Photometric redshift estimation of strongly lensed galaxies

Langeroodi, D.; Sonnenfeld, A.; Hoekstra, H.; Agnello, A.

Note: To cite this publication please use the final published version (if applicable).

Astronomy & Astrophysics

Photometric redshift estimation of strongly lensed galaxies

Danial Langeroodi[1,2] , Alessandro Sonnenfeld[1] , Henk Hoekstra[1] , and Adriano Agnello[2]

[1] Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[2] DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen Ø, Denmark
    e-mail: danial.langeroodi@nbi.ku.dk

ABSTRACT

Context. Around 10^5 strongly lensed galaxies are expected to be discovered with upcoming wide-field surveys such as *Euclid* and the LSST. Utilising these large samples to study the inner structure of lens galaxies requires source redshifts, which are needed to turn lens models into mass measurements. However, obtaining spectroscopic source redshifts for large samples of strong lenses is prohibitive with the current capacity of spectroscopic facilities.

Aims. As an alternative to spectroscopy, we study the possibility of obtaining source photometric redshifts (photo-zs) for large samples of strong lenses. We pay particular attention to the problem of blending between the lens and the source light.

Methods. Our strategy consists of deblending the source and lens light by simultaneously modelling the lens galaxy and the background source in all available photometric bands, and then feeding the derived source colours to a template-fitting photo-z algorithm. We describe the lens and the source light with a Sérsic profile, and the lens mass with a singular isothermal ellipsoid. We first test our approach on a simulated sample of lenses. Then, we apply it to 23 real systems with broad-band photometry from the Hyper Suprime-Cam survey.

Results. We identify the deviations of the lens light from a Sérsic profile and the contrast between the lens and source image as the main drivers of the source colour measurement error. Although the former is challenging to measure directly for real lenses, we find the latter to be sufficient for evaluating the accuracy of a measured source colour. We split the real sample based on the ratio Λ of the lens to source surface brightness measured at the image locations. In the $\Lambda < 1$ regime, the photo-z outlier fraction is 20%, and the accuracy of photo-z estimation is limited by the performance of the template-fitting process. In the opposite regime, the photo-z outlier fraction is 75%, and the errors from the source colour measurements dominate the photo-z uncertainty.

Conclusions. Measuring source photo-zs for lenses with $\Lambda < 1$ poses no particular challenges compared to the isolated galaxy case. For systems with significant lens light contamination, however, improvements in the description of the surface brightness distribution of the lens are required: a single Sérsic model is not sufficiently accurate.

Key words. gravitational lensing: strong

1. Introduction

Strong gravitational lensing is a rare phenomenon in which the close alignment of a background light source with a massive foreground lens results in multiple images of the former. The positions and surface brightness distribution of the multiple images tightly constrain the total projected mass (dark matter and baryonic component) of the lens. In virtue of this feature, strong lensing has been used in the last two decades to investigate important aspects of the inner structure of massive galaxies, such as their total density profile (Treu & Koopmans 2004; Auger et al. 2010; Barnabè et al. 2011; Bolton et al. 2012; Sonnenfeld et al. 2013), their stellar mass-to-light ratio (Treu et al. 2010; Sonnenfeld et al. 2015, 2019a; Smith et al. 2015), their dark matter density profile (Sonnenfeld et al. 2012; Schuldt et al. 2019; Newman et al. 2015; Oldham & Auger 2018; Shajib et al. 2021), and the presence of substructure (Vegetti et al. 2010; Hezaveh et al. 2016; Gilman et al. 2020).

One of the crucial pieces of information on which most strong lensing studies rely is the knowledge of the redshift of the lens and source galaxies; both are needed in order to convert angular measurements of image positions into measurements of mass. Traditionally, these redshifts have been obtained by means of dedicated spectroscopic observations. However, while lens galaxies are usually very bright, obtaining spectroscopic redshifts of strongly lensed sources can be relatively expensive, typically requiring approximately one hour of integration time on an eight-metre class telescope (Ruff et al. 2011). Moreover, a large fraction of lensed galaxies are found in the redshift range $1.5 < z < 3$, where they do not contribute with any strong emission line features to the optical part of the spectrum, thus requiring near-infrared observations (Sonnenfeld et al. 2013).

The number of known strong lenses is growing rapidly, thanks to wide-field photometric surveys such as the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP, Aihara et al. 2018a), the Kilo-Degree Survey (KiDS, Kuijken et al. 2015), and the Dark Energy Survey (DES, Dark Energy Survey Collaboration 2016), which are leading to the discovery of hundreds of lenses (Sonnenfeld et al. 2018, 2020; Petrillo et al. 2019; Jacobs et al. 2019; Cañameras et al. 2021; Rojas et al. 2022; Storfer et al. 2022). However, the number of lenses with available spectroscopic data is only a fraction of the total. The largest sample of strong lenses with a spectroscopic measurement of the redshift of both the lens and the source galaxy is the Sloan ACS Lens Survey (SLACS, Bolton et al. 2006), which consists of ∼100 systems (Auger et al. 2009). Although there are ongoing spectroscopic follow-up programmes (Tran et al. 2022), this lack of spectroscopic observations is already limiting our ability to take

advantage of the large number of currently known lenses, and will become an increasingly severe issue in the next decade; upcoming surveys such as *Euclid*[1], the LSST[2], and the Chinese Space Station Telescope (CSST) are predicted to lead to the discovery of 10^5 new strong lenses (Collett 2015). While our capabilities for spectroscopic observations will also increase with upcoming facilities such as the 4-m Multi-Object Spectrograph Telescope (4MOST; de Jong et al. 2019), the Maunakea Spectroscopic Explorer (The MSE Science Team 2019), and MegaMapper (Schlegel et al. 2019), most of these strong lenses will realistically not be followed up.

Photometric redshift (photo-z) measurements are a valid alternative to spectroscopic observations, when highly precise redshift estimates of individual galaxies are not strictly needed. This is the case for most strong lensing applications. The lensing deflection angle is a weak function of source redshift; therefore, redshift uncertainties do not propagate catastrophically to the mass measurements[3]. Photometric redshift techniques have developed significantly in the last few years (see Salvato et al. 2019, for a review), especially because accurate photo-zs are a crucial ingredient for many experiments in cosmology (see e.g., Tanaka et al. 2018; Hildebrandt et al. 2021; Myles et al. 2021). Carrying out photometric redshift measurements on strong lenses, however, is generally more difficult than doing so on isolated objects. The typical angular separation of a strongly lensed image from the centre of its lens galaxy is ~1′; as a result the lens and source light are often severely blended, especially when observed with ground-based imaging. While this blending is usually a minor issue for the determination of the lens light distribution, it can have a strong impact on the ability to recover the magnitude and colours of the background source, which are needed to determine its photo-z.

In this work we investigate how accurately we can measure the photo-zs of strongly lensed galaxies, when observed with photometric data similar to the HSC survey. We chose the HSC survey because it is to date both the deepest and the one with the best image quality among the ground-based surveys used to find strong lenses. It also resembles the expected depth and image quality from LSST. Our procedure for measuring the source photo-z consists of a two-step process: first we fit the multi-band image of a strong lens with a simply parametrised model describing the lens light, the lens mass, and the source light; then we apply a template fitting-based photo-z method to the measured magnitude and colours. We first test this method on simulated lenses, then apply it to a set of 23 strong lenses from the Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI; Sonnenfeld et al. 2019a). We pay particular attention to our ability in measuring accurate colours. For this purpose we investigate the importance of systematic errors associated with a non-perfect description of the surface brightness profile of the foreground galaxy.

Section 2 covers the details of our sample of strongly lensed galaxies, consisting of both simulated and real systems. In Sect. 3 we explain our methodology for modelling the source and lens surface brightness and lens mass in multi-band images of galaxy-galaxy lens systems. In Sect. 4 we identify the main drivers of the source colour measurement error by applying our method to the simulated sample. We estimate the source redshifts

for the simulated and the real samples in Sect. 5. We discuss and summarise our results in Sects. 6 and 7.

2. Data

Our data consist of a real and a simulated sample of galaxy-galaxy strong lens systems, both based on imaging data in the broad-band *grizy* photometry of the HSC survey. In this section we give a brief description of the HSC photometric data and describe our real and simulated samples.

2.1. HSC photometry

For our experiment we use *grizy* imaging data from the second public data release of the HSC survey (HSC PDR2; Aihara et al. 2019). For each object we obtain (from the HSC PDR2) a 101×101 pixel (0.164″ per pixel) sky-subtracted cutout (centred on the lens galaxy), a pixelised model of the point spread function (PSF), and the variance map, in each band. The median i-band seeing is 0.7″ and its magnitude limit is 26.2 (5σ detection limit for a point source).

The photometry of HSC is calibrated against the Pan-STARRS 1 (PS1) 3π survey (Magnier et al. 2013). At bright magnitudes, HSC coadd-measured star PSF fluxes have ~0.03 mag RMS residuals with respect to the PS1-based measurements and ~0.04 mag RMS residuals with respect to the SDSS-based (SDSS DR9: the ninth data release of the Sloan Digital Sky Survey; Ahn et al. 2012) measurements (see Fig. 8 and Table 6 from Aihara et al. 2018b). While the latter consists of both the absolute (the linear offset between the measured and true magnitudes) and relative (the standard deviation of the offset-corrected measured magnitudes around the true magnitudes) zero-point calibration inaccuracies, the former mostly consists of the relative calibration inaccuracies (since the zero-points of HSC are calibrated against the PS1). Internal PS1 3π comparisons and comparisons with the SDSS indicate that the photometry of both PS1 and SDSS surveys is accurate to ~1% in all bands (Schlafly et al. 2012; Finkbeiner et al. 2016). However, because of spatial variations of the HSC filter transmissions, the RMS residuals of the HSC coadd-measured star PSF fluxes exceed the expected ~1% scatter inherited from the inaccuracies in PS1 and SDSS photometry (see Kawanomoto et al. 2018).

Photo-z estimation in this work (see Sect. 3.2 for a detailed description of the method) is mostly sensitive to the colour measurement accuracy (as opposed to magnitude measurement accuracy), and therefore is only affected by the relative calibration accuracy. In general, a calibration of the colour–redshift relation compensates for this inaccuracy (see e.g., Ilbert et al. 2010). However, strong lenses are sufficiently rare that colours are measured from different pointings/observations, and thus are nonetheless affected by the relative calibrations. Therefore, the relative calibration accuracy (~0.03 mag for HSC) provides the lower threshold for the uncertainty on colour measurement of strongly lensed galaxies in HSC photometry. The stage IV surveys (e.g., ESA-*Euclid* and Rubin-LSST) are expected to perform better than the HSC (we discuss this in more detail in Sect. 6.1).

2.2. Real sample

The real sample contains 23 galaxy–galaxy strong lenses from the study of Sonnenfeld et al. (2019a). All of these lenses

---

[3] Strictly speaking, this does not apply to lenses in which redshift of the lens is close to that of the source. However, such systems are rare because the lensing cross section decreases with decreasing source redshift.

consist of a foreground galaxy that belongs to the constant mass (CMASS) sample of the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013). They all have publicly available HSC *grizy* imaging data, and spectroscopic measurements of both the source and the lens redshift. The lens redshifts range from 0.46 to 0.80, with a 0.58 median. The Einstein radii, measured by Sonnenfeld et al. (2019a), range from 0.72″ to 2.14″. The source redshifts cover a range between 0.94 and 3.12, with a median of 1.82 (Sonnenfeld et al. 2018, 2019a; Wong et al. 2018). The average source-plane (i.e., de-lensed) half-light radii (inferred from the best-fit models of Sonnenfeld et al. 2019a) range from 0.05″ to 0.95″, with a 0.20″ median. The source *i*-band magnitudes are between 22.86 and 27.02.

2.3. Simulated sample

We simulated each lens system in our mock sample by taking the five-band *grizy* HSC images of a real massive galaxy as the lens and adding the simulated images of a lensed source, closely following the method that Sonnenfeld et al. (2020) used to produce the training sample for the citizen science lens finding programme Space Warps-HSC[4]. We selected galaxies from the CMASS sample for this purpose since this is the sample from which the foreground galaxies of our real lenses are drawn (see Sect. 2.2). In particular, we drew elliptical galaxies from the subset of visually selected CMASS galaxies with stellar mass above $10^{11} M_\odot$ of Sonnenfeld et al. (2019b).

Modelling the surface brightness distribution of the simulated lensed source in the lens plane requires associating a lensing potential with the foreground galaxy. We assign a singular isothermal ellipsoid (SIE, Kormann et al. 1994) mass profile to each lens, constructed from aligned co-centric elliptical isodensity contours of equal axis ratio (q_{SIE}). Bolton et al. (2008) showed that the SIE mass model can provide accurate reproductions of the source light distribution in the lens plane. The SIE is a five-parameter model fully described by the position of its centroid pixel ($x_{centroid}$ and $y_{centroid}$), axis ratio (q_{SIE}), position angle (p_A), and angular Einstein radius (b_{SIE}; defined as the circularised radius of the isodensity contour that encloses an average surface mass density equal to the critical surface mass density for lensing).

We drew the angular Einstein radii from a truncated normal distribution between 0.5″ and 3.0″, centred on 1.5″, and with a 0.5″ width. We set the lower limit because systems with angular Einstein radii significantly smaller than the seeing of HSC are difficult to detect as strong lenses and are usually missed by photometry-based lens searches. The upper limit is imposed because in this work we are mostly interested in galaxy-scale lenses as opposed to group-scale lenses, which dominate the image separations larger than 3″ (Oguri 2006).

We do not intend to align the lens mass profile to its surface brightness distribution because our aim is to span a broad range of parameter space with our simulation. Therefore, we randomly draw the axis ratios between 0.4 and 1.0 and position angles between 0° and 180°, regardless of the lens light. As we explain in Sect. 4, if we can reproduce the strong lensing image configuration in this general setting (where the lens mass and light profiles are not aligned), we should also be able to reproduce them in the more realistic case where the lens mass and light profiles are closely aligned. Furthermore, we select the polar coordinates (r, ϕ) of the SIE centroids randomly in one-pixel radius circles centred on the centre of the lens cutout in each photometry band
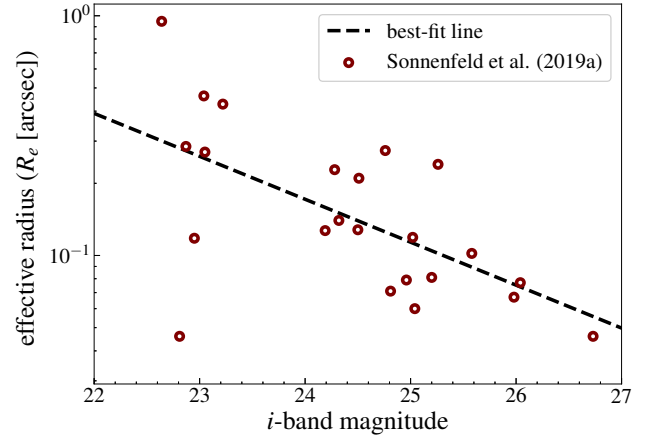
[4] https://www.spacewarps.org

Fig. 1. Effective radius (R_e [arcsec]) plotted against the *i*-band magnitude (m_i) for the sources in our real sample of strongly lensed galaxies. The dashed line shows the best-fit line, given by Eq. (3). This magnitude–size relation (with scatter) is used to assign the effective radii corresponding to the drawn *i*-band magnitudes of the source galaxies in our simulated sample of strongly lensed galaxies.

(i.e., $x = 50$, $y = 50$ pixels). In other words, we randomly draw $r^2_{centroid}$ between 0 and $1''^2$ and the $\phi_{centroid}$ between 0 and 2π.

We describe the surface brightness of source galaxies in the source plane with elliptical Sérsic surface brightness profiles, constructed from co-centric elliptical isophotes (Sersic 1968). The Sérsic light profile is defined as

$$I(x, y) = I_0 \exp\left\{ - b(n) \left(\frac{R}{R_e}\right)^{1/n} \right\}, \tag{1}$$

where I_0 is the amplitude, n is the Sérsic index describing the central concentration of the light profile, and

$$b(n) \approx 2n - \frac{1}{3} + \frac{4}{405n} + O(n^{-2}) \tag{2}$$

is defined such that the effective radius (R_e) encloses half the light (Ciotti & Bertin 1999).

We assume that the colours of the background source are spatially uniform. Therefore, for each simulated source, we use the same values of structural parameters (axis ratio, position angle, Sérsic index, and half-light radius) in all photometric bands. We randomly draw the axis ratios of the co-centric elliptical isophotes between 0.3 and 1.0, and the position angles between 0° and 180°. We draw the Sérsic indices randomly between 0.5 and 2 because the detected lensed sources are typically late-type galaxies, best fit by a Sérsic index between 0.5 and 2 (see e.g., Kormendy et al. 2009; Kelvin et al. 2012). The prevalence of late-type galaxies among the lensed source population is partly intrinsic and partly a result of selection effects: blue late-type galaxies can be more easily detected in photometry-based search algorithms as they can be more easily separated from the light of the lens galaxy, which is typically red (see e.g., Gavazzi et al. 2014; Sonnenfeld et al. 2018).

We assigned effective radii assuming a magnitude–size relation, with scatter, which we inferred from the Sonnenfeld et al. (2019a) sample (Fig. 1). In particular, we fitted the unlensed *i*-band magnitude m_i and source effective radius R_e with the model

$$\log_{10}(R_e) \sim -0.179m_i + 3.533 + \mathcal{D}(0, 0.272), \tag{3}$$

where $\mathcal{D}(0, 0.272)$ is a normal distribution centred at 0 with a dispersion equal to the standard deviation of the best-fit line (0.272). Motivated by the distribution of source magnitudes in the sample of real strongly lensed galaxies (Sect. 2.2 and Table 4 in Sonnenfeld et al. 2019a), we draw the source i-band magnitudes from a normal distribution centred on 24.4 mag, with a 1.2 mag dispersion. We assign effective radii corresponding to the drawn i-band magnitudes using Eq. (3).

The first requirement for producing strongly lensed images is to place the source within the region that is mapped into multiple images, which for an SIE mass model corresponds to the radial caustic and has a circularised radius similar to the value of the Einstein radius. However, placing the source too close to the radial caustic produces three images, two of which are very faint and close to the deflector, and one that has minimal distortion. In practice, these image configurations cannot be identified as strongly lensed. To prevent the production of such image configurations, following Sonnenfeld et al. (2020), we assign r_{source} by drawing x from the $p(x) \propto x \exp(-4x)$ distribution, where $x = r_{source}/b_{SIE}$. Multiplying the drawn x by the previously drawn Einstein radius (b_{SIE}) gives the r_{source}. We draw the ϕ_{source} randomly between 0 and 2π. For each drawn source position ($r_{source}, \phi_{source}$), we calculate the number and magnifications of the strongly lensed images. If the drawn source position results in multiple images, we accept it. Otherwise, we draw a new source position and repeat the same procedure.

Unlike its structural parameters and position, the amplitude of the source Sérsic profile varies with photometry band. This reflects the spectral energy distribution (SED) of galaxy, which depends on both its redshift and spectral type. We draw the source redshifts uniformly between 1.0 and 3.0, while selecting the spectral types randomly from the Bayesian photometric redshift (BPZ; Benítez 2000) SED templates (Coleman et al. 1980; Kinney et al. 1996; Sawicki et al. 1997; Fernández-Soto et al. 1999; Coe et al. 2006). Since the majority of source galaxies in detected strongly lensed samples are late-type galaxies, we only draw late-type spectral types from the BPZ SED templates. Each drawn redshift–spectral type pair corresponds to an SED continuum, which we convert to multi-band $grizy$ magnitudes, using the $grizy$ response functions. We re-scale the calculated $grizy$ magnitudes, such that the i-band magnitude matches its previously drawn value (m_i).

Provided a source position, source light profile, and a lens mass profile, we calculate the surface brightness distribution of the strongly lensed image in each photometric band. For each band, we convolve the resulting image with the PSF of the real lens galaxy, and add the PSF-convolved lensed image to the lens image.

We sift through our simulated sample to discard the cases that would not be detected as strongly lensed galaxies with conventional detection techniques. This is the case when the lensed source is not sufficiently bright compared to the background sky noise, σ_{sky}. We define the source footprint as the pixels in the source-only image with pixel values larger than $3\sigma_{sky}$. Since the sources in Sonnenfeld et al. (2019a) are typically brightest in g-band, this band often has the largest number of source region pixels. We assume that detection in one band (g here) is sufficient to identify the system as strongly lensed. Therefore, we discard any simulated source for which the number of g-band source region pixels is smaller than 20 as undetectable. We make a sample of 73 galaxy–galaxy strong lens systems in the $grizy$ HSC broad-band photometry. Figure 2 shows the HSC gri colour-composite images of these systems.

3. Methods

In this section we describe our method for measuring the source photometry and using the measured photometry to estimate the source photometric redshift (photo-z). Source photometry is measured by deblending the lens and source light, which is done by modelling the lensed systems (Sect. 3.1). We use a photometric redshift estimation algorithm (photo-z algorithm) to convert the measured colours to redshift probability distribution functions (PDFs). We describe the photo-z algorithm used in this work in Sect. 3.2.

3.1. Modelling the lensed system

Our lens model was designed to reproduce the image configuration of the strongly lensed system by simultaneously modelling the lensing potential, the surface brightness distribution of the foreground lens galaxy, and the unlensed surface brightness distribution of the source galaxy. We describe them with a SIE mass profile, an elliptical Sérsic profile in the image plane, and an elliptical Sérsic profile in the source plane, respectively.

For both the lens and the source galaxy, we assume that the structural parameters of the light profile (axis ratio, Sérsic index, half-light radius, position angle, and centroid position) are the same in each band. Additionally, we allow the four colours to vary independently of each other. For the five-band HSC $grizy$ photometry used in this work, the colours are $g-i$, $r-i$, $i-z$, and $i-y$. For each set of source and lens light profile parameters (2×9 parameters), and lens mass profile parameters (five parameters), we find the source and lens i-band magnitudes that minimise

$$\chi^2 = \sum_\lambda \sum_j \left(\frac{I_{\lambda,j}^{mod, L} + I_{\lambda,j}^{mod, S} - I_{\lambda,j}^{obs}}{\sigma_{\lambda,j}^2} \right), \tag{4}$$

in which $I_{\lambda,j}^{mod, L}$ ($I_{\lambda,j}^{mod, S}$) is the λ-band lens (source) PSF-convolved Sérsic model, evaluated at the jth pixel, $I_{\lambda,j}^{obs}$ is the observed value of the jth pixel, and $\sigma_{\lambda,j}$ is the observational uncertainty provided by the variance file.

We use the EMCEE package (Foreman-Mackey et al. 2013) to explore the parameter space with a Markov chain Monte Carlo (MCMC) sampler, searching for the parameter values that minimise the χ^2 in Eq. (4). Since we are only interested in obtaining a good description of the system in the area that includes the multiple images of the source, we only fitted data within a circular region with a 20 pixel radius (3.36″) around the lens centroid in each photometric band.

In addition to the lens and the source, image cutouts typically contain contaminating bright objects. These are either satellite galaxies and extended features in the lens plane or accidental alignments of other bright objects that do not belong to the lens or source planes. These contaminants can bias the measurements if not taken into account. If the contaminant is sufficiently faint and not closely aligned with the lensed system, its surface brightness at the lens or source image position is negligible. Hence, it is sufficient to discard the pixels at which the contaminant dominates the image cutout. To do this we identify the contaminants by running the Source Extractor (SExtractor; Bertin & Arnouts 1996) on the i-band cutouts. We exclude (i.e., mask) these segments from the chi-squared optimisation by excluding their pixels from the sum in Eq. (4).

However, if the contaminant is too bright or closely aligned with the lensed system, its surface brightness at the lens or source

Fig. 2. 101 × 101 pixel (0.164″ per pixel) *gri* colour-composite images of the 73 systems in our simulated sample of strongly lensed galaxies. Source *i*-band magnitudes increase from 20.50 to 25.90 (from left to right and top to bottom). For each system, the *i*-band ratio Λ_i of the lens to source surface brightness measured at the image locations (introduced and discussed in Sect. 4.2) is overwritten on its image.

image position becomes significant. This additional contaminating light results in systematic errors in modelling the lensed system. Moreover, often `SExtractor` cannot separate these contaminants from the lensing system. We flagged these cases by running `SExtractor` on the original lens image (for the simulated sample) and by visual inspection (for both the simulated and the real samples). In these cases we modelled the contaminant light explicitly, with an additional Sérsic component. Figure 3 shows an example of a fitted model to the *grizy* images of a simulated strong lens system.

3.2. Source photometric redshift estimation

We use BPZ (Benítez 2000) to obtain photo-*z* estimates from the measured magnitudes. BPZ fits the PSF-corrected multi-band magnitude measurements of the galaxy surface brightness with a grid of SED templates varying in both redshift *z* and spectral type *T*. The outcome is a probability distribution $P(z, T)$. The marginal posterior probability in redshift, $P(z)$, is calculated by marginalising over the spectral type *T*. BPZ contains

elliptical, spiral, irregular, and starburst spectral types, including the templates from Benítez (2000; see also Coleman et al. 1980; Kinney et al. 1996; Sawicki et al. 1997; Fernández-Soto et al. 1999), re-calibrated in Benítez et al. (2004) with the addition of two bluer templates of Coe et al. (2006), all corrected for intergalactic absorption according to Madau (1995). These are the same spectral types that we used for the simulations (Sect. 2.3).

The BPZ adopts a magnitude-dependent prior that gives the redshift and spectral type probability distribution for a reference-band magnitude. Benítez (2000) calibrated this prior empirically on the spectroscopic redshifts and spectral type classifications of the Hubble Deep Field North observations (HDF-N; Williams et al. 1996) and Canada-France Redshift Survey (CFRS; Lilly et al. 1995). For simulated systems, instead of this prior we adopt a flat prior that is truncated below the lens redshift and above a maximum value of $z = 6$. This is appropriate because the source redshifts in our simulated sample are not correlated with their reference-band magnitudes; this choice was made to span a broad range of parameter space (similar to the choice made in Sect. 2.3 for not aligning the lens mass profile to
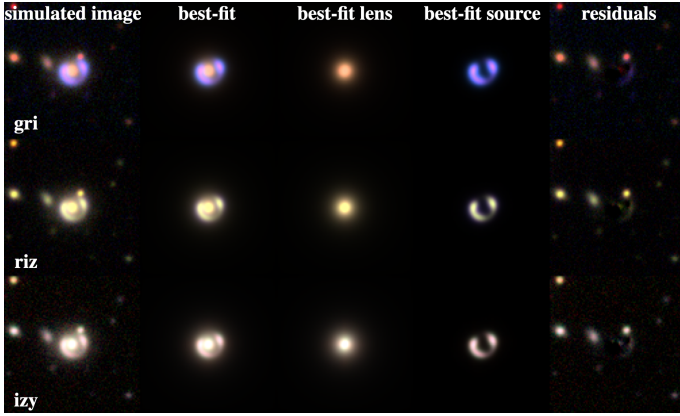
Fig. 3. Example of a two-component best-fit model to the *grizy* images of a simulated strongly lensed galaxy–galaxy system. From top to bottom: colour-composite *gri*, *riz*, and *izy* images. The first column from the left shows the simulated image. The second column shows the PSF-convolved best-fit image. The third and fourth columns show the best-fit lens-only and the best-fit source-only light profiles. The last column shows the residuals (best-fit-subtracted original image). The contaminants were masked by running the `SExtractor` on the *i*-band image.

the lens light). In the case of our 23 real strongly lensed galaxies, using either the same prior as used for the simulated sample or the magnitude-dependent prior of BPZ, truncated below the lens redshift, does not change the accuracy of photo-*z* estimation.

4. Systematics in source colour measurement

In this section we identify the main drivers of the source colour measurement error. We do this by assessing the accuracy of the method of Sect. 3.1 in measuring the source colours of our simulated sample. We measure the source colours and evaluate the uncertainties and biases of these measurements (Sect. 4.1). We hypothesise that the colour measurement errors are mainly correlated with the deviations of the lens light from a Sérsic profile and with the contrast between the lens and source image. We investigate the former in Sect. 4.2 and the latter in Sect. 4.3.

4.1. Colour measurement

Following the method of Sect. 3.1, we measure the source *i*-band magnitude and colours for each system in our simulated sample. The right-hand panel in Fig. 4 shows the best-fit source *i*-band magnitudes plotted against the true values. The error bars show the 1σ uncertainties inferred from the MCMC chains. The best-fit values have a -0.011 bias with respect to the true values and a 0.29 standard deviation. The inaccuracy of the *i*-band magnitude measurements is largely due to the lens model degeneracies: there is not a unique lens mass model that can reproduce a given strong lensing image configuration. In particular, the best-fit source magnitudes are degenerate with the best-fit lens mass model, resulting in large magnitude measurement inaccuracies.

The remaining four panels in Fig. 4 show the $g - i$, $r - i$, $i - z$, and $i - y$ best-fit source colours plotted against their true values. The error bars show the statistical 1σ uncertainties inferred from the MCMC chains. The best-fit $g - i$, $r - i$, $i - z$, and $i - y$ source colours have 0.007, -0.002, -0.006, and -0.015 bias around their true values and 0.03, 0.02, 0.03, and 0.06 standard deviations, respectively. The colour measurement is much more accurate than the magnitude measurement because gravitational lens-

ing is achromatic, which is why lens model degeneracies do not affect the colour measurement. The colour measurement errors are, on average, as low as the zero-point systematic uncertainty of the HSC (see e.g., Sect. 2.1, and Aihara et al. 2019).

The formal uncertainties inferred from the MCMC chains significantly underestimate the errors on the magnitude and colour measurements. In other words, the 1σ error bars of each plot in Fig. 4 (0.12 for the *i*-band magnitude and 0.011, 0.009, 0.007, and 0.008, respectively for the $g - i$, $r - i$, $i - z$, and $i - y$ colours) are significantly smaller than the standard deviation of the corresponding best-fit measurements around their true values (as measured earlier: 0.29 for the *i*-band magnitude and 0.03, 0.02, 0.03, and 0.06, respectively, for the $g-i, r-i, i-z$, and $i - y$ colours). This suggests that non-negligible systematic uncertainties are present in magnitude and colour measurement. Although the colour measurement errors and biases are significantly smaller than those of the magnitude measurement (*i*-band here), they can translate into significant systematics in photo-*z* estimation. In the following section, we investigate the main drivers of source colour measurement inaccuracy.

4.2. Accuracy of the lens light model

The cause for the biases in the measured colours must lie in the differences between the true surface brightness distribution of the lens systems and that of the model fitted to them. We describe the source galaxies with the same model family used to generate them, a Sérsic profile, and therefore that aspect of the model is accurate by construction. In contrast, the lens galaxies in our simulated sample are real objects whose surface brightness in general deviates from a Sérsic model. These deviations cause positive or negative lens light residuals in the model-subtracted images. The source light model overfits the leftover light at the source image position to minimise the χ^2 (Eq. (4)). Since the lens colours (and hence, the colours of their residuals) differ from the source colours, this overfitting results in best-fit source colours that deviate from the true values.

In this section we investigate how deviations of the lens light from a Sérsic profile affect the source colour measurement. First, we derive the expected correlation between the lens residuals at the source image position and the source colour measurement error by assuming that all the lens residuals at the source image position get overfitted by the best-fit source light model. We then use this correlation to derive the expected colour measurement error in terms of the deviations of the lens light from a Sérsic profile and the contrast between the lens and source image. We focus on the $g - i$ colour, C_{gi}, but the equations for the other colours are identical.

Since lensing is achromatic, in the absence of lens residuals the source $g - i$ colour is simply the difference between the g-band and i-band magnitudes of the source image[5]

$$C_{gi} = -2.5(\log_{10} I_0^g - \log_{10} I_0^i), \qquad (5)$$

where I_0^{band} is the source image flux in each band. In the presence of lens residuals, however, the measured colour becomes

$$C'_{gi} = -2.5\left[\log_{10}(I_0^g + \delta I^g) - \log_{10}(I_0^i + \delta I^i)\right], \qquad (6)$$

where δI^{band} is the lens residual flux at the source image position. It is straightforward to show that the source colour measurement

---

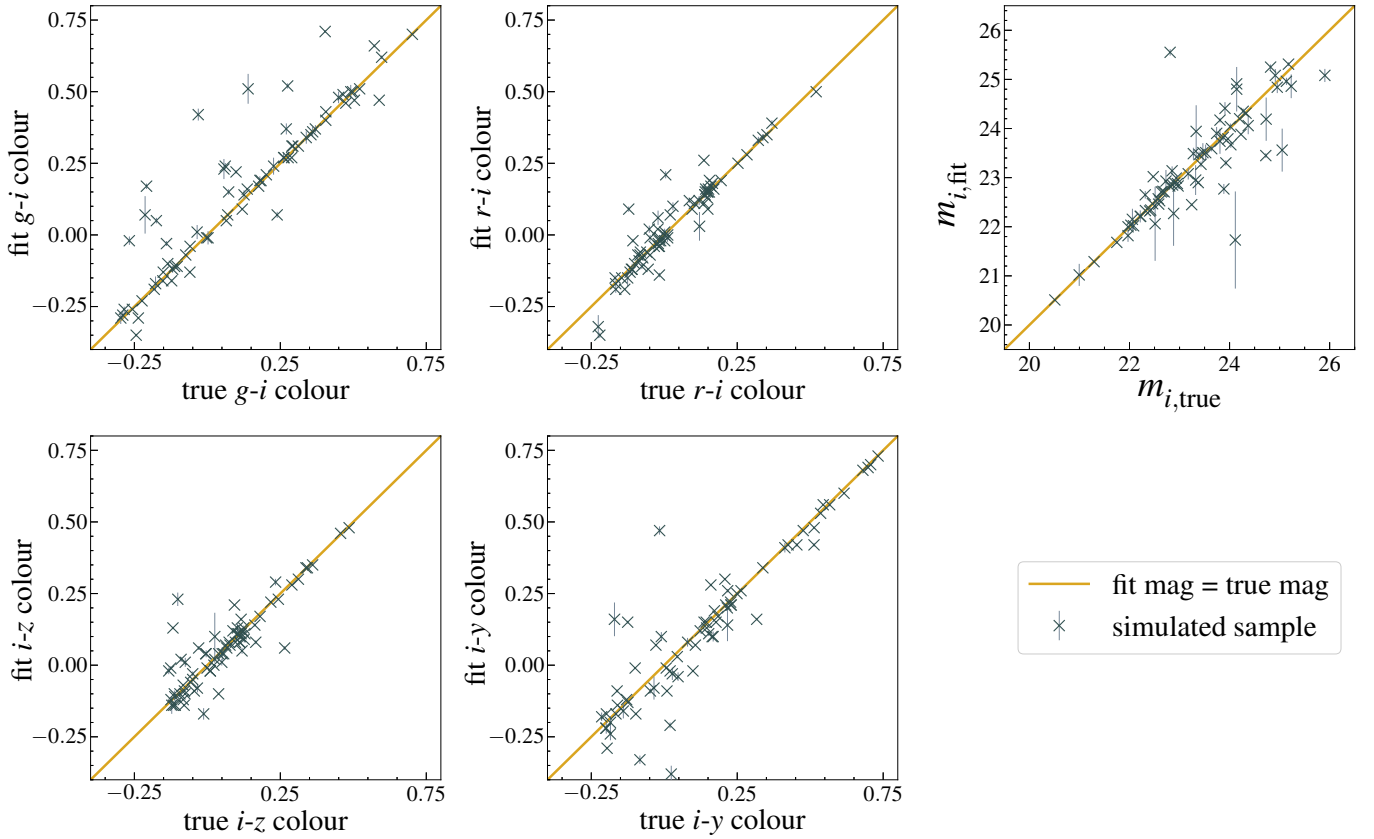[5] Throughout this paper the 'source image' refers to the image at the lens plane.

Fig. 4. Best-fit vs. true $g - i$, $r - i$, $i - z$, and $i - y$ source colours (four left panels), and the best-fit vs. true source i-band magnitudes (top right panel), for 73 strongly lensed galaxy–galaxy systems in our simulated sample. The dark yellow lines indicate where the fitted values equal the true values. The error bars show the 1σ uncertainties inferred from the MCMC chains. The best-fit i-band magnitudes have a 0.07 bias around their true values and a 0.29 standard deviation. The best-fit $g - i$, $r - i$, $i - z$, and $i - y$ colours have 0.007, -0.002, -0.006, and -0.015 bias around their true values and 0.03, 0.02, 0.03, and 0.06 standard deviations, respectively. The colour measurement is much more accurate than the magnitude measurement.

error ($\delta C_{gi} = C'_{gi} - C_{gi}$) is given by

$$\delta C_{gi} = -2.5 \log_{10} \frac{1 + \psi_g}{1 + \psi_i}, \tag{7}$$

where we define

$$\psi_{\text{band}} = \frac{\delta I^{\text{band}}}{I_0^{\text{band}}}. \tag{8}$$

In Appendix A we measure the same correlation from the best-fit models of Sect. 4.1 and compare it with the expected correlation from Eq. (7).

We then define the deviation of the lens light from a Sérsic profile at the source image position, α, as

$$\alpha_{\text{band}} = \frac{I_{\text{res}}^{\text{band}}(x_c, y_c)}{I_{\text{Sersic}}^{\text{band}}(x_c, y_c)}. \tag{9}$$

In each band, $I_{\text{res}}^{\text{band}}(x_c, y_c)$ is the lens residual flux and $I_{\text{Sersic}}^{\text{band}}(x_c, y_c)$ is the best-fit lens light model flux, both at a 3×3 pixel cutout (x_c, y_c) centred on the pixel that is brightest in the i-band (lensed) source image. Moreover, we introduce the λ-band contrast between the lens and source, Λ_λ, defined as the λ-band ratio of the best-fit lens light model to the best-fit (lensed) source light model at the (x_c, y_c) pixels (Λ_i of each system in our simulated sample is overwritten on its gri colour-composite

image in Fig. 2). With the above definitions, ψ_{band} in Eq. (8) can be re-written as

$$\psi_{\text{band}} = \alpha_{\text{band}} \Lambda_{\text{band}}. \tag{10}$$

Using Eq. (10) to insert for ψ_g and ψ_i, Eq. (7) becomes

$$\delta C_{gi} = -2.5 \log_{10} \frac{1 + \alpha_g \Lambda_g}{1 + \alpha_i \Lambda_i}. \tag{11}$$

In order to verify the validity of Eq. (11), we use it to calculate the expected δC_{gi} for each set of α_i, α_g, Λ_i, and Λ_g measured for the simulated systems, and compare it with the true δC_{gi} of the best-fit models (Sect. 4.1). Figure 5 shows the close agreements between the predicted δC_{gi} and the true δC_{gi} of the best-fit models. The 1σ scatter of the $|\delta C_{gi}|$ predicted with Eq. (11) is 0.02 and their bias is 0.002, both in close agreement with the $g - i$ colour measurement error and bias found in Sect. 4.1 (0.03 and 0.007, respectively).

Equation (11) can also be used to calculate the required accuracy of the lens light model (i.e., the maximum allowed α) as a function of Λ, to achieve source colour measurements that are more accurate than a desired threshold. For HSC-like data, a reasonable choice for the colour error threshold is $\delta C_{gi} \approx 0.03$ since this is the typical zero-point systematic uncertainty of HSC photometric measurements. For simplicity, we do this in the limiting case where Λ_g approaches 0, which allows us to reduce the dimensionality of the problem. This is a good approximation for
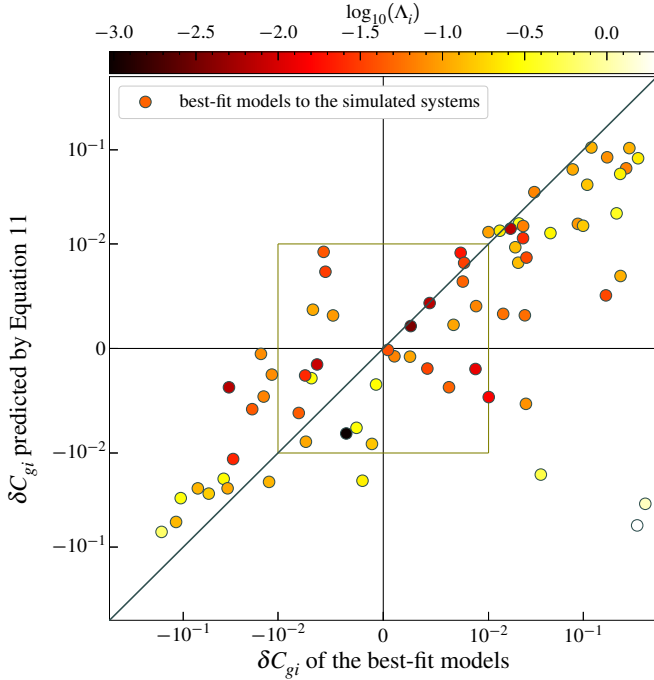
Fig. 5. Source $g-i$ colour measurement error δC_{gi} as predicted by inserting the α_i, α_g, Λ_i, and Λ_g (i-band and g-band deviation of the lens light from a Sérsic profile, and i-band and g-band ratio of the lens to source surface brightness measured at the image locations, respectively) of the best-fit models into Eq. (11) (derived correlation between δC_{gi}, α, and Λ), plotted against the true δC_{gi} of the best-fit models. The diagonal solid black line indicates were the predicted and true δC_{gi} are equal. The data point for each simulated system is colour-coded with its Λ_i value. The data points are plotted linearly within the inner olive box.

Fig. 6. Highest allowed deviation of the lens light from a Sérsic profile (α_i) as a function of the i-band contrast between the lens and source image (Λ_i) to achieve source colour measurements that are more accurate than a set threshold. The solid lines are generated from the derived correlation between δC_{gi}, Λ_i, and α_i (Eq. (12)). Each solid line shows α_i as a function of Λ_i, for a constant δC_{gi}. The value of δC_{gi} increases from 0.01 to 0.1, in 0.01 intervals (from bottom to top). The scattered data points show the distribution of α_i and Λ_i for the best-fit models to the simulated systems. The scattered data points are colour-coded with the δC_{gi} values of the best-fit models. In agreement with the derived correlation (Eq. (12)), the δC_{gi} of the best-fit models increases perpendicularly to the solid constant δC_{gi} lines. This figure indicates that Λ_i is the main parameter that determines the maximum allowed α_i to have δC_{gi} below a desired threshold.

our lens set since the lens galaxy is typically very faint in the g-band, compared to the source (see Sect. 3.2, Figs. 1 and 2 in Sonnenfeld et al. 2019a). We write α_i in terms of δC_{gi} and Λ_i:

$$\alpha_i = \frac{10^{\delta C_{gi}/2.5} - 1}{\Lambda_i}. \tag{12}$$

This formulation is particularly useful for understanding the correlations between δC_{gi} and α_i, and δC_{gi} and Λ_i, independently. The latter is of particular importance because Λ_i can be measured on real strongly lensed systems, while α_i is challenging to measure without the lens-only image (which is only available for the simulated sample).

Equation (12) shows that large deviations from a Sérsic profile (large values of α_i) can potentially lead to large errors on the colour. The impact of α_i, however, is modulated by the lens-source contrast Λ_i. For large values of Λ_i, for example, even small values of α_i can propagate to large values of δC_{gi}, and vice versa, for $\Lambda_i \approx 0$ the impact of the lens light contamination becomes negligible.

Figure 6 shows the required α_i as a function of Λ_i for a set of values of δC_{gi} threshold (solid lines). The error δC_{gi} increases from 0.01 to 0.1 from the bottom to the top in 0.01 intervals. The scattered data points show α_i vs. Λ_i for the best-fit models to the simulated systems, colour-coded with their δC_{gi}. As predicted, δC_{gi} of the best-fit models increase perpendicularly to the $\delta C_{gi} = $ constant lines (i.e., the same direction in which the predicted δC_{gi} of the solid lines increase). Figure 6 indicates that for a realistic range of lens light parameters and colours, such as those spanned by our simulation, the contrast between the lens
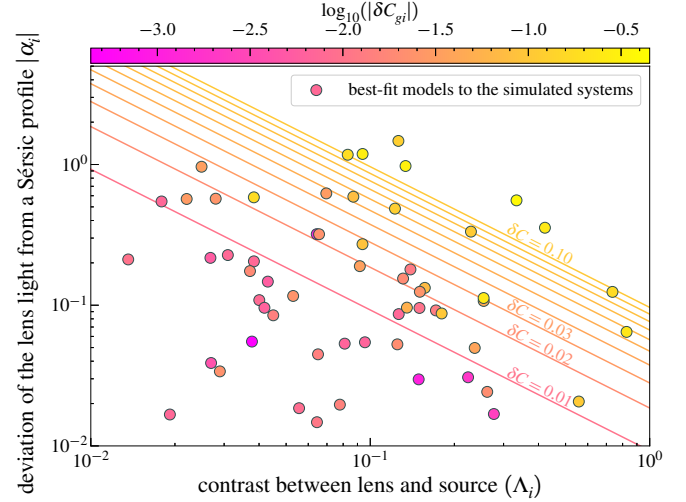
and source Λ_i is the main parameter that determines the required accuracy on the lens light model. In other words Λ_i is the main parameter that determines the maximum allowed α_i to have δC_{gi} below a desired threshold.

Equation (12) only applies to cases in which the lens residuals in one band are negligible. In the general case, which is relevant for the colours other than $g - i$ in our sample, it is necessary to use Eq. (11), which places joint constraints on the α_λs and Λ_λs in two bands. In this case, Eq. (11) cannot be simplified further since α_i and $\alpha_{\rm band}$ are not necessarily correlated. This is so because of the colour gradients in the lens light residuals, and is shown in Fig. 7, where α_g is plotted against α_i.

4.3. A proxy for the lens light contamination

As mentioned in Sect. 4.2, α_i is challenging to measure for real strongly lensed systems. Therefore, Eqs. (11) and (12) cannot be used to evaluate the accuracy of the best-fit colours (and consequently, the accuracy of the estimated photo-zs) for real strongly lensed galaxies. However, the contrast between the lens and source (Λ; see Sect. 4.2) depends only on the best-fit lens and the best-fit source light models, making it straightforward to measure for real strongly lensed systems. As indicated in Eqs. (11) and (12), although Λ_i is not the only parameter determining the δC_{gi}, we expect the δC_{gi} to be correlated with Λ_i. Hence, Λ_i can be used to evaluate the accuracy of the inferred colours of real strongly lensed galaxies. Here we investigate the correlation between δC_{gi} and Λ_i on the best-fit models.

Figure 8 shows the correlation between δC_{gi} and Λ_i. The slope of the best-fit power law (dashed line) is 0.563 ± 0.014. The 1σ scatter of the $\log_{10} \delta C_{gi}$ around the best-fit $\log_{10} \delta C_{gi}(\Lambda_i)$
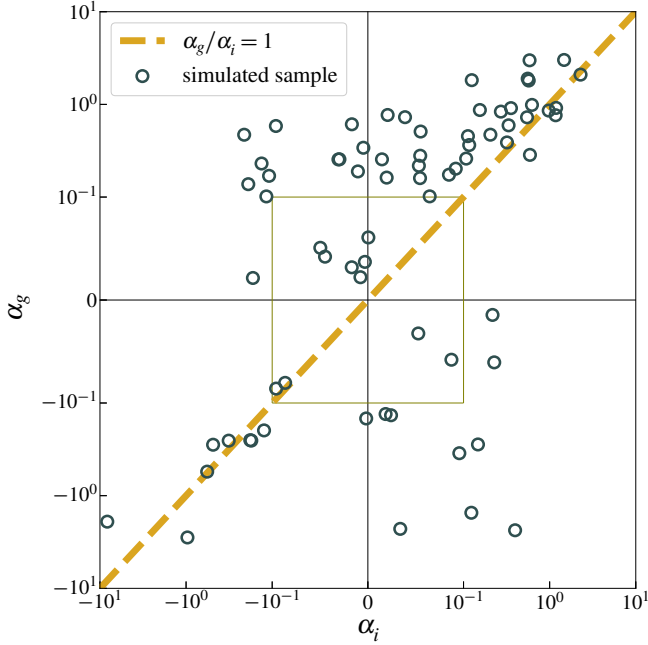
Fig. 7. Lens light residuals can have non-negligible colour gradients. This figure shows the distribution of the g-band and the i-band deviations of the lens light from a Sérsic profile (α_g and α_i) for the systems in our simulated sample. The dashed line indicates where residuals have no colour gradients (i.e., where $\alpha_g = \alpha_i$). The data is plotted linearly within the inner olive box. This figure shows that the deviations of the lens light from a Sérsic profile in different bands are not necessarily correlated.
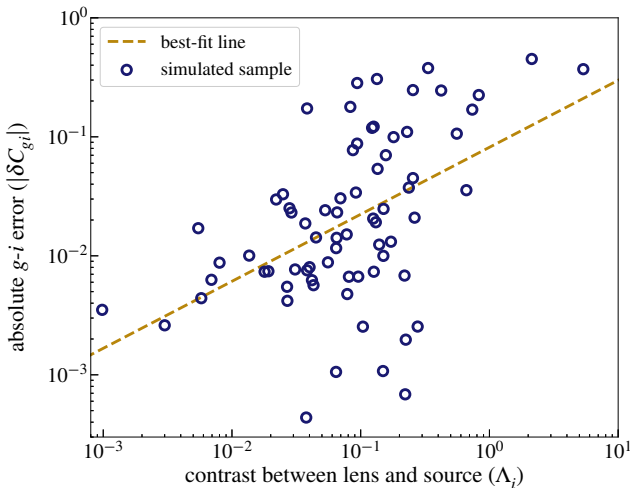

Fig. 8. Source colour measurement error $|\delta C_{gi}|$ plotted against the i-band contrast between the lens and source (Λ_i). The dashed line shows the best-fit power law, with a 0.563 ± 0.014 slope. For real strongly lensed systems this correlation can be used to select a Λ_i threshold below which the accuracy of source colour measurement is sufficient for a given application (e.g., source photo-z estimation).

is 0.508. This correlation is much weaker than the correlation between δC_{gi} and the deviations of the lens light from a Sérsic profile or the lens residuals (e.g., see Sect. 4.2, and Figs. 5, 6, and A.1), and cannot be used to estimate the δC_{gi} with high precision for individual systems. However, we can use this correlation to select a Λ_i threshold, below which the accuracy of source colour measurement is sufficient for a given application (e.g., source
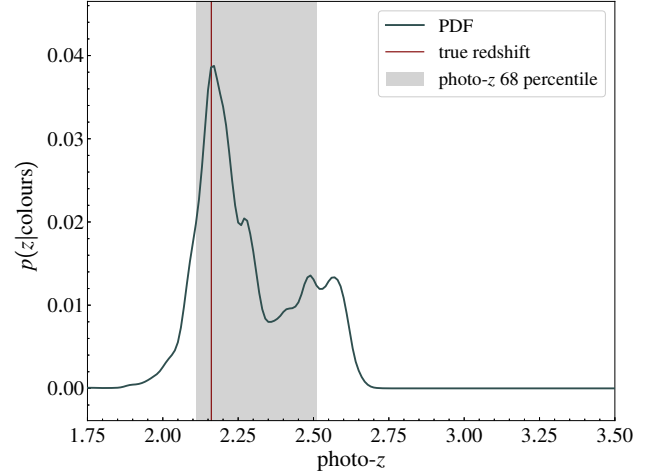

Fig. 9. Example of a normalised photo-z probability distribution function (PDF), obtained by running a photo-z algorithm (BPZ is used here) on the best-fit source colours and their uncertainties. The red line shows the true redshift, and the grey region shows the 68th percentile of the redshift PDF. The redshift with the highest probability is used as the estimated photo-z.

photo-z estimation). This is shown in Sect. 5.2, where we estimate the source photo-zs for our real lens sample.

5. Results: Photo-z estimation

5.1. Simulated sample

In this section we use the BPZ algorithm to estimate the source photo-zs of our simulated sample from the best-fit source magnitudes measured in Sect. 4.1. We showed in Sect. 4.1 that the colour measurement is significantly more accurate than the magnitude measurement. In other words, magnitude uncertainties are strongly correlated. The correct way to take these correlations into account would be to provide BPZ with the 5×5 covariance matrix of our magnitude measurements. This feature, however, is not supported in BPZ.

Hence, we instead ran the BPZ algorithm with the source colours and their uncertainties as follows. We selected the i-band as the reference band of the colour measurement by assuming a small i-band uncertainty (0.01), and used the 1σ scatter of the colour measurement errors (measured in Sect. 4.1) as the uncertainty on the non-reference bands: 0.03, 0.02, 0.03, and 0.06 on the g-, r-, z-, and y-bands, respectively.

Figure 9 shows an example of a normalised photo-z PDF, obtained by running the BPZ on a set of best-fit source colours (from Sect. 4.1) and uncertainties. We define the estimated photo-z as the redshift with the highest probability in this PDF. Figure 10 shows the estimated photo-zs for all the source galaxies in our simulated sample. The error bars show the 68% credible region of the redshift PDFs (corresponding to the shaded area in Fig. 9). The outlier fraction for the estimated photo-zs, which we define as the fraction of cases with $\delta z/(1 + z_{\text{true}}) > 0.15$, is 8.3%, and the 1σ scatter of the redshift errors ($\delta z/(1 + z_{\text{true}})$) is 0.032.

As discussed in Sect. 4, blending with the lens light is the main factor that complicates the source colour measurement, and consequently the photo-z estimation. We evaluated the impact of blending on the accuracy of estimated source photo-zs by testing our algorithm on a sample of simulated isolated galaxies consisting of all the unlensed sources from our simulated
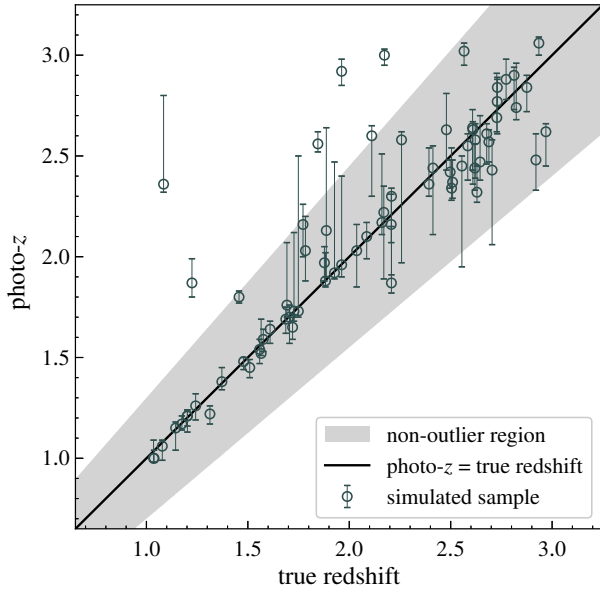
Fig. 10. Estimated photo-zs vs. true redshifts for the source galaxies in our simulated sample. The photo-zs were obtained by running the BPZ algorithm on the best-fit colours shown in Fig. 4. The black line shows where the photo-z equals the true redshift. The grey shaded region indicates where $\delta z/(1 + z_{\text{true}}) < 0.15$ (i.e., where the estimated redshift is not an outlier). The outlier fraction (fraction of the cases with $\delta z/(1 + z_{\text{true}}) > 0.15$) is 8.3% and the 1σ scatter of redshift errors ($\delta z/1 + z_{\text{true}}$) is 0.032. As in Fig. 9, the redshift with the highest probability in the redshift PDF was used as the estimated photo-z. The error bars in this figure correspond to the 68th percentile of the redshift PDFs (grey shaded region in Fig. 9.)

Fig. 11. Estimated photo-zs vs. the spectroscopic redshifts for the source galaxies in our sample of real strongly lensed galaxy–galaxy systems. The error bars show the 1σ uncertainty on the estimated redshift, and the grey shaded region indicates where the estimated redshift is not an outlier. The outlier fraction for the 15 systems with $\Lambda_i < 1$ (blue points) is 20%, while the outlier fraction for the remaining 8 systems with $\Lambda_i > 1$ (yellow points) is 75%. The source photo-z estimation is much more accurate for the systems with $\Lambda_i < 1$, due to their significantly more accurate source colour measurements (compared to the systems with $\Lambda_i > 1$).

strongly lensed sample. We fitted these simulated isolated galaxies with Sérsic light profiles to measure their colours, and ran the BPZ algorithm on the measured colours. We found that there are no outliers amongst the estimated photo-zs, and the 1σ scatter of the redshift errors ($\delta z/(1 + z_{\text{true}})$) is 0.018 (Fig. B.1). The details of these simulations, method, and results are available in Appendix B. This highly accurate photo-z estimation of simulated isolated galaxies implies that for simulated systems, where source colours are drawn from the SED templates of BPZ (see Sect. 2.3), blending (and not the performance of the photo-z algorithm) is the most important source of uncertainty in photo-z estimation; we discuss this in more detail in Sect. 6.1.

5.2. Real sample

In this section we estimate the source photo-zs for our sample of real strongly lensed galaxies. We measured the source magnitudes for our 23 real strongly lensed systems (Sect. 2.2) by following the approach described in Sect. 4.1. Here, unlike Sect. 4.1, visual inspection is our only means for identifying the contaminating objects missed by SExtractor. We identified six systems with one SExtractor-missed contaminant and one system with two SExtractor-missed contaminants, and modelled these contaminants with additional Sérsic light components.

We estimated the source photo-zs by running BPZ on the best-fit source colours. As in Sect. 5.1, we chose the i-band as the reference band by assuming a small i-band uncertainty (0.01). The uncertainty in the remaining bands consists of the uncertainties from the photometric zero-points and the uncertainties from the best-fit models. The zero-point uncertainties of the HSC fil-
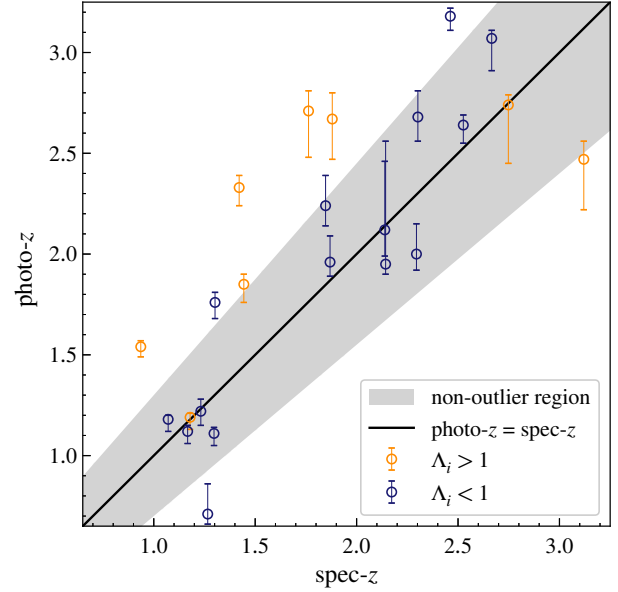
ters are available from Table 6 in Aihara et al. (2018b; see also Sect. 2.1). We used the colour measurement uncertainty estimates that were obtained by running our deblending algorithm on the simulated sample (Sect. 4.1) as the uncertainties from the best-fit models. These independent uncertainties add up to a total 0.04, 0.03, 0.04, and 0.06 uncertainty respectively for the g, r, z, and y filters. Figure 11 shows the estimated source photo-zs and the 68% credible region of their PDFs plotted against the spectroscopic redshifts. The outlier fraction (fraction of cases with $\delta z/(1 + z_{\text{spec}}) > 0.15$) is 39% and the 1σ scatter of redshift errors ($\delta z/(1 + z_{\text{spec}})$) is 0.157. The performance is thus much worse than that found for the simulated sample.

The source photo-z estimation inaccuracies are driven by the source colour measurement inaccuracies and the performance of the BPZ algorithm. We estimated the contribution of the latter by comparing the performance of our source photo-z estimation (on real strongly lensed galaxies) with that of BPZ on isolated galaxies in a similar $1 < z < 3$ redshift range. We did this by selecting a large sample (~2000) of such galaxies in the HSC PDR2 *grizy* photometry with spectroscopic redshifts available from The VIMOS VLT Deep Survey (VVDS, Le Fèvre et al. 2013) and 3D-HST (Skelton et al. 2014; Momcheva et al. 2016). Running BPZ on the CModel-measured *grizy* magnitudes results in a 20%–30% photo-z outlier fraction.

On the one hand, this result suggests that the higher outlier fraction in the photo-zs of the real strongly lensed sources, compared to the simulation case, is due to limitations of BPZ when dealing with relatively faint sources at $z > 1$. On the other hand, the performance is still significantly worse for the strongly lensed sources compared to the isolated galaxy case. Hence, systematic uncertainties of source colour measurement propagate significantly into source photo-z estimation, and are an

important driver of the higher photo-z outlier fraction of real lensed galaxies.

We show in Sect. 4.3 that the source colour measurement errors are correlated with Λ_i, and suggested that a Λ_i threshold can be used to identify the high-quality colour measurements before estimating the photo-zs. Here, we choose $\Lambda_i = 1$ as the threshold, and investigate whether it improves the source photo-z estimation; $\Lambda_i = 1$ is the limit above which the lens becomes brighter than the images at the image positions.

Figure 8 shows that $\Lambda_i = 1$ roughly corresponds to $\delta C_{gi} = 0.1$. In order to estimate the expected photo-z outlier fraction corresponding to this scatter on δC_{gi}, we conducted the following experiment. Following the approach of Sect. 2.3, we drew a large sample (10 000) of *grizy* magnitudes from the BPZ SED templates, with $1 < z < 3$. We added Gaussian scatters to the drawn magnitudes and then used the BPZ algorithm to estimate the photo-zs. We found that adding a 0.1 mag Gaussian scatter to each drawn magnitude results in a 31% outlier fraction and a 0.16 1σ scatter of the redshift errors. Since the 0.1 mag scatter roughly corresponds to $\Lambda_i = 1$ (Fig. 8), we conclude that achieving sub-30% photo-z outlier fractions with our method (comparable to the performance of the BPZ on isolated galaxies in the same redshift range; see further details in Sect. 6.1) requires sub-unity Λ_i.

Following the approach described in Sect. 4.3, we measured the Λ_i for all the systems in our real sample from their best-fit lens and source light models. We find 15 systems with $\Lambda_i < 1$ (in blue in Fig. 11), for which the photo-z outlier fraction is 20% and the 1σ scatter of redshift errors is 0.089. The remaining eight systems with $\Lambda_i > 1$ (in yellow in Fig. 11) have a 75% photo-z outlier fraction and a 0.312 1σ scatter of the redshift errors.

6. Discussion

6.1. Real sample vs. the simulated sample

The source photo-z estimation of the systems in the real sample is significantly less accurate than that of the systems in the simulated sample (39% vs. 8.3% photo-z outlier fraction). Even after imposing the $\Lambda_i < 1$ threshold, the photo-z outlier fraction of the real sample (20%) remains significantly larger than that of the simulated sample. The uncertainties in modelling the lens light are not expected to propagate significantly into the photo-z estimation of the real systems with $\Lambda_i < 1$ (see Sects. 4.3 and 5.2). Since the uncertainties on modelling the lens light are the main drivers of the uncertainties on source colour measurement (see Sect. 4.2), the less-accurate photo-z estimation of these systems can mainly be attributed to the performance of the BPZ algorithm.

It is straightforward for BPZ to fit the source colours of the simulated sample because these colours have been drawn from its own SED templates. This is evident from our experiment at the end of Sect. 5.1 (detailed in Appendix B); in the absence of lensing there are no photo-z outliers for simulated isolated galaxies in the $1 < z < 3$ range. This indicates that even the small 8.3% outlier fraction of the photo-z estimation of simulated strongly lensed galaxies in Sect. 5.1 was mainly caused by lensing. In contrast, real colours are much more challenging to fit by BPZ, and a 20−30% photo-z outlier fraction is typical for galaxies in $1 < z < 3$ with *grizy* photometry of the HSC (see Sect. 5.2). This is partly due to the template incompleteness of the BPZ at this redshift range. Moreover, the performance of the BPZ and similar SED-fitting algorithms drop when the 4000 Å break gets redshifted out of the reddest filter. This occurs because

the 4000 Å break is one of the primary spectral features used for fitting the spectra of galaxies in SED-fitting photo-z algorithms. The 4000 Å break gets shifted redwards of the *y*-band (the reddest filter of HSC, 9300−10 700 Å) at $z \sim 1.75$. As can be seen for the $z > 1.75$ galaxies in Fig. 10, this effect even propagates noticeably into the photo-z estimation of the simulated sources. The template incompleteness exacerbates these effects for the real galaxies at $z > 1.75$.

The higher photo-z outlier fraction of the real sample could also be caused by complications in modelling the surface brightness of real sources compared to the simulated ones; the source model is accurate by construction in the simulated sample, but not for the real sample. Additionally, simulating the lensing potential with a simple SIE model (and making it easy to fit by the SIE mass model in our deblending algorithm) contributes to the difference between the photo-z outlier fractions of the simulated and the real samples, although to a much lesser degree than the source light modelling (since lensing is achromatic).

Nevertheless, the inaccuracies of lens light modelling govern the source colour measurement inaccuracy. In Sect. 4, we showed that this is the case for our simulated lensed systems. We can argue that this is also the case for the real lensed systems; the estimated source photo-zs are dramatically more accurate for the real systems with $\Lambda_i < 1$ (for which the lens light modelling inaccuracies are not expected to propagate significantly into photo-z estimation) compared to the real systems with $\Lambda_i > 1$ (20% vs. 75% photo-z outlier fraction). Moreover, the accuracy of the estimated photo-zs for the real sources with $\Lambda_i < 1$ is consistent with the expected accuracy of BPZ for real galaxies in a similar redshift range (20−30% photo-z outlier fraction, if the *grizy* photometry of the HSC is used). Hence, other systematics (such as modelling the light of the real source with a Sérsic profile or modelling the lensing potential with an SIE) propagate into colour measurement to a much lesser degree than the lens light modelling inaccuracies.

Stage IV surveys are expected to improve the accuracy of estimated source photo-zs, thanks to their better-calibrated photometry and broader spectral coverage. In this work the relative zero-point calibration accuracy of HSC photometry sets the lower limit on the accuracy of source colour measurement (see Sects. 2.1, 4.2, 5.2, and 6.2). The better-calibrated photometry of *Euclid* and the LSST will lower this threshold, making the uncertainties from lens light modelling even more dominant than their current level with HSC photometry. Moreover, at $z > 1.7$, stage IV surveys (*Euclid* in particular) will enable photo-z estimations (for lensed sources as well as isolated galaxies) that are significantly more accurate than the HSC-based measurements. This is because stage IV surveys will provide photometry in near-infrared filters redder than the *y*-band of HSC, which will be instrumental in breaking the redshift degeneracies by detecting the 4000 Å break. Unless the contamination from the lens galaxy is properly dealt with, however, these improvements will only translate into more accurate photo-zs for the systems with a low Λ_i. This emphasises the importance of improving the deblending algorithms (lens light modelling in particular) in order to estimate accurate source photo-zs for large samples of lenses.

6.2. Lens light modelling

Since the uncertainties from the lens light model dominate the source colour measurement uncertainties (see Sect. 6.1), the accuracy of the source colour measurement and photo-z estimation can be estimated for any sample of real strongly lensed galaxies based on their Λ_i distributions. The large scatter on the best-fit

line in Fig. 8 prevents high-precision estimations of the accuracy of measured colours and estimated photo-zs for individual systems. However, systems with accurately measurable colours and photo-zs can be identified by imposing a Λ_i threshold, as shown in Sect. 5.2. Although it dramatically increases the accuracy of photo-z estimation, this practice hinders population studies that require complete and representative lens samples (e.g., for studying the inner structure of lens galaxies) by introducing non-negligible selection biases (e.g., towards the lens galaxies that are fainter or have a larger Einstein radius). Although followed-up source spectroscopic redshifts can compensate for the lack of accurate photo-zs at high Λ_i, follow-up spectroscopy campaigns can be carried out only for a small fraction of the large samples expected to be discovered with the LSST and *Euclid*. Hence, without accurate photo-zs at high Λ_i, our ability to take full advantage of these large samples remains limited.

Therefore, it is essential to improve this and similar methods by improving the deblending and photo-z algorithms. Improvements in the photo-z algorithm result in an overall increase in the accuracy of the estimated photo-zs. However, these improvements do not translate into increased source photo-z accuracy for the systems with a large Λ_i: inaccurate colours can hardly yield accurate photo-zs. Hence, improving the deblending algorithm is a necessity for systems with a high Λ_i.

For samples of lenses with similar properties to those considered in this work, modelling the lens light with a single Sérsic profile is the main source of colour measurement error. More complex lens light models, such as multi-component or non-parametrised descriptions of the surface brightness distribution, can potentially yield smaller lens light residuals (i.e., lower α_i) and a more accurate source colour measurement (see Sect. 4.2, Appendix A, and Figs. 6 and A.1 for the correlation between δC_{gi} and residuals).

6.3. Contaminating bright objects

One of the potential complications arising in applying the methods of this paper to estimate the source photo-zs of large samples of strong lenses (more than a few hundred systems) is the identification of contaminating bright objects. Contaminating objects that are faint and not closely aligned with the lensed system can be automatically identified and masked with segmenting algorithms such as SEXTRACTOR. However, these algorithms cannot distinguish the bright contaminants that are closely aligned with the lensed system from the lensed system itself. These missed contaminants, if unidentified (and not modelled), significantly increase the source colour measurement (and photo-z estimation) errors.

In this work visual inspection has been our only means for identifying the SExtractor-missed contaminating objects. However, repeating the same practice for the large upcoming samples is inefficient. Currently, this is an open problem for the analysis of large samples of lenses, and one that affects not only the measurement of the source photo-z, but more generally the lens modelling process.

6.4. Prior on the source properties

Inferring the redshift of a galaxy requires making prior assumptions on its probability distribution. In a Bayesian approach, such as the one adopted in this paper, the prior is explicitly included in the formalism. In methods based on machine learning, the choice of the distribution of the training sample acts as a prior on the

inference. In our experiment we assumed a flat prior in redshift–magnitude space. The rate of catastrophic failures in photo-z estimates, quantified by the outlier fraction, is not very sensitive to this choice: we verified that the results do not change significantly when adopting the default magnitude-dependent redshift prior of BPZ. However, assuming the wrong prior can systematically bias the inferred photo-zs, potentially limiting our ability to take advantage of the statistical combination of large samples of lenses.

Choosing the correct prior for a sample of strongly lensed sources is not a trivial task. The distribution in redshift-magnitude space of lensed galaxies differs from that of isolated objects because of strong lensing selection effects: the probability of a galaxy being strongly lensed increases as a function of redshift, and the lensing magnification allows us to detect intrinsically fainter objects. For this reason, a prior calibrated on a sample of isolated galaxies is not a good choice for strongly lensed sources. The main challenge in determining the correct prior distribution of lensed sources is that the redshift of a source is correlated with the mass distribution of the lens, and both of these properties are, in all practical cases, unknown. The correct way of solving this problem is to jointly describe the properties of the foreground and background galaxy population with a hierarchical model, following the method proposed by Sonnenfeld (2022b). This approach allows the mass distribution of the lenses and the redshift distribution of the sources to be inferred simultaneously; however, it requires the lensing selection effects to be explicitly accounted for.

7. Conclusion

We showed that the accuracy of the lens light model and the contrast between the lens and source are the main drivers of the uncertainties on source colour measurement and photo-z estimation. We modelled the galaxy-galaxy strong lens systems by simultaneously modelling the lensing potential with an SIE mass profile, the lens surface brightness with a Sérsic profile, and the source surface brightness with a Sérsic profile that is lensed through the lensing potential. If the lens is not brighter than the source at the source image position ($\Lambda_i < 1$), this modelling step provides source colours that are accurate enough to render the performance of the photo-z algorithm the main driver of the uncertainties on source photo-z estimation.

If the lens becomes brighter than the source at the source image position ($\Lambda_i > 1$), the propagation of the lens residuals at the source image position into the best-fit source light model reduces the ability of our deblending algorithm in measuring source colours that are sufficiently accurate for an accurate source photo-z estimation (i.e., δC_{gi} becomes larger than 0.1). In this scenario, the uncertainties in modelling the lens light become the dominant source of the uncertainties on source photo-z estimation.

We showed that the maximum allowed deviation of the lens light from a Sérsic profile (i.e., the maximum allowed α_i) that is required to measure source colours that are more accurate than a given threshold, decreases with increasing Λ_i (Sect. 4.2 and Fig. 6). Hence, δC_{gi} increases with increasing Λ_i. We inferred this correlation from a simulated sample of strongly lensed systems. Despite the large scatter on this relation, it can be used to filter the objects with accurately measurable source colours.

However, filtering introduces non-negligible selection biases towards lenses that are fainter or have a larger Einstein radius. This in turn hinders our ability to take full advantage of the current and upcoming large samples of strongly lensed galaxy-galaxy

systems (from the ESA-*Euclid*, the *Rubin*-LSST, and the CSST) in studying complete and representative samples of lens galaxies (which are needed to map the properties of the lens population to those of the general galaxy population, as shown by Sonnenfeld 2022a). Obtaining complete lens samples requires improvements in both the photo-*z* and the deblending algorithms. In particular, more advanced lens light models (compared to the single Sérsic profile used here) that leave behind smaller residuals will raise the Λ_i threshold below which source colours can be measured accurately. In the case that obtaining accurate photo-*z*s of complete samples of strong lenses proves difficult, an alternative approach is to use information on the population distribution of galaxy redshifts, as opposed to individual source redshift measurements, and to incorporate it into a Bayesian hierarchical model that takes lensing selection effects into account. This approach is discussed in detail by Sonnenfeld (2022b).

Significant improvements in both the deblending and photo-*z* algorithms remain essential before the launch of Euclid and the LSST. Nonetheless, this work is a first exploratory step in constructing the pipeline for estimating the source redshifts of the upcoming samples of strong lenses.

References

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, PASJ, 70, S4
Aihara, H., Armstrong, R., Bickerton, S., et al. 2018b, PASJ, 70, S8
Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, ApJ, 705, 1099
Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, ApJ, 724, 511
Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., & Bolton, A. S. 2011, MNRAS, 415, 2215
Benítez, N. 2000, ApJ, 536, 571
Benítez, N., Ford, H., Bouwens, R., et al. 2004, ApJS, 150, 1
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Bolton, A. S., Brownstein, J. R., Kochanek, C. S., et al. 2012, ApJ, 757, 82
Cañameras, R., Schuldt, S., Shu, Y., et al. 2021, A&A, 653, L6
Ciotti, L., & Bertin, G. 1999, A&A, 352, 447
Coe, D., Benítez, N., Sánchez, S. F., et al. 2006, AJ, 132, 926
Coleman, G. D., Wu, C. C., & Weedman, D. W. 1980, ApJS, 43, 393
Collett, T. E. 2015, ApJ, 811, 20
Dark Energy Survey Collaboration (Abbott, T., et al.) 2016, MNRAS, 460, 1270
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3
Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34
Finkbeiner, D. P., Schlafly, E. F., Schlegel, D. J., et al. 2016, ApJ, 822, 66
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144
Gilman, D., Birrer, S., Nierenberg, A., et al. 2020, MNRAS, 491, 6077
Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ApJ, 823, 37
Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, A&A, 647, A124
Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, ApJ, 709, 644
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17
Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. 2018, PASJ, 70, 66
Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, MNRAS, 421, 1007
Kinney, A. L., Calzetti, D., Bohlin, R. C., et al. 1996, ApJ, 467, 38
Kormann, R., Schneider, P., & Bartelmann, M. 1994, A&A, 284, 285
Kormendy, J., Fisher, D. B., Cornell, M. E., & Bender, R. 2009, ApJS, 182, 216
Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Lilly, S. J., Le Fevre, O., Crampton, D., Hammer, F., & Tresse, L. 1995, ApJ, 455, 50
Madau, P. 1995, ApJ, 441, 18
Magnier, E. A., Schlafly, E., Finkbeiner, D., et al. 2013, ApJS, 205, 20
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Myles, J., Alarcon, A., Amon, A., et al. 2021, MNRAS, 505, 4249
Newman, A. B., Ellis, R. S., & Treu, T. 2015, ApJ, 814, 26
Oguri, M. 2006, MNRAS, 367, 1241
Oldham, L. J., & Auger, M. W. 2018, MNRAS, 476, 133
Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879
Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73
Ruff, A. J., Gavazzi, R., Marshall, P. J., et al. 2011, ApJ, 727, 96
Salvato, M., Ilbert, O., & Hoyle, B. 2019, Nat. Astron., 3, 212
Sawicki, M. J., Lin, H., & Yee, H. K. C. 1997, AJ, 113, 1
Schlafly, E. F., Finkbeiner, D. P., Jurić, M., et al. 2012, ApJ, 756, 158
Schlegel, D., Kollmeier, J. A., & Ferraro, S. 2019, Bull. Am. Astron. Soc., 51, 229
Schuldt, S., Chirivì, G., Suyu, S. H., et al. 2019, A&A, 631, A40
Sersic, J. L. 1968, Atlas de Galaxias Australes (Cordoba: Observatorio Astronomico)
Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. 2021, MNRAS, 503, 2380
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Smith, R. J., Lucey, J. R., & Conroy, C. 2015, MNRAS, 449, 3441
Sonnenfeld, A. 2022a, A&A, 659, A132
Sonnenfeld, A. 2022b, A&A, 659, A133
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2012, ApJ, 752, 163
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013, ApJ, 777, 98
Sonnenfeld, A., Treu, T., Marshall, P. J., et al. 2015, ApJ, 800, 94
Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29
Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019a, A&A, 630, A71
Sonnenfeld, A., Wang, W., & Bahcall, N. 2019b, A&A, 622, A30
Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148
Storfer, C., Huang, X., Gu, A., et al. 2022, ArXiv e-prints [arXiv:2206.02764]
Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9
The MSE Science Team (Babusiaux, C., et al.) 2019, ArXiv e-prints [arXiv:1904.04907]
Tran, K.-V. H., Harshan, A., Glazebrook, K., et al. 2022, AJ, 164, 148
Treu, T., & Koopmans, L. V. E. 2004, ApJ, 611, 739
Treu, T., Auger, M. W., Koopmans, L. V. E., et al. 2010, ApJ, 709, 1195
Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969
Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, AJ, 112, 1335
Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107
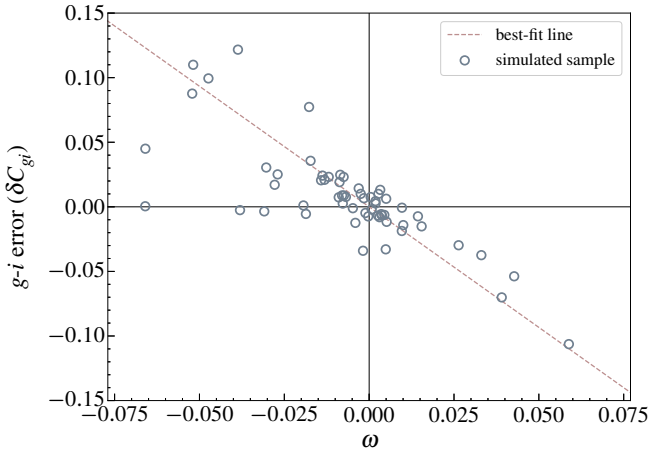
Appendix A: Lens residuals

Fig. A.1. Source colour measurement error correlates with the lens light residuals. This figure shows the error of the measured source $g - i$ colours (δC_{gi}), plotted against the residual ratio parameter (ω, defined in Equations A.1 and A.3). The dashed line shows the best-fit line passing through the origin with a -1.867 ± 0.028 slope. The thin grey lines show the $\omega = 0$ and the $\delta C_{gi} = 0$ lines.

Equation 7 can be used to predict the $g - i$ colour measurement error, δC_{gi}, for each best-fit model to the simulated systems. Here, we compare these predicted δC_{gi} against the true δC_{gi} of the best-fit models (measured in Section 4.1). We start by re-writing Equation 7. Defining the residual ratio parameter ω as

$$\omega = \log_{10} \frac{1 + \psi_g}{1 + \psi_i} \, , \qquad (A.1)$$

Equation 7 can be re-written as

$$\delta C_{gi} = -2.5\omega \, . \qquad (A.2)$$

From Equation A.2, we expect the colour measurement error to be inversely proportional to ω. We want to compare this prediction with the results from our fit to the simulated lenses.

Measuring ω on the best-fit models of Section 4.1 requires specifying the definitions of the source image position and the lens residuals at the source image position. As in Section 2.3, we define the source footprint as the g-band source-only pixels with values higher than $3\sigma_{\text{sky}}$. With this definition, the source footprint can consist of multiple disconnected patches. We define the total lens residuals at the source image position in each band ($\Sigma_{\text{res,band}}$) as the sum of all the pixels of the lens residual image that lie within the source footprint (the lens residual image is produced by subtracting the best-fit lens light model from the lens original image). Similarly, we define the total source light in each band ($\Sigma_{\text{source,band}}$), as the sum of all the pixels of the source-only image that lie within the source footprint. Using these definitions, ψ_{band} in Equation 8 can be re-written as

$$\psi_{\text{band}} \approx \frac{\Sigma_{\text{res,band}}}{\Sigma_{\text{source,band}}} \, , \qquad (A.3)$$

which allows ω to be measured on the best-fit models.

Figure A.1 shows the correlation between the δC_{gi} and ω. We fitted a linear model with zero intercept to this distribution, finding a best-fit slope of -1.867 ± 0.028 (dashed line). This correlation simply indicates that $|\delta C_{gi}|$ grows with the ratio of the
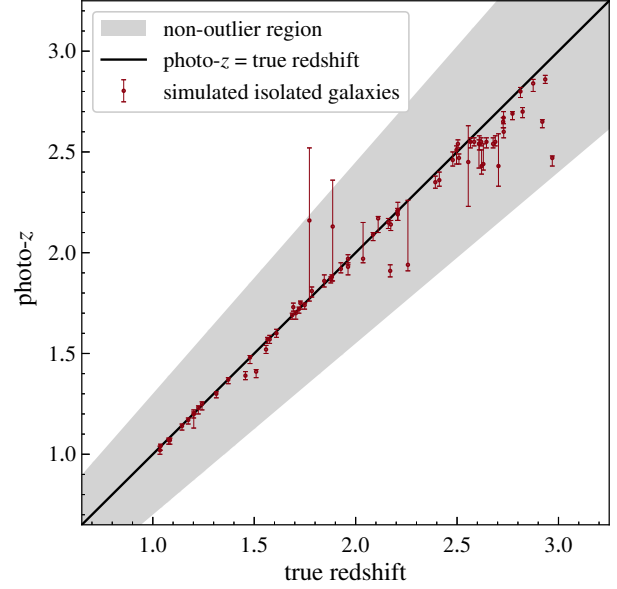
Fig. B.1. Estimated photo-zs vs the true redshifts for our simulated isolated galaxies. The black line shows where photo-zs equal true redshifts. The error bars correspond to the 68th percentile of the redshift PDFs. The grey shaded region shows where $\delta z/(1 + z_{\text{true}}) < 0.15$. There are no photo-$z$ outliers, and the 1σ scatter of redshift errors ($\delta z/1 + z_{\text{true}}$) is 0.018.

lens residual flux to the source flux at the source image position, and is positive (negative) if the $g - i$ colour of the residuals is bigger (smaller) than the true source $g - i$ colour. The slope inferred from the best-fit models is shallower than the expected slope from Equation A.2 (-2.5). This discrepancy is rooted in our initial assumption that all the lens residuals at the source image position get overfitted by the best-fit source light model (see Section 4.2). The weaker correlation probed here indicates that this is not necessarily the case, and that the -2.5 slope should be treated as the most conservative limit.

Appendix B: Photometric redshift estimation of simulated isolated galaxies

In Section 5.1 we evaluate the performance of our algorithm in estimating the photo-zs of simulated strongly lensed galaxies. Here we use a simulated sample of isolated galaxies to evaluate the impact of lensing on the accuracy of photo-z estimation. For each strongly lensed system in our simulated sample (Section 2.3), we use the unlensed image of the source galaxy in each *grizy* band as the simulated isolated galaxy. In each band, we convolve this image with the PSF of the lens galaxy that was originally assigned to this source in our simulated sample of strongly lensed systems (Section 2.3). We do this to limit the variations between the sample of simulated isolated galaxies and the sample of simulated strongly lensed systems to lensing. In each band we centre the PSF-convolved image of the simulated isolated galaxy on the central pixel of a 101×101 pixel cutout.

We measure the colours by fitting a single Sérsic profile to each isolated galaxy, using a similar method to that used in Section 3.1 to fit the surface brightness of lens galaxies. The best-fit $g - i$, $r - i$, $i - z$, and $i - y$ source colours have ~ 0.004 standard deviations and ≤ 0.0005 bias around their true values. We estimate the photo-zs by running the BPZ algorithm with the best-fit colours and their uncertainties (0.004 for all the colours,

see Section 5.1). In order to isolate the variations between the experiment here and that of Section 5.1 to lensing, as in Section 5.1, we use a flat prior for the reference-band magnitude dependence of redshift. Figure B.1 shows the estimated photo-zs for all the simulated isolated galaxies. The error bars show the 68th percentile of the redshift PDFs. There are no outliers amongst the estimated photo-zs, and the 1σ scatter of the redshift errors ($\delta z/(1 + z_{\mathrm{true}})$) is 0.018. In contrast, the photo-z estimation of simulated strongly lensed galaxies had an 8.3% outlier fraction and a 0.032 1σ scatter of redshift errors (Section 5.1).