocity dispersion and spectroscopic redshift information are obtained, which will require follow-up observations from *JWST* to obtain the required precision at the redshifts in question.

It is also important to note that the `lenspop` model has a number of simplifications: first, the images are convolved with a Gaussian PSF rather than the more complex PSF of *JWST*; secondly, that all the lenses are modelled as SIEs, meaning that the images produced are typically simple rings, arcs, or multiple images, unlike realistic lenses; and lastly that the weak lensing by line-of-sight objects (i.e. ‘external shear’) is not taken into account. However, it is unlikely that these assumptions significantly impact the forecast abundance of strong lenses. As described in Section 5.3, the signal-to-noise ratio plays the largest role in the detectability, and signal-to-noise ratio is not strongly correlated with the complexity of lensed image morphology.

Lastly, we reiterate that the MIRI observations of COSMOS-Web have not yet been inspected for strong lenses. Made using the single *F770W* filter, the MIRI observations cover a smaller sky area, 0.19 deg², than those using NIRCам, but we may expect another handful of lenses to be detected in these data. This is because MIRI observes even longer wavelengths than NIRCам, which will reveal even redder and potentially higher redshift systems than those in the current COWLS catalogue, akin to how NIRCам has revealed very red lensed sources that are invisible in *HST* imaging (G. Mahler et al. 2025). The MIRI-bright lenses are likely to have Einstein radii comparatively larger than the majority identified in the NIRCам data, due to the lower spatial resolution of MIRI, which means that smaller Einstein radius lenses are less likely to be resolved.

In conclusion, we have shown that the `lenspop` strong lens forecast model can be successfully adapted for *JWST* observations, and that the abundance of strong lenses and the distribution of the Einstein radii of observed lenses in COSMOS-Web are in good agreement with `lenspop` predictions. This implies that our choice of source galaxy population, the JAGUAR mock catalogue, is a good match to strong lens sources from the current day to high redshifts, $0 \leq z \leq 11$, spanning 92 per cent of the Universe’s history. Furthermore, it supports the idea that the majority of strong lenses can be described by elliptical isothermal mass profiles.

However, the distributions of the magnitudes of the observed deflexors and sources do not align as well with those from our forecast, with lenses in the observed data being systematically brighter than expected from the forecast. This result motivates detailed follow-up of the entire COWLS catalogue to obtain spectroscopic lens and source redshifts and lens velocity dispersions, in order to better understand the source of these discrepancies, as well as targeted re-inspection of the COSMOS-Web field to search for the undiscovered population of faint high-redshift lenses.

ACKNOWLEDGEMENTS

We are grateful to Tom Collett for useful discussions about `lenspop`, Giovanni Ferrami for discussions about G. Ferrami & J. S. B. Wyithe (2024), and to Philip Holloway for providing us with the modified version of the JAGUAR catalogue created for P. Holloway et al. (2023) along with his catalogues of forecast lens properties. We thank the referee for their detailed comments that helped improve the paper.

NBH is supported by a postdoctoral position funded by IN2P3. JWN is supported by an [STFC](http://dx.doi.org/10.13039/501100000271) (0:funding-source)/UKRI Ernest Rutherford Fellowship, Project Reference: ST/X003086/1. AA and QH acknowledge support from the European Research Council (ERC) through Advanced Investigator grant DMIDAS (GA 786910). GM and RM were supported in Durham by STFC via grant ST/X001075/1, and the UK Space Agency via grant ST/X001997/1. DS acknowledges the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004). OI acknowledges the funding of the French Agence Nationale de la Recherche for the project iMAGE (grant ANR-22-CE31-0007).

This research has made use of the SVO Filter Profile Service ‘Carlos Rodrigo’, funded by MCIN/AEI/10.13039/501100011033/ through grant PID2023-146210NB-I00. This work was made possible by utilizing the CANDIDE cluster at the Institut d’Astrophysique de Paris. The cluster was funded through grants from the PNCG, CNES, DIM-ACAV, the *Euclid* Consortium, and the Danish National Research Foundation Cosmic Dawn Center (DNRF140). It is maintained by Stephane Rouberol. The French contingent of the COSMOS team is partly supported by the Centre National d’Études Spatiales (CNES).

This work is based on observations made with the NASA/ESA/CSA *James Webb Space Telescope*. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–03127 for JWST. These observations are associated with program 1727.

DATA AVAILABILITY

The forecast data underlying this article were generated with the publicly available modified version of `lenspopgraphic`; the forecast lens catalogue is available on request. The observational data are publicly available in the COWLS repository [graphic](#).

REFERENCES

- Abbott T. M. C. et al., 2018, *ApJS*, 239, 18
 Acevedo Barroso J. A. et al., 2025, *A&A*, 697, A14
 Adelman-McCarthy J. K. et al., 2007, *ApJS*, 172, 634

- Aihara H. et al., 2018, *PASJ*, 70, S8
 Anderson T. W., Darling D. A., 1952, *Ann. Math. Stat.*, 23, 193
 Asada Y. et al., 2023, *MNRAS*, 523, L40
 Beckwith S. V. W. et al., 2006, *AJ*, 132, 1729
 Birrer S. et al., 2021, *J. Open Source Softw.*, 6, 3283
 Birrer S., Amara A., 2018, *Phys. Dark Universe*, 22, 189 <https://doi.org/10.1016/j.dark.2018.11.002>
 Birrer S., Refregier A., Amara A., 2018, *ApJ*, 852, L14
 Birrer S., Welschen C., Amara A., Refregier A., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 049
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
 Cabanac R. A. et al., 2007, *A&A*, 461, 813
 Cao X. et al., 2024, *MNRAS*, 533, 1960
 Casey C. M. et al., 2023, *ApJ*, 954, 31
 Casey C. M. et al., 2024, *ApJ*, 965, 98
 Chevallard J., Charlot S., 2016, *MNRAS*, 462, 1415
 Choi Y.-Y., Park C., Vogeley M. S., 2007, *ApJ*, 658, 884
 Collett T. E., 2015, *ApJ*, 811, 20
 Connolly A. J. et al., 2010, in Angeli G. Z., Dierickx P. eds, *Proc. SPIE Conf. Ser. Vol. 7738, Modeling, Systems Engineering, and Project Management for Astronomy IV*. SPIE, Bellingham, p. 773810
 De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, 366, 499
 de Vaucouleurs G., 1948, *Annales d’Astrophysique*, 11, 247
 Djorgovski S., Davis M., 1987, *ApJ*, 313, 59
 Drakos N. E. et al., 2022, *ApJ*, 926, 194
 Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, *ApJ*, 313, 42
 Duboscq T., Hogg N. B., Fleury P., Larena J., 2024, *J. Cosmol. Astropart. Phys.*, 2024, 021
 Eisenstein D. J. et al., 2023, preprint ([arXiv:2306.02465](https://arxiv.org/abs/2306.02465))
 Faure C. et al., 2008, *ApJS*, 176, 19
 Ferrami G., Wyithe J. S. B., 2024, *MNRAS*, 532, 1832
 Fleury P., Larena J., Uzan J.-P., 2021, *J. Cosmol. Astropart. Phys.*, 2021, 024
 Franco M. et al., 2024, *ApJ*, 973, 23
 Garvin E. O., Kruk S., Cornen C., Bhatawdekar R., Cañameras R., Merín B., 2022, *A&A*, 667, A141
 Gavazzi R., Marshall P. J., Treu T., Sonnenfeld A., 2014, *ApJ*, 785, 144
 Gelli V., Salvadori S., Ferrara A., Pallottini A., Carniani S., 2023, *ApJ*, 954, L11
 Hartigan J. A., Hartigan P. M., 1985, *Ann. Stat.*, 13, 70
 He Q. et al., 2024, *MNRAS*, 532, 2441
 Heymans C. et al., 2021, *A&A*, 646, A140
 Hezaveh Y. D. et al., 2013, *ApJ*, 767, 132
 Hogg N. B., 2024, *MNRAS*, 528, L95
 Hogg N. B., Fleury P., Larena J., Martinelli M., 2023, *MNRAS*, 520, 5982
 Hogg N. B., Shajib A. J., Johnson D., Larena J., 2025, preprint ([arXiv:2501.16292](https://arxiv.org/abs/2501.16292))
 Holloway P. et al., 2025, preprint ([arXiv:2503.15328](https://arxiv.org/abs/2503.15328))
 Holloway P., Verma A., Marshall P. J., More A., Tecza M., 2023, *MNRAS*, 525, 2341
 Huterer D., Keeton C. R., Ma C.-P., 2005, *ApJ*, 624, 34
 Hyde J. B., Bernardi M., 2009a, *MNRAS*, 394, 1978
 Hyde J. B., Bernardi M., 2009b, *MNRAS*, 396, 1171
 Jackson N., 2008, *MNRAS*, 389, 1311
 Jin S. et al., 2018, *ApJ*, 864, 56
 Jin S. et al., 2024, *A&A*, 690, L16
 Kochanek C. S., 2004, in Meylan G., Jetzer P., North P., Schneider P., Kochanek C., Wambsganss J., eds, 33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro. preprint ([arXiv:astro-ph/0407232](https://arxiv.org/abs/astro-ph/0407232))
 Koekemoer A. M. et al., 2007, *ApJS*, 172, 196
 Kormann R., Schneider P., Bartelmann M., 1994, *A&A*, 284, 285
 Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Poczós B., 2018, *MNRAS*, 473, 3895
 Lee D.-W., Kim S.-J., 2014, *MNRAS*, 443, 328
 Li T. et al., 2025, preprint ([arXiv:2503.15327](https://arxiv.org/abs/2503.15327))
 Li T., Collett T. E., Krawczyk C. M., Enzi W., 2024, *MNRAS*, 527, 5311

- Lin X. et al., 2023, *ApJ*, 944, L59
- Lines N. E. P. et al., 2025, preprint (arXiv:2503.15326)
- Mahler G. et al., 2025, *MNRAS*, 544, L8
- Marshall P. J. et al., 2016, *MNRAS*, 455, 1171
- Marshall P. J., Hogg D. W., Moustakas L. A., Fassnacht C. D., Bradač M., Schrabback T., Blandford R. D., 2009, *ApJ*, 694, 924
- Mason C. A. et al., 2015, *ApJ*, 805, 79
- Mellier Y. et al., 2025, *A&A*, 697, A1
- Mercier W. et al., 2024, *A&A*, 687, A61
- More A. et al., 2016, *MNRAS*, 455, 1191
- More A. et al., 2024, *MNRAS*, 533, 525
- More A., Cabanac R., More S., Alard C., Limousin M., Kneib J. P., Gavazzi R., Motta V., 2012, *ApJ*, 749, 38
- Nagam B. C. et al., 2025, preprint (arXiv:2502.09802)
- Newman S. L. et al., 2025, preprint (arXiv:2501.03133)
- Nightingale J. et al., 2025, *MNRAS*, 543, 203
- Nightingale J. W. et al., 2021, *J. Open Source Softw.*, 6, 2825
- O’Riordan C. M., 2025, *A&A*, 694, A145
- O’Riordan C. M., Despali G., Vegetti S., Lovell M. R., Moliné A., 2023, *MNRAS*, 521, 2342
- Pearce-Casey R. et al., 2025, *A&A*, 696, A214
- Pearson J. et al., 2023, *MNRAS*, 527, 12044
- Pourrahmani M., Nayyeri H., Cooray A., 2018, *ApJ*, 856, 68
- Razali N. M., Wah Y. B., 2011, *J. Stat. Model. Anal.*, 2, 21
- Refsdal S., 1964, *MNRAS*, 128, 307
- Rodrigo C., Solano E., Bayo A., 2012, IVOA Working Draft
- Rojas K. et al., 2023, *MNRAS*, 523, 4413
- Rojas K. et al., 2025, preprint (arXiv:2503.15325)
- Sainz de Murieta A., Collett T. E., Magee M. R., Pierel J. D. R., Enzi W. J. R., Lokken M., Gagliano A., Rychanowski D., 2024, *MNRAS*, 535, 2523
- Scoville N. et al., 2007, *ApJS*, 172, 1
- Shajib A. J., Treu T., Birrer S., Sonnenfeld A., 2021, *MNRAS*, 503, 2380
- Shuntov M. et al., 2025, *A&A*, 696, L14
- Silverman B. W., 1981, *J. R. Stat. Soc.*, 43, 97
- Sonnenfeld A. et al., 2020, *A&A*, 642, A148
- Sonnenfeld A., 2022, *A&A*, 659, A132
- Sonnenfeld A., 2024, *A&A*, 690, A325
- Sonnenfeld A., Cautun M., 2021, *A&A*, 651, A18
- Sonnenfeld A., Li S.-S., Despali G., Gavazzi R., Shajib A. J., Taylor E. N., 2023, *A&A*, 678, A4
- Sprent P., 1998, *Data Driven Statistical Methods*. Chapman and Hall, London
- Suyu S. H. et al., 2014, *ApJ*, 788, L35
- Swinbank A. M. et al., 2015, *ApJ*, 806, L17
- Tan C. Y. et al., 2024, *MNRAS*, 530, 1474
- Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V. E., 2012, *Nature*, 481, 341
- Walmsley M. et al., 2025, preprint (arXiv:2503.15324)
- Weiner C., Serjeant S., Sedgwick C., 2020, *Res. Notes Am. Astron. Soc.*, 4, 190
- Williams C. C. et al., 2018, *ApJS*, 236, 33
- Williams R. E. et al., 1996, *AJ*, 112, 1335
- Wong K. C. et al., 2020, *MNRAS*, 498, 1420
- York D. G. et al., 2000, *AJ*, 120, 1579

APPENDIX A: SUMMARY OF STATISTICAL TESTS

In this appendix we tabulate the results of the statistical tests used to compare the forecast and observed data. For each test, the null hypothesis is that the forecast and observed data were drawn from the same underlying distribution. Table A1 shows the test results for the comparison between the forecast and the 17spectacular lenses; Table A2 shows the test results for the comparison between the forecast and the 59 high confidence lenses; Table A3 shows the test results for the comparison between the forecast and the 116 high and medium confidence lenses. For the KS test, the null hypothesis is

Table A1. Results of the statistical tests comparing the forecast with the 17 spectacular lenses. The ‘H₀’ column indicates when the null hypothesis is sustained with a tick (graphic) and when it is rejected with a cross (graphic).

Quantity	Statistic	KS		H ₀	AD		H ₀
		<i>p</i> -value			<i>p</i> -value		
θ_E	0.398	6.1×10^{-3}	graphic		8.790	1.0×10^{-3}	graphic
z_{lens}	0.218	3.8×10^{-1}	graphic		0.190	2.5×10^{-1}	graphic
M_{lens}^{F115W}	0.361	2.2×10^{-2}	graphic		2.864	2.2×10^{-2}	graphic
M_{lens}^{F150W}	0.418	3.3×10^{-3}	graphic		4.229	6.6×10^{-3}	graphic
M_{lens}^{F277W}	0.658	9.5×10^{-8}	graphic		33.390	1.0×10^{-3}	graphic
M_{lens}^{F444W}	0.807	1.7×10^{-12}	graphic		48.953	1.0×10^{-3}	graphic
$M_{\text{source}}^{F115W}$	0.328	5.0×10^{-2}	graphic		4.704	4.4×10^{-3}	graphic
$M_{\text{source}}^{F150W}$	0.218	3.5×10^{-1}	graphic		1.595	7.1×10^{-2}	graphic
$M_{\text{source}}^{F277W}$	0.304	6.8×10^{-2}	graphic		2.915	2.1×10^{-2}	graphic
$M_{\text{source}}^{F444W}$	0.195	4.8×10^{-1}	graphic		1.023	1.2×10^{-1}	graphic

Table A2. Results of the statistical tests comparing the forecast with the 59 high confidence lenses. The ‘H₀’ column indicates when the null hypothesis is sustained with a tick (graphic) and when it is rejected with a cross (graphic).

Quantity	Statistic	KS		H ₀	AD		H ₀
		<i>p</i> -value			<i>p</i> -value		
θ_E	0.101	5.6×10^{-1}	graphic		0.750	1.6×10^{-1}	graphic
z_{lens}	0.224	5.3×10^{-3}	graphic		3.479	1.3×10^{-2}	graphic
M_{lens}^{F115W}	0.238	2.8×10^{-3}	graphic		5.511	2.3×10^{-3}	graphic
M_{lens}^{F150W}	0.268	4.3×10^{-4}	graphic		5.370	2.6×10^{-3}	graphic
M_{lens}^{F277W}	0.517	6.3×10^{-15}	graphic		69.439	1.0×10^{-3}	graphic
M_{lens}^{F444W}	0.602	1.2×10^{-20}	graphic		96.381	1.0×10^{-3}	graphic
$M_{\text{source}}^{F115W}$	0.255	1.3×10^{-3}	graphic		10.647	1.0×10^{-3}	graphic
$M_{\text{source}}^{F150W}$	0.197	2.1×10^{-2}	graphic		5.966	1.6×10^{-3}	graphic
$M_{\text{source}}^{F277W}$	0.309	2.0×10^{-5}	graphic		21.328	1.0×10^{-3}	graphic
$M_{\text{source}}^{F444W}$	0.293	6.7×10^{-5}	graphic		22.360	1.0×10^{-3}	graphic

Table A3. Results of the statistical tests comparing the forecast with the 116 high and medium confidence lenses. The ‘H₀’ column indicates when the null hypothesis is sustained with a tick (graphic) and when it is rejected with a cross (graphic).

Quantity	Statistic	KS		H ₀	AD		H ₀
		<i>p</i> -value			<i>p</i> -value		
θ_E	0.061	7.8×10^{-1}	graphic		0.036	2.5×10^{-1}	graphic
z_{lens}	0.171	2.3×10^{-3}	graphic		4.714	4.4×10^{-3}	graphic
M_{lens}^{F115W}	0.160	7.3×10^{-3}	graphic		3.138	1.7×10^{-2}	graphic
M_{lens}^{F150W}	0.103	1.8×10^{-1}	graphic		0.832	1.5×10^{-1}	graphic
M_{lens}^{F277W}	0.366	7.1×10^{-14}	graphic		69.647	1.0×10^{-3}	graphic
M_{lens}^{F444W}	0.436	1.0×10^{-19}	graphic		101.870	1.0×10^{-3}	graphic
$M_{\text{source}}^{F115W}$	0.302	9.8×10^{-9}	graphic		39.868	1.0×10^{-3}	graphic
$M_{\text{source}}^{F150W}$	0.269	2.0×10^{-7}	graphic		26.340	1.0×10^{-3}	graphic
$M_{\text{source}}^{F277W}$	0.419	3.6×10^{-18}	graphic		91.412	1.0×10^{-3}	graphic
$M_{\text{source}}^{F444W}$	0.410	2.2×10^{-17}	graphic		92.232	1.0×10^{-3}	graphic

rejected at 95 per cent confidence if the p -value is less than 0.05; for the AD test, the null hypothesis is rejected if the test statistic is greater than the critical value, which for 95 per cent confidence is 1.961. We note that the `scipy.stats.anderson_ksamp` method we use to compute the AD test in Python returns p -values of 1.0×10^{-3} as a proxy for very small p -values that cannot be computed exactly.

¹Laboratoire Univers et Particules de Montpellier, CNRS & Université de Montpellier, Parvis Alexander Grothendieck, F-34090 Montpellier, France

²School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, Newcastle-upon-Tyne NE1 7RU, UK

³Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK

⁴Department of Physics, Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA

⁵STAR Institute, Quartier Agora – Allée du six Août, 19c B-4000 Liège, Belgium

⁶Department of Physics, Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK

⁷Department of Computer Science, Aalto University, PO Box 15400, FI-00076 Espoo, Finland

⁸Faculty of Science, Department of Physics, University of Helsinki, FI-00014 Helsinki, Finland

⁹Jet Propulsion Laboratory, California Institute of Technology, 4800, Oak Grove Drive, Pasadena, CA 91109, USA

¹⁰Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

¹¹Purple Mountain Observatory, Chinese Academy of Sciences, 10 Yuanhua Road, Nanjing 210023, China

¹²Cosmic Dawn Centre (DAWN), DTU Space, Elektrovej 327, 2800 Kgs. Lyngby, Denmark

¹³Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200 Copenhagen, Denmark

¹⁴Institut d'Astrophysique de Paris, UMR 7095, CNRS, Sorbonne Université, 98 bis boulevard Arago, F-75014 Paris, France

¹⁵Laboratoire d'astrophysique de Marseille, Aix Marseille University, CNRS, CNES, Marseille, F-13388, France

¹⁶Department of Physics, University of California Santa Barbara, Santa Barbara, CA 93106, USA

¹⁷Laboratory for Multiwavelength Astrophysics, School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA

¹⁸Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

¹⁹Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

²⁰Department of Physics and Astronomy, University of Hawaii, Hilo, 200 W Kawili St, Hilo, HI 96720, USA

²¹Caltech/IPAC, 1200 E. California Blvd., Pasadena, CA 91125, USA

²²Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA

²³Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciamia Building, Burnaby Road, Portsmouth PO1 3FX, UK

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