4. Simulating a training dataset

We follow the data-driven approach introduced by Cañameras et al. (2020) as well as Rojas et al. (2022) and Savary et al. (2022). This requires to select a set of real r-band images of potential deflector galaxies. A composite mass model is then associated to each galaxy, composed of a baryonic mass profile plus dark matter halo. We then simulate lensing of real galaxies from deep HST imaging data, and add the resulting arcs on the CFIS images of our selection of deflectors. For all the lensing calculations and image simulations we use the Python package `Lenstronomy` (Birrer & Amara 2018; Birrer et al. 2021), and we use `Astropy` (Astropy Collaboration et al. 2022) to handle the imaging data. Figure 1 gives a visual summary of the whole simulation chain. We describe below the different elements of the process.

4.1. Selection of potential CFIS deflectors

The first step of the process is to select edge-on late-type galaxies in our CFIS data. In doing this, we apply the same ellipticity cut as Sygnet et al. (2010), keep the range $0.6 < e < 0.92$. We use the same definition of the ellipticity as these authors:

$$e \equiv \frac{a^2 - b^2}{a^2 + b^2}, \tag{4}$$

where a and b are respectively the semi-major and semi-minor axes of the light distribution. The lower limit on the ellipticity removes most of the non-edge-on galaxies while the upper limit removes artefacts such as dead columns, stellar spikes, fast asteroids or any other spurious objects with extreme ellipticities.

Then, we cross-match the selected galaxies with the Sloan Digital Sky Survey (SDSS) data release 18 (Almeida et al. 2023) to obtain photometric redshifts (photo-z), K-corrections, and distance moduli. We keep only those galaxies with `photoErrorClass` = 1 in the SDSS catalog, ensuring the best possible estimation of the photo-z error.

Finally, we remove bad deflectors, which we define as galaxies that, based on their r band image, would require a mass-to-light ratio larger than 12 to generate a lens with Einstein radius of $R_E = 1.1''$, assuming a Single Isothermal Ellipsoid (SIE) model for the mass and a source at redshift $z_s = 3.0$. Our lower-limit on the Einstein radius corresponds to an angular distance between the lens light and the simulated lensed source that allows to see obvious signs of lensing. In total, we have 649 340 galaxies useful as potential deflectors. For consistency with the K-corrections and photo-z, we use the same cosmology as SDSS for all relevant calculations, adopting $\Omega_m = 0.2739$ and $h = 0.705$.

4.2. Light and mass models

To simulate the mass distribution of the deflector galaxies we use a composite model where we treat baryons and dark matter independently. This is preferable to a power-law or an isothermal model because for this type of galaxies the bulge and disc components dominate the lensing mass over the dark halo.

For the baryonic part of the composite model we use the light profile, as seen in the actual data,normalized by a constant stellar mass-to-light ratio, Υ. For the dark halo component we use a spherical Navarro-Frenk-White profile (NFW, see Navarro et al. 1996). Since the light profiles of the deflectors are highly elliptical, we approximate an elliptical Sersic light profile as the difference between two cored elliptical isothermal profiles. This is also known as a "Chameleon" profile used in (Dutton et al. 2011). We follow the parametrization of this profile introduced by Suyu et al. (2014):

$$I(\boldsymbol{\xi}, w_i, q) = \frac{1}{\sqrt{\xi_1^2 + \xi_2^2/q^2 + 4w_i^2/(1+q)^2}}, \tag{5}$$

$$L_c(\boldsymbol{\xi}, w_c, w_t, q, L_0) = \frac{L_0}{1+q}\left[I(\boldsymbol{\xi}, w_c, q) - I(\boldsymbol{\xi}, w_t, q)\right], \tag{6}$$

where $\boldsymbol{\xi} \equiv (\xi_1, \xi_2) = \boldsymbol{\theta} - \boldsymbol{\theta_0}$ is the angular position relative to the center of the profile, w_i is the size of the core, q is the ellipticity, and where L_0 is the surface brightness at $1''$ from the center, along the major axis. In addition, w_c and w_t are respectively the sizes of the two profiles that are subtracted from each other, with the additional condition that $w_t > w_c$.

Because in practice our lens galaxies can display a prominent bulge component, we use a Double Chameleon profile to represent their light distribution. The Double Chameleon is parameterized as the sum of two concentric Chameleon profiles, each one with its own ellipticity and core sizes:

$$L_d(\boldsymbol{\xi}, \boldsymbol{X_1}, \boldsymbol{X_2}, r, L_0) = L_c\left(\boldsymbol{\xi}, \boldsymbol{X_1}, \frac{L_0}{1+r}\right) + L_c\left(\boldsymbol{\xi}, \boldsymbol{X_2}, \frac{L_0}{1+1/r}\right), \tag{7}$$

where the vector $\boldsymbol{X_i} \equiv (w_{ic}, w_{it}, q_i)$ corresponds a the single Chameleon parameters, and where r is to the ratio between the profiles along the major axis, $1''$ away from the common centers of each profile. We the double Chameleon profile is used as a proxy for the baryonic mass, we normalize it by replacing L_0 with α_1, where α_1 is now the deflection angle along the major axis at $1''$ away from the center.

Similarly, the spherical NFW profile depends only on its characteristic radius, R_s, and on the deflection angle at its characteristic radius, α_{Rs}. `Lenstronomy` offers conversion back and forth between dimensionless angular units, R_s and α_{Rs}, and physical units, M_{200} and c. As usual, the latter are defined as the mass enclosed by the radius at which the density is 200 times the critical density, R_{200}, and as the concentration $c = R_{200}/R_s$. We use the scaling relations described in Duffy et al. (2008) to estimate the concentration c for every M_{200}.

We fit the Chameleon and Double Chameleon models to the light of every potential deflector using the r-band stamps and a circular pixel mask with radius $4.5''$ around the barycenter of the light distributions. To do so, we use a log-likelihood that combines the standard pixel-wise noise-normalized square error with penalties for ellipticities larger than $e > 0.98$ and for negative fluxes. We minimize the log-likelihood using first a Particle Swarm Optimizer (Kennedy & Eberhart 1995) with 200 particles and 50 iterations, and then improve the minimization using Downhill simplex optimization (Nelder & Mead 1965) with a maximum of 3000 iterations. Then, we use the reduced χ^2 to choose the best option between the Chameleon and Double Chameleon. With no surprize, deflectors with prominent bulges are better modeled by the Double Chameleon, while for bulgeless discs the single Chameleon is sufficient. Good models easily reach a reduced χ^2 close 1 since we have reliable estimates of the pixel-wise noise via the RMS maps provided by the CFIS reduction pipeline. Finally, we discard deflectors with $\chi^2 > 3$. The whole process leaves us with a total of 558 199 potential deflectors with good light models.

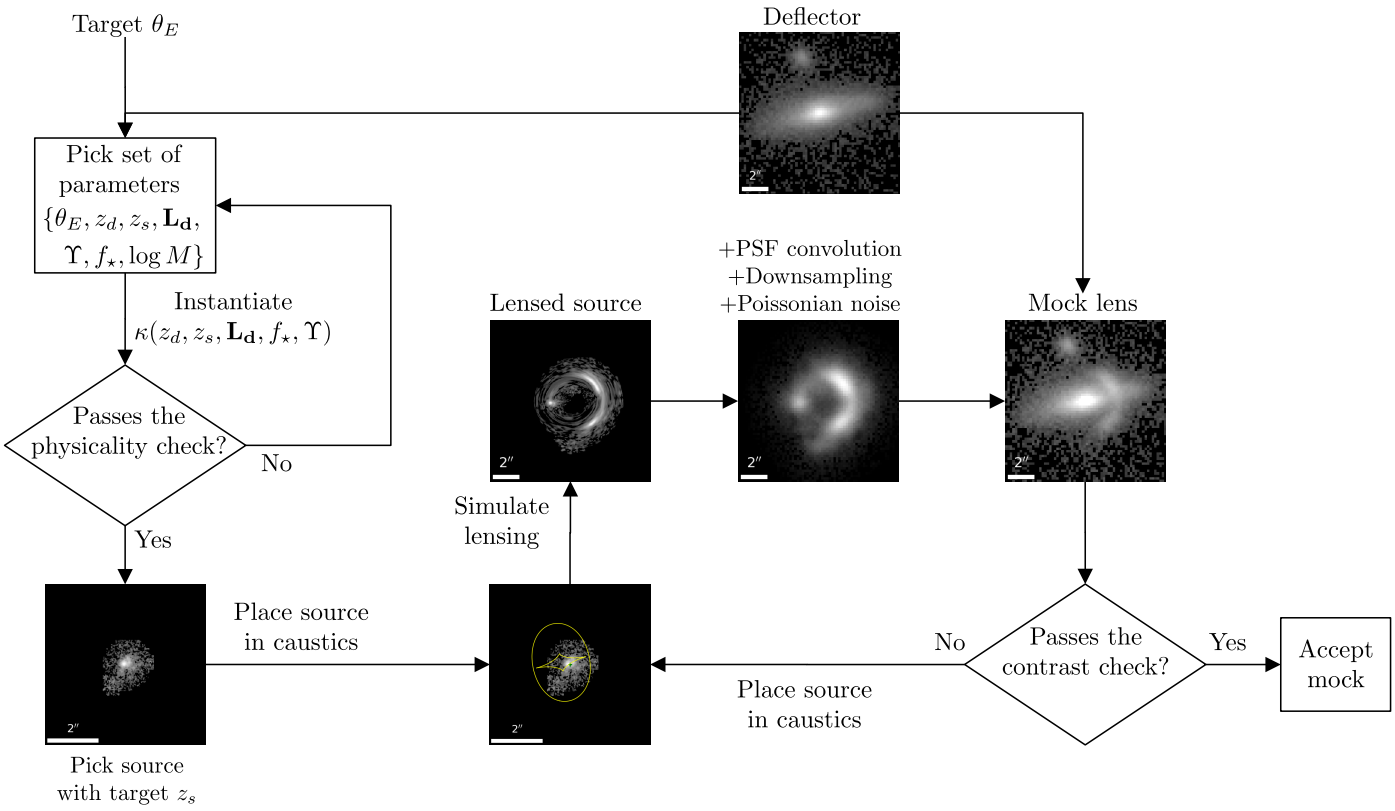

Fig. 1. Schematic description of our lens simulation procedure. To produce each simulated stamp, we start with a potential deflector and a target Einstein radius, θ_E. Then, we pick a set of parameters for a mass model consistent with the target θ_E and the deflector's light, and perform a check where we make sure that $\log M$, $f_\star$, and Υ are within physical ranges. After that, we pick a source using the target z_s provided by the mass model parameters, and place it randomly around the caustics before simulating its lensing by the mass model. Next, we convolve the lensed source image with the deflector's psf, then downsample to match the pixel size of CFIS, and add Poissonian noise. At this point, we add the original stamp of the deflector to get the final mock. Lastly, we perform a contrast check to make sure that the lensing features are visible. If it passes the check, we accept the mock and proceed to the next deflector. Otherwise, we reposition the source relative to the caustics and try a second time.

4.3. Reparametrizing the mass models

The default parametrization of mass models in `Lenstronomy` is dimensionless. This means that models are parameterized in terms of the deflection angle caused by their associated convergence map, rather than in terms of their total mass or mass density. This allows for the modeling of lens systems while remaining agnostic about the exact cosmology, redshifts, and masses involved. Of course, once a cosmology and redshifts are considered, the dimensionless convergences take on a physical meaning and can be expressed as surface mass densities, which is needed if we want to use scaling relations to associate mass to our light maps of the lenses. Since we adopt the same cosmology used in SDSS, the total convergence of each potential deflector is completely determined by the redshifts of the deflector, z_d, and of the source ,z_s, as well as by the deflection due to the (Double) Chameleon profile α_1, the M_{200} parameter of the NFW, and the light model of the deflector, $\boldsymbol{L}$.

However, in order to avoid "unphysical" or unreal-looking lenses, we impose several physical constrains on the mass models. More precisely, we force the stellar mass-to-light ratio to be in the range $4.0 < \Upsilon < 9.0$, the fraction of stellar to total mass to be $0.4 < f_\star < 0.8$, and the total mass enclosed by the tangential critical lines to be $10^{10} < M < 10^{12}\,M_\odot$. In order to do this efficiently we re-parametrize the total convergence from $\kappa\,(z_d, z_s, \mathbf{L}, \alpha_1, M_{200})$ to $\kappa\,(z_d, z_s, \mathbf{L}, \Upsilon, f_\star)$.

For every potential deflector, we create a regular grid of parameters:

- $z_s = 10^x$, with x uniformly sampled between $\log(z_d)$ and 3.0
- $\Upsilon \in \{4, 5, 6, 7, 8, 9\}$
- $f_\star \in \{0.4, 0.48, 0.56, 0.64, 0.72, 0.8\}$

For every combination of parameters in the above grid, z_s, Υ, $f_\star$, we find the equivalent combination of z_s, α_1 and $\log(M_{200})$ and we then compute $\log(M)$ and θ_E. For these searches we use the bisection method as implemented in `scipy` (Virtanen et al. 2020). We save all the compatible values of z_s, $\mathbf{L}$, Υ, $f_\star$, α_1, $\log(M_{200})$, $\log(M)$, and θ_E for each deflector. Following that, we estimate the values in-between the grid sampling using the "Random Sample Consensus" algorithm with a linear regression as base regressor (Fischler & Bolles 1981), as implemented in `scikit-learn` (Pedregosa et al. 2011). For this regression we normalize the dataset such that each parameter has a mean of 0 and a standard deviation of 1, and we generate polynomial features using a polynomial of degree 3, rendering a total of 120 features. We take two-thirds of the dataset for training and leave the remaining one-third for testing. We keep the interpolated values only if the accuracy on the test dataset is larger than 98%. This has to be done independently for each deflector since the relationship between the seven parameters depends on the redshift of the deflector and its luminosity. The output of this procedure is a well populated grid of reasonable mass models, such that for every potential deflector and every combination of Υ, $f_\star$, $\log(M)$, and θ_E; we have the corresponding z_s, α_1, and M_{200}.

4.4. Generating a mock lens from a potential deflector

There is evidence that ML classifiers benefit from a "flat" Einstein radius distribution when training on simulated lenses (Canameras et al. 2023). Consequently, we generate mocks following a flat Einstein radius distribution between 1.2″ and 2.6″. To do so, we fix the target Einstein radius, θ_E, of every mock lens down to a precision of 0.1″ while respecting the constrains described in Sect. 4.3. Then, for every target Einstein radius we generate the same number of mocks, thus ensuring a flat distribution down to a precision of 0.1″. potential To produce a single mock lens, we start from a deflector with a good light model, a good photometric redshift, and a list of good mass models. We randomly select a mass model that complies with the physical constrains and target θ_E. This becomes the "target mass model". Then, we choose a potential source at a redshift equal to the value given by the target mass model with a tolerance of ±0.02. Afterwards, we instantiate the convergence using `Lenstronomy`'s standard parametrization and the parameters α_1 and M_{200} given by the target mass model. We calculate Υ, $f_\star$, $\log(M)$ and θ_E, and check that they are consistent with the target mass model. If any of the parameters is off by more than 25%, or if the sum of the relative errors was larger than 50%, we try again with a different target mass model. If this check fails for a second time, we skip the deflector and proceed to the next one. Otherwise, we go on to calculate the caustic lines of the system.

We randomly select between the tangential and radial caustics and place the central pixel of the source inside a polygon with the same shape as the selected caustic, but with a 15% larger perimeter. We favor the selection of the tangential over the radial caustic with a probability of 75% against 25%. Before simulating the lensing of the source by the convergence of the deflector, we interpolate the light of the source from the HST/HSC image and rotate it by a random angle. The lensed image is generated with a pixel size exactly 6 times smaller than that of the deflector. We convolve the high resolution lensed source image with the corresponding PSF model of the deflector. Then, we downsample the resulting image to CFIS' pixel size of 0.1857″and add Poissonian noise. The final mock lens is the sum of the deflector with the simulated image of the lensed source.

At this stage all the simulated lenses are physically motivated images mocking CFIS observations of gravitational lensing. However, the lensing features can be too faint or too blended with the deflector's light, up to the point that it becomes impossible for human experts to discern the lensing features. This is often the case when simulating small Einstein radius systems and it even affects known lenses when they are seen in worse seeing. For example, the SLACS lenses (Bolton et al. 2006), which were discovered spectroscopically and thus are rather compact, were not identified as lenses by human experts when seen on Dark Energy Survey (DES) imaging (Rojas et al. 2023). Consequently, even if there are arcs or multiple images after simulating the lensing, that by itself is not enough to guarantee that the lensing features are visible and identifiable as such in the final image. This is especially problematic when simulating lensing by edge-on late-type galaxies because their light profiles tend to be extended and to have structure; thus, increasing the chances of blending between the deflector's light and the lensed source. To address this problem, we implement a "contrast check" detecting when the lensing features are invisible in the final image. If a mock lens fails the contrast check, we change the position of the source regarding the caustic lines and generate a new mock lens. If the new mock also fails the contrast check, we reject the mock and go to the next potential deflector.

Since both the image of the lensed source and the image of the deflector galaxy are sampled on the same pixel grid, the contrast check consists of comparing the equivalent pixels between the images and seeing which ones are "dominated by the lensed source light". To do so, we use the procedure illustrated in Fig. 2:

1. Estimate the background level μ and background dispersion σ of the deflector image. We use the median to estimate μ and the normalized median absolute deviation for σ.
2. Make masks for both images selecting pixels above $\mu + 3\sigma$. The second column of Fig. 2 shows example masks for the deflector and lensed source introduced in the first column.
3. Find and count the selected pixels in the lensed source mask that do not overlap with the deflector mask. They are immediately considered as dominated by the lensed source, and are shown as white in the third image of the first row of Fig. 2. If there are less than five such pixels, then the contrast check fails.
4. Find the pixels that do overlap between the two masks. They are shown in gray on the third image of the first row of Fig. 2.
5. Count how many of the overlapping pixels are dominated by the lensed source light. That is, the pixel value in the lensed source image is at least 3 times larger than in the deflector image. They are shown in red on the third image of the second row of Fig. 2.
6. The total number of pixels dominated by the lensed source light, which is counted in step 3 and 5, must be larger than 20; and they must amount to at least 30% of the pixels in the lensed source mask.