Uncertainties driven by neighbor contamination

Source contaminations (blendings) from neighbors is a fundamental situation in extragalactic survey, which can often occur not only in the highly crowded cluster environments, but also in general fields with deep observations, where the targets overlap with background faint objects along the line of sight. This problem affects all of the main processing steps in our pipeline, from source detection through to galaxy profile fitting, potentially leading to additional uncertainties. However, the literature lacks detailed analysis on this challenging aspect, with the exception that Häußler et al. (2007) and Barden et al. (2012) present, in a similar fashion, only rough evaluations for the influence of neighbor contamination on the galaxy fitting quality with GALFIT. Due to the exceptionally crowded nature in HFF, the effects of blending on Sérsic parameter uncertainties merit careful exploration. To address this issue, we have performed detailed simulations using single Sérsic model to comprehensively characterize random and systematic uncertainties introduced by neighboring source contaminations. Our method is naturally applicable to other less crowded fields, such as CANDELS.

As explicitly described previously, GALAPAGOS2 defined any close neighbor with potentially significant light influence on primary as secondary, which will be fit simultaneously with primary. Note that secondary has to be fainter than primary. In contrast, if a secondary is brighter than primary, GALAPAGOS2 will hold its parameters fixed to the best-fitting values for saving CPU time, because the fitting results for brighter source already exist. During the fitting process, these two types of secondaries are determined based on an automated basis, and each type may affect the primaries in a different manner, introducing scatters and biases in the recovered parameters. We therefore will look into both cases separately in the following sections.

1   fainter secondaries

In order to remove any light contaminations, fainter secondaries will be fitted simultaneously with the primary galaxy. Such contamination is a complex profile superpositions from all selected fainter secondaries at the position of current primary. For this case, the fitting accuracy of primary depends on the GalfitM deblending performance. Since the extent of source blending is typically a stronger function of distance to the primary, we expect that the nearest secondary will have the dominant contribution to the fitting results of primary, while the others only exert second-order effects. We only consider the impact of nearest secondary

in this work. In this sense, the fitting uncertainties of primary are primarily dependent on the intrinsic structural parameters of both primary $\vec{p_1} = (m_1, r_{e,1}, n_1, q_1, PA_1)$ and secondary $\vec{p_2} = (m_2, r_{e,2}, n_2, q_2, PA_2)$ and the spatial separation d between these two objects, which can be summarized as following formula:

$$(\vec{b_1}, \vec{\sigma_1}) = f(\vec{p_1}; \vec{p_2}; d) \tag{1}$$

where the two error vectors $\vec{b_1} = [b(m_1), b(r_{e,1}), b(n_1), b(q_1), b(PA_1)]$ and $\vec{\sigma_1} = [\sigma(m_1), \sigma(r_{e,1}), \sigma(n_1), \sigma(q_1), \sigma($ are the systematic biases and random/statistical uncertainties of the primary Sérsic parameters, respectively. As such, we fit simulations of primary with only one secondary, i.e., primary-secondary pair. We do not account for the case of including multiple secondaries in the simulations.

Similarly to what was done previously, we create the simulated primary-secondary pair again using simple single Sérsic functions. Figure 1 illustrates the method for creating the input frame of simulated primary-secondary pair (i.e., model image), for both fainter and brighter secondaries. For the pair with known input parameters (Sérsic parameters and separation), we start by generating a simulated *primary frame* with only Sérsic model of primary placed in the image center, as illustrated in the middle panels of Figure 1. To proceed, we first use GalfitM to create the an initial PSF-convolved primary model image (the first middle panel), with the same PSF image as used before. We subsequently add the Poisson noise to the primary model (the second middle panel). The resulting image is then combined/added with a real sky frame of the same size (shown as the gray background in Figure 1), to form the input *primary frame* (the last middle panel). The secondary model was produced following the same way as applied to primary, and then combined with *primary frame* to create the ultimate input frame. Since we intend to investigate the impact of secondary on the fitting parameters of primary as a function of distance, in order to obtain the frames of pairs with various separations, we repeat the above procedures by increasing the separation, in steps of $0.5r_{e,1}$, over a range of $d = 1 - 20\, r_{e,1}$, where the $r_{e,1}$ is the input effective radius of primary. This is achieved through simply modifying the position of secondary model (blue circle), and during the process we repeatedly use the same *primary frame* as the pedestal. This method has the advantage to ensure that the Poisson noise from primary model is kept 'fixed', as we vary the position of secondary. Some readers may argue that a more simplified and efficient way is to create both models simultaneously in a single frame, but this also implies that the Poisson noise has to be added to both models each time when changing the separation. We warn the reader that in the latter way the non-constant/unpredictable Poisson noise will slightly oscillate the primary parameters, which bias the subsequent model fitting and thus the interpretation of fitted parameters. Next, the input frame is provided to GALFTM to perform simultaneous fit.

Below, we describe how we select the input parameters for the models. We first produce a large mock sample with parameters following the same distribution as the real galaxy sample To select the input primary models, the mock sample is then divided into five magnitude bins, 19-21 mag, 21-23 mag, 23-25 mag, 25-27 mag, and 27-29 mag. In each bin, we randomly select five galaxies, but requiring that their Sérsic indexes span the range of values found in real data, which can ensure the coverage of different types of galaxy profiles. Our primary sample therefore consists of 25 models covering the full parameter range. For each primary, we further select another five models in the corresponding magnitude bin of the mock sample as

2

its secondaries, whose Sérsic indexes cover the whole range of primaries. As a consequence, our input sample for simulations comes out be consisted of 125 primary-secondary pairs, representing a complete combinations of various galaxy profile types. It is necessary to stress that the secondaries are chosen by design to be slightly fainter than the primary, in which we expect that the impact of secondaries on primaries is most significant.

For each simulated primary-secondary pair, we follow the method as illustrated in Figure 1 to create the input frame, for which we run simultaneous fit with Galfitm to derive primary parameters. The same procedure is repeated as we change the separation of the pair. Figure 2 shows the parameter deviations of primary as a function of spatial separation of two components, for a pair with typical parameter combination. From top to bottom panels, we show the differences between measured and input values for magnitude, effective radius, and Sérsic index. The upper legend in the top panel denotes the input parameters of primary with a Sérsic index of n=4, representative of spheroidal galaxy. The lower counterpart labels secondary parameters, which is an exponential disc profile (n=1). Each gray line corresponds to a simulation *cycle* performed at a random location of the frame. As mentioned above, one simulation will be conducted after changing the pair separation. The input positions of primary are identical in different simulations (namely using the same *primary frame*), while the secondary is allowed to be gradually shifted along the direction with azimuthal angle $\phi = 45$ degree (see Figure 1 and Figure 3). Hereafter, we refer to the simulation processes through the defined range of separations (from $1r_{e,1}$ to $20r_{e,1}$) as *cycle*. For each pair, we run 20 cycles by placing its primary at 20 random positions in the frame, which aims to approach how the local pixel-to-pixel variation translates into the measurement uncertainties of primary. Figure 25 shows the input frames of cycles for the representative pair exemplified in Figure 24. Each row in Figure 25 denotes a cycle, where we show the input frames with pair separations evenly sampled at an interval of $2r_{e,1}$, for illustrative purposes. We can see that these cycles have different primary positions (lower-left object of the pair in each panel) that are randomly chosen.

Therefore, the gray lines in Figure 2 are results derived from simulation cycles that are conducted in succession by randomly placing the primary. It is clear that those lines are noisy, even at larger separations where the local background variation dominates the fitting uncertainties. This suggests that local background variation is a non-negligible effect when characterizing the dependences of primary parameter uncertainties on secondary properties. To statistically explore this effect, we inject only the primary model at random into the empty sky frame. This *primary frame* is then measured using Galfitm with single Sérsic function. We repeat the above computation 50 times to produce a distribution of Δmag, Δlog(R_e), and Δlog(n). The associated 3σ deviations are displayed as yellow shaded regions in Figure 2, representative of an overall influence from background pixels. Note that we have kept the Poisson noise constant across the different runs so that the resulting fitting parameters are not deteriorated by its unpredictable perturbations, but purely driven by the background pixels. The red vertical line in Figure 2 delineates the separation limit to deblend the pair, which is calculated using SExtractor cold+hot mode with same configuration parameters as used in the real image onto the input frames from a single simulation cycle. In the left side of this limiting boundary, our source detection procedure fails to deblend the pair due to small separations. Moreover, the fitted Sérsic parameters are highly uncertain below this limit, namely more scattered gray lines, demonstrating severe profile overlapping. In contrast,

towards the right side, the separation of the pair is large enough, such that we are able to distinguish both models, and the gray lines become more stable and flatter.

The red points with error bars indicate the median and 1σ deviation of the difference (i.e., for all the gray lines) with equally spaced bins. This gives a quantitative measure of systematic and random uncertainties because of the presence of secondary. As expected, the random uncertainties increase when the separations become small. From Figure 24, for three key Sérsic parameters we see that the systematic uncertainties are close to zero, and therefore the error budget is dominated by the random uncertainties, but the latter are much smaller than that induced by sky pixel variation, even around limiting separation. Regarding the specific pair in Figure 24, it is obvious that the parameter uncertainties owing to secondary are negligible compared to that resulting from sky pixel variations. Following the same steps, we examine a total of 125 pairs (as described above) covering a wide range of parameter combinations via generating diagnostic figures with same format as Figure 24. We found that they display the same behavior and are in excellent agreement with those found in Figure 24. We conclude that fainter secondaries have no significant impact on the recovery of input primary parameters. Finally, it is important to note that our model selection for the pairs demonstrates a worst/extreme case scenario, because we have intentionally required the secondary models to be slightly fainter than the primaries, where we would expect the impact to be most significant (i.e., upper limits). However, this is not true in reality, as the nearest secondaries much fainter than primaries are more common in our science image, hence resulting in smaller uncertainties. In conclusion, we opt to ignore the uncertainties introduced by fainter secondaries.

2 brighter secondaries

In this section, we quantify the primary parameter uncertainties caused by the brighter secondaries. Following the assumption in previous section, we only focus on the nearest brighter secondary. Since parameters of the brighter secondaries are held fixed when fitting the primaries, the involved simulation procedures are slightly more complicated than the fainter secondaries. Starting from the input frame, we first fit the pair simultaneously to derive their best-fitting parameters. We then rerun the model fitting while keeping secondary parameters fixed, to measure primary parameters for the second time. The input frame used for the second run (top left panel of Figure 1) is constructed by adding the best-fitting model of secondary to *primary frame*. These steps are illustrated in the top panels of Figure 1. To assign meaningful uncertainties to the real objects, we create a new set of mock pairs with parameter combinations as complete as possible to the real galaxy sample. For a given pair, the systematic and random uncertainties of primary, as well as the corresponding error plot with the same style as Figure 24, are obtained following the same procedures as described before. Each error plot is essentially equivalent to an error template, with uncertainties as a function of separation. Such collection of templates will serve as a database which we use to assign the uncertainties to the real objects, as explained below.

In previous section, the selection of input pair parameters is carried out by specially picking up individual object from mock sample, which is generated based on the real distribution. However, in this section the mock pairs are selected in a different way, to ensure that the
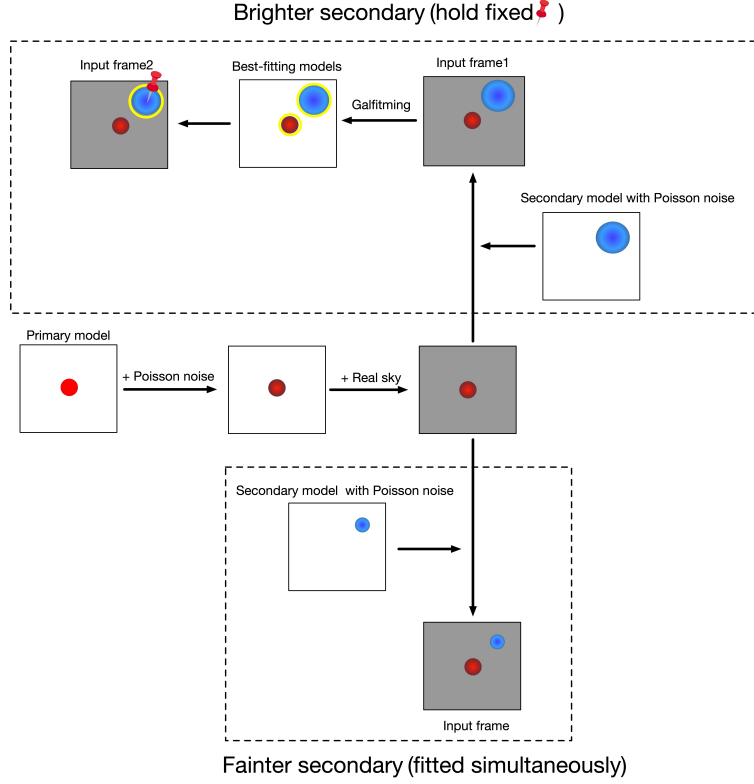
Figure 1: Cartoons that illustrate the strategies for generating the input simulated frame containing primary-secondary model pair. The middle panels describe the processes for generating *primary frame*, i.e., the image cutout with gray background (real sky) and centered on the dark red circle (primary model with added poisson noise). Then the procedures diverge from *primary frame*: the lower flow shows how the fainter secondary is added, while the upper flow is for brighter secondary. In both cases, the associated secondaries (blue circles) are generated in separate frames, and then added to *primary frame*. The goal is to keep the Poisson noise of primary constant. The top panels show the fitting strategy for the case of brighter secondary. Starting with input frame, we first perform a simultaneous fit using two Sérsic components to recover the initial best-fitting models, which are illustrated by encircling the models with yellow rings. Next, in order to derive the primary parameters, we redo the model fitting by forcing the secondary parameters fixed to its prior best-fitting values. The associated input frame, i.e., Input frame2 (top left panel), is constructed by adding the best-fitting model of secondary into the *primary frame*.
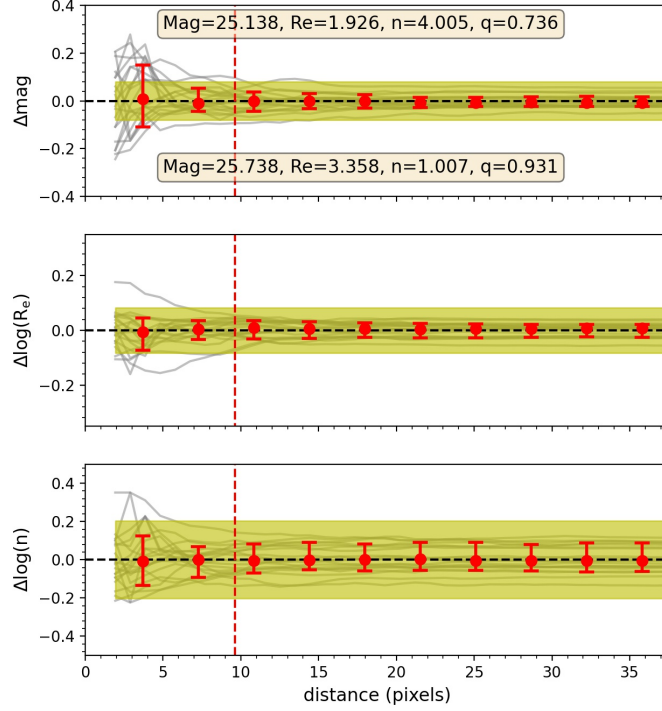
Figure 2: Parameter differences (output − input) of primary for our simulated primary-secondary pair as a function of the distance to the nearest secondary. The upper and lower legends in the top panel display input primary and secondary parameters, respectively. The gray lines represent the simulation *cycles* with randomly chosen primary positions, in order to investigate the potential effects of background pixel-to-pixel variation on the fitting results. The yellow shaded regions are representative of the 3σ deviations, which are derived from the *primary frames* (in the absence of secondaries) with randomly placed locations for primaries, during which we have fixed the Poisson noise to be constant in different runs, such that these scatters purely result from background pixel variations (see text for details). The red vertical line marks the limiting separation where we can barely distinguish between the primary and secondary, using SExtractor cold+hot mode with the same configuration parameters. In its left side (i.e., lower than the limit), both models are indistinguishable owing to close overlapping, while in the right side they can be separated. Red points with error bars indicate the running median and 1σ deviation (16/84 percentiles) for gray lines in evenly spaced bins. This presents a quantitative measure of systematic and random uncertainties due to secondary. The systematic errors are close to zero, while the random uncertainties get larger as the separation decreases, as anticipated. The horizontal dashed line marks a difference of zero. This figure highlights that the uncertainty in background pixel variation greatly exceeds the statistical uncertainties associated with secondaries.
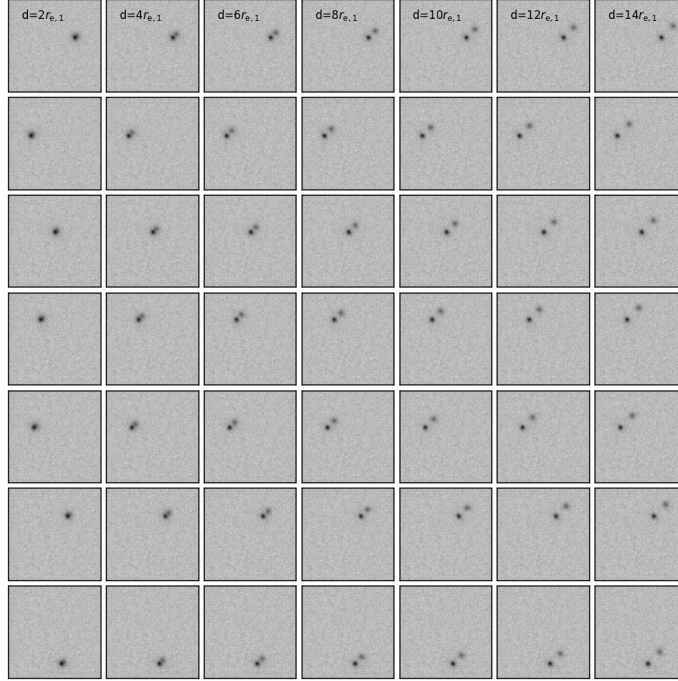
Figure 3: Input frames for a model pair with typical galaxy parameters, see the top panel of Figure 2. The rows indicate simulation cycles (corresponding to gray lines in Figure 2), whose primary positions are randomly determined within the frame (lower-left object of the pair). The panels in each row are selected with an interval of $2r_{e,1}$ in separation, and are only shown for illustration. Each column of panels show pairs with same separation but random primary locations, while the row of panels represent identical/constant primary locations with gradually increased separations.

input parameters are uniformly distributed throughout entire parameter space of the real sample. We start by describing the determination of the input magnitude. From the real sample, we select all galaxies with brighter secondaries that are defined by GALAPAGOS2 during the fitting process. We emphasize that a primary does not necessarily have brighter secondaries. It depends on the image cutout size and spatial distribution of the neighbors. Table 1 lists the best-fitting Sérsic parameters for objects with brighter secondaries, as well as those of the latter. The total number of these objects is 1571, accounting for $\sim 57\%$ of the total sample. Figure 5 demonstrates the magnitude differences between primaries and its nearest bright secondaries (secondary minus primary) for the real sources as a function of primary magnitude (shown in red circles). Note that the values of the differences (y-axis) are negative, because the secondaries are defined to be brighter. We also aware that the distribution is constrained within a well-defined triangular region, indicating that brighter primaries tend to have fewer brighter secondaries. This is to be expected, as the brighter objects are commonly less abundant than the fainter ones. We utilize a simple linear relation (yellow dashed line) to roughly depict its boundary. We then overlay an equally spaced mesh grid (blue dashed lines) across the plane in Figure 5, with a mesh size of 2 mag $\times$ 2 mag. The grid points (black stars) in the triangular zone that is populated with real data, are used to assign the primary and secondary magnitudes to the pair. Given that this type of secondaries are required to be brighter than primaries, we systematically shift the bottom row of grid points towards negative (vertical) direction, such that these secondaries are 0.5 mag brighter than the corresponding primaries. There is a total of 14 stars/pairs in the figure, for which we know the input magnitudes.

The effective radius r_e is determined from an empirical magnitude-size relation. To do so, we take advantage of the existing mock sample obtained in previous section (i.e., blue stars in Figure 6), by fitting a 3-order polynomial to its running medians. The derived formula is shown in Figure 6. Given the input magnitudes, we use this formula to empirically calculate the associated effective radii. To assign the Sérsic index, we choose four representative values of $n = [0.5, 1, 2.5, 4]$, where n=2.5 is a reasonable midpoint between late-type and early-type galaxies. This leads to 16 combinations of Sérsic index for our pairs. Given that the measured uncertainties are not correlated with axis ratio q (van der Wel et al., 2012), we only adopt two different values of $q = [0.25, 0.75]$ (4 combinations) to reduce the computation time. The position angle is chosen randomly. Consequently, we have a total of $14 \times 16 \times 4 = 896$ parameter combinations, namely input pairs of 896, each of which provides an error template generated using aforementioned simulation procedures. To properly assign uncertainties to our real sample, we assume that galaxies/pairs with similar structural properties share similar uncertainties. For each real pair, we identify the closest input pair in the four-dimensional parameter space (m, r_e, n, q), and then the associated error template will be used to assign the parameter uncertainties for the target primary. Table 2 exemplifies an error template for a pair with input parameters labelled in Figure 4 (wheat-color legends), as well as that of the matched/closest real pair (cyan-color legends), while Figure 4 presents the equivalent error plot following the same layout as Figure 2. The assigned uncertainties for our real sources are listed in Table 3, where only those objects associated with brighter secondaries are considered. These uncertainties are purely driven by brighter secondaries.

Table 1: The best-fitting Sérsic parameters of primary-secondary pairs in our sample

id1	mag1 (mag)	re1 (pixels)	n1	q1	pa1 (degree	id2	mag2 (mag)	re2 (pixels)	n2	q2 (degree)	pa2	d (pixe
2	26.043	3.219	0.497	0.069	-89.644	3	25.586	0.302	3.433	0.05	-84.272	26.19
4	26.519	3.392	0.522	0.391	84.189	3	25.586	0.302	3.433	0.05	-84.272	22.5
8	25.292	7.298	0.37	0.25	1.99	10	23.557	0.667	3.688	0.772	-84.636	32.9
10	23.557	0.667	3.688	0.772	-84.636	1	20.464	15.032	0.348	0.212	21.288	38.1
13	23.821	3.918	7.945	0.928	-76.046	1	20.464	15.032	0.348	0.212	21.288	25.5
16	25.981	3.144	0.399	0.657	-40.586	13	23.821	3.918	7.945	0.928	-76.046	18.5
17	25.687	1.871	2.979	0.583	87.918	20	25.142	4.479	0.826	0.737	-57.812	27.9
19	26.29	0.543	1.836	0.525	60.016	23	26.053	2.223	0.963	0.821	47.053	18.7
24	26.084	4.336	0.781	0.601	-61.165	21	24.797	7.634	0.888	0.241	-74.982	29.5
25	26.536	1.408	0.974	0.512	-49.698	20	25.142	4.479	0.826	0.737	-57.812	17.19
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Notes. This table lists the derived Sérsic parameters for both primaries and the associated nearest brig[ht] secondaries. Column 1: ID number of the primary in the catalog. Column 2-6: primary Sérsic paramet[ers] total magnitude, half-light radius, Sérsic index, axis ratio, and position angle. Column 7: ID number of matched secondary. Column 8-12: secondary Sérsic parameters. Column 13: distance between the prim[ary] and secondary, in units of pixels.

Table 2: An example of error template for one input pair

d (pixels)	mag_sys (mag)	mag_err (mag)	re_sys (pixels)	re_err (pixels)	n_sys	n_err	q_sys	q_err	pa_sys (degree)	pa_err (degree)
5.39	-0.0488	0.1199	0.1626	0.422	-0.0042	0.3111	-0.0149	0.068	0.5754	3.2633
10.57	-0.0222	0.0509	0.0508	0.2505	0.0104	0.2637	-0.0073	0.0501	0.576	2.521
15.75	-0.0238	0.0396	0.0863	0.189	0.0672	0.2201	-0.0155	0.0474	-0.0292	2.4946
20.93	-0.0213	0.0338	0.1053	0.1541	0.0963	0.2013	-0.0138	0.0455	0.0811	2.4801
26.11	-0.0251	0.0329	0.1091	0.1688	0.1143	0.2334	-0.0135	0.0467	0.2069	2.5212
31.29	-0.0251	0.0313	0.1098	0.1689	0.1124	0.2294	-0.0143	0.0495	0.1573	2.4356
36.47	-0.0227	0.0287	0.1059	0.1624	0.109	0.2138	-0.0152	0.0481	0.1082	2.4268
41.65	-0.0251	0.0285	0.1112	0.1657	0.1079	0.2188	-0.0168	0.0477	0.1061	2.4394
46.83	-0.0234	0.0292	0.1077	0.1615	0.1148	0.2201	-0.0154	0.0459	0.1504	2.3865
52.01	-0.0236	0.0281	0.1123	0.163	0.1189	0.2149	-0.0134	0.047	0.1523	2.4259

Notes. This table lists parameter uncertainties as a function of separation (Column 1), which is agai[n] in units of of primary r_e. It includes systematic and random uncertainties of single Sérsic parameters namely magnitude (column 2, 3), effective radius (column 4, 5), Sérsic index (column 6, 7), axis rati[o] (column 8, 9), and position angle (column 10, 11). The input parameters of the pair are labelled i[n] Figure 1.
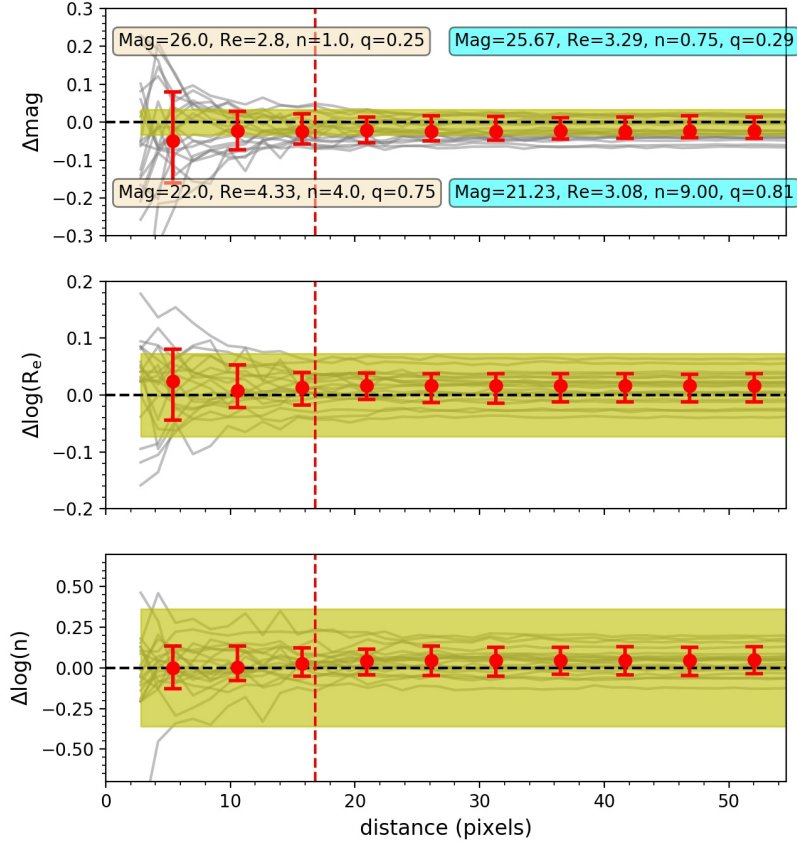
Figure 4: Identical to Figure 2 but for a pair with brighter secondary. The wheat-color legends in the top panel mark the input parameters, and the cyan-color legends in the right show the best-fitting parameters for the closest real pair (top: primary, bottom: secondary), for which the uncertainties are assigned. Table 2 gives the equivalent data for the same input pair.
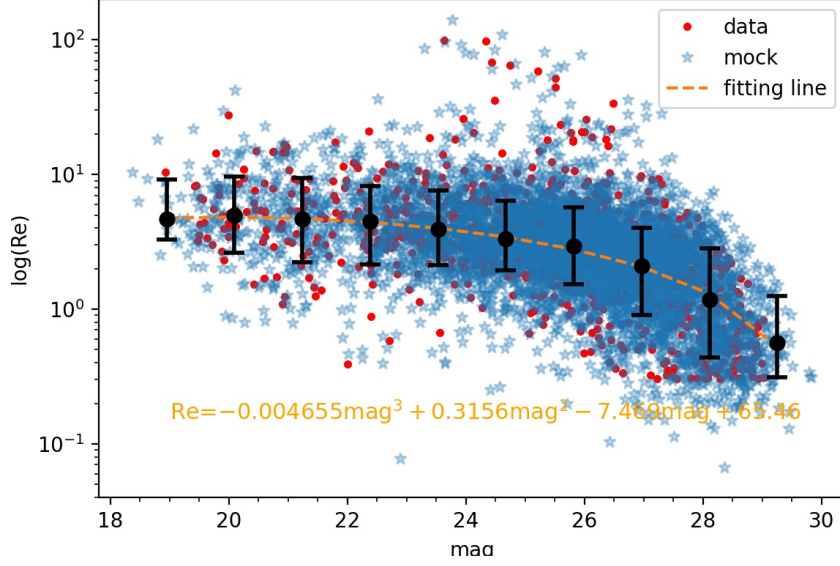
Figure 5: Magnitude-size relations for our real and mock sample. The red points present the real sample, and the blue stars show the mock sample with parameters following the real distributions. The black points represent the running medians for the mock sample, and errorbars indicate 16 and 84 percentiles. The yellow dashed curve is the best-fitting relation for the black points, using 3-order polynomial. For given magnitude, we use this empirical relation to calculate the input effect radius.

Table 3: Systematic and random uncertainties of the Sérsic parameters for the real sources, driven by the brighter secondaries.

id	mag_sys (mag)	mag_err (mag)	re_sys (pixels)	re_err (pixels)	n_sys	n_err	q_sys	q_err	pa_sys (degree)	pa_err (degree)
2	-0.0118	0.0289	0.0471	0.1075	-0.0197	0.1801	0.0089	0.0664	1.0115	2.0917
4	-0.0069	0.0296	0.0425	0.1165	-0.0219	0.1759	0.0086	0.0643	0.9948	2.1174
8	-0.0084	0.0212	0.0026	0.1052	-0.0302	0.147	0.0106	0.0412	-0.2897	3.0147
10	-0.0013	0.0114	0.0058	0.0972	0.0301	0.1663	0.0078	0.0216	0.5729	2.5303
13	-0.001	0.0144	0.0043	0.1227	0.0379	0.1663	0.0067	0.0224	0.4925	2.6171
16	-0.023	0.0238	0.0799	0.107	0.0067	0.1534	-0.0462	0.0847	1.1285	8.7588
17	-0.0095	0.0298	0.1925	0.2762	0.0336	0.5228	-0.0335	0.0708	-74.797	83.3482
19	-0.0064	0.0308	0.1691	0.2453	-0.0271	0.5652	-0.0362	0.0701	-74.8806	83.9215
24	-0.0103	0.0352	0.0461	0.2213	-0.009	0.1185	-0.0428	0.0999	-1.3175	9.4701
25	-0.0055	0.0406	0.0122	0.1638	-0.0595	0.228	-0.0105	0.0749	0.7256	10.4215
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Notes. The first column gives the ID of real source with brighter secondaries defined by our pipeline. The rest columns follow the same format as Table 2.

References

Barden, M., Häußler, B., Peng, C. Y., McIntosh, D. H., & Guo, Y. 2012, MNRAS, 422, 449

Häußler B. et al., 2007, ApJS, 172, 615

van der Wel, A., Bell, E. F., Häußler B., et al. 2012, ApJS, 203, 24