THE UNIVERSITY OF CHICAGO

MEASUREMENTS OF MORPHOLOGY IN STRONGLY LENSED GALAXIES IN THE IMAGE PLANE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ASTRONOMY AND ASTROPHYSICS

BY
MICHAEL KENNETH FLORIAN

CHICAGO, ILLINOIS
DECEMBER 2017

For my father, Kenneth Lee Florian, who introduced me to astronomy.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I have been fortunate that, from an early age, I have had supportive parents, teachers, and friends that have encouraged me and made it possible for me to get to this point. As I reach the end of my formal education and reflect on the nearly three decades that have led me to this point, I would like to take a moment to thank the people in my life who have had the most influence on my education my decision to take on graduate studies in astronomy.

My parents are, of course, the first on this list. They understood the importance of a good education and stressed that importance until I came to understand and accept it as well. My dad, Kenneth Florian, and his love of space (both astronomy and science fiction) helped point me toward astronomy by the time I was in first grade. While the help and support I have gotten from my mom, Jane Florian, are not as directly related to astronomy as that which I got from my dad, the degree to which she's helped me pursue my education and grow as a person cannot be understated. And of course, my sister, my extended family, and our close family friends have all provided tremendous support for me in one way or another countless times.

I suppose next in line, chronologically at least, are the teachers that I have had over my lifetime. Primary education is extraordinarily important. I cannot imagine how much harder it would have been to even go to college, let alone graduate school, had it not been for having good teachers early on. Mrs. Sider, my first grade teacher at Portage Park Elementary School in Chicago, was one of the best teachers I have had, and I was extremely lucky to have gotten such a good teacher so early. She always encouraged me and gave me the resources and the push that I needed to always improve (and just as importantly, to want to improve) in all subjects, and to never rest on what I already knew. Another teacher, Mr. Matsumoto, who I had several times for math and science classes between fifth and eighth grade, was also instrumental in keeping me on the path toward higher education, keeping me ever curious and future-looking. His patience, dedication to teaching, and encouragement of his students are unsurpassed.

Because I listened to teachers like Mrs. Sider and Mr. Matsumoto, and because of a good dose of luck as well (we all know there's a component of randomness to application processes), I was able to get into Northside College Preparatory High School, where most of my teachers were excellent. But I'd like to mention a few of them by name, briefly. Mr. Koreman, my freshman year physics teacher, helped re-focus me on astronomy and on my studies. Mr. Dahlberg and Mr. Harada, my senior year A.P. calculus and A.P. physics teachers respectively, went above and beyond their responsibilities as teachers, staying after school and being available during their lunch and prep periods to help students like me—both when we had difficulties with class material, and when we wanted to learn about topics outside of the curriculum that would have otherwise had to wait for college. And I would also like to thank Ms. Montgomery (now Mrs. Driver) for all of the work she did with our "advisory" to help us with our college searches and our preparation for the college application process and also for holding me to the claim I made during freshman orientation that I wanted to be an astrophysicist.

I would like to thank Prof. Don York at the University of Chicago for helping me get experience with research in astronomy when I was a second year undergraduate and for his help and guidance as I continued my undergraduate work and transitioned into the graduate program. On a similar note, I would like to thank Prof. Jim Truran for helping me make the transition to graduate school and being an important scientific mentor during that transition as well as in the years that followed. Of course, a special thanks goes to my advisor Prof. Mike Gladders, for all of the work that goes in to helping a student like me develop into a scientist. And thanks, also, to the other members of my committee, Prof. Brad Benson, Prof. Nick Gnedin, and Prof. Rich Kron, for their helpful questions, comments, and advice as I conducted the work presented in this thesis.

While I am acknowledging those whose help directly contributed to this thesis, I should mention the agencies whose funding allowed me to carry out the work in this dissertation. Partial support for this work was provided by the Hubble Space Telescope program number

GO-13003, which was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Furthermore, this material is based upon work supported in part by the National Science Foundation under Grant No. 1616551. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

And last, but certaintly not least, I would like to thank Kathleen Beilsmith, my girfliend of the last eight years, who has, among other things, taught me to put important statements in places of emphasis, like the last paragraph of the section, or the first sentence of the paragraph. She has requested that I not right anything sappy, so I won't, but I do want to thank her for her help organizing and outlining this dissertation, brainstorming approaches and solutions to some of the problems presented here, and for generally making the last several years much more pleasant than they otherwise could have been. Graduate school has its low points, but she has made them more bearable. And the high points have been all the more enjoyable because she was there with me.

ABSTRACT

The peak of star formation in the universe, the so-called "cosmic noon," occurs around redshift 2. Therefore, to study the physical mechanisms driving galaxy assembly and star formation, and thus the bulk morphological appearances of present day galaxies, we must look to galaxies at this redshift and greater. Unfortunately, even with current space-based telescopes, the internal structures of these galaxies cannot be resolved. The point spread function of the Hubble Space Telescope (HST), for example, corresponds to scales of about 0.5 kpc at redshift 2. Even the next generation of telescopes (e.g., the James Webb Space Telescope, the Wide-Field Infrared Survey Telescope, and the new thirty meter class of ground-based telescopes) will not be able to access the spatial scales—tens of parsecs or less—on which star formation has been shown to occur in the local universe. Fortunately, strong gravitational lensing can magnify these spatial scales to angular scales comparable to, or larger than, the HST point spread function. However, this increased access to small scales comes at the cost of strong distortions of the underlying image. To deal with this, I use simulations to show that some morphological measurements (e.g., the Gini coefficient) are preserved by gravitational lensing and can be measured in the image plane. I further show how such measurements can aid image family identification and thus improve lens models and source reconstructions. I explore a method to measure the fraction of a lensed galaxy's light that is contained in star-forming clumps in the image plane, which would bypass the need for lens modeling and source reconstruction to carry out similar measurements. I present a proof of concept for a simple case, and show where the major uncertainties lie—uncertainties that will need to be dealt with in order to expand this technique for use on more image configurations and tighten the relationship between the intrinsic values and the measured values. I suggest several ways in which these uncertainties can be overcome. Finally, I discuss the potential for future application of these techniques, particularly in the context of explaining star formation processes across cosmic time, and the associated implications for galaxy mass assembly mechanisms and galaxy evolution.

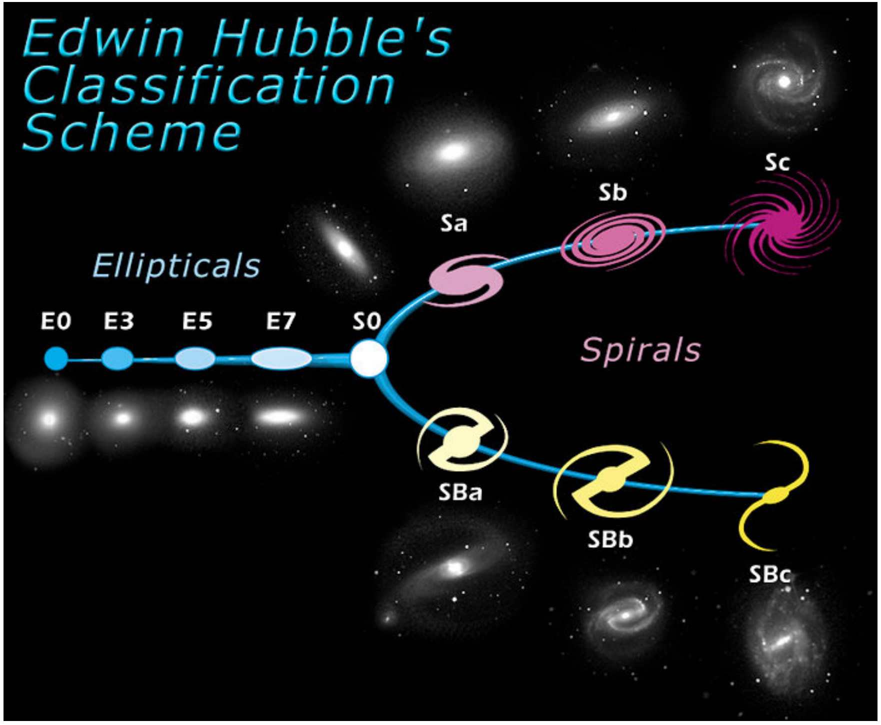
CHAPTER 1

INTRODUCTION

When Edwin Hubble published his famous 1926 paper on the morphologies of galaxies outside of the Milky Way (Hubble 1926), the paper that motivated the famous "tuning fork" diagram (Fig. 1.1), the astronomical community still called such galaxies "extra-galactic nebulae." We have come a long way in our understanding of galaxies since then (and have, of course, reserved the word nebulae for an entirely different class of astronomical objects), but explaining the stark differences between the various components of the tuning fork—unbarred and barred spiral galaxies, and elliptical galaxies—and the irregular galaxies that do not fall neatly into any category remains a major focus of both theoretical and observational work in extragalactic astronomy (e.g., Kereš et al. 2005, Dekel & Birnboim 2006, Porter et al. 2014, etc.). Our understanding of the relationship between these two major classes of galaxies has evolved substantially over the past 90 years. Though we have, perhaps unfortunately, inherited the "late-type" and "early-type" nomenclature for spiral and elliptical galaxies respectively, astronomers no longer believe that these types represent an evolutionary sequence. But how *are* they related? Or are they related at all? Why are these two types so prevalent in the local universe even though, as we know, other galaxy morphologies do exist (e.g., irregular galaxies, ultra compact disks, etc.)? What are the physical processes that drive these differences? How can we explain the vast diversity of morphologies within these classes themselves, like the existence of bars or bulges in some spirals? How do these morphologies, and their relative frequencies, change over time? While some progress has been made toward answering these fundamental questions (for a detailed review, see, for example, Somerville & Davé 2015 or Baugh 2006), they continue to be major drivers of research in this field today.

A key question among these is: Which physical processes have most strongly driven the mass assembly histories of galaxies at different points in cosmic time? For example, were they predominantly built through minor mergers or major mergers? Or are today's galaxies

Figure 1.1: The Hubble Tuning Fork Diagram



The famous Hubble tuning fork diagram, a classification scheme that encompasses the most common galactic morphologies in the local universe. (Image credit: NASA & ESA)

mainly the result of a slow accretion process of hot or cold gas over time? And do different processes dominate at different times or for different types of galaxies? To begin to answer these questions we first need a set of morphological metrics that can both be observed in real galaxies and be predicted in simulations, and that are sensitive to these various mass assembly scenarios. A great deal of work has been done to predict, via simulation, the effects of processes like minor mergers, major mergers, cold mode accretion, and hot mode accretion on galaxy morphology. The following is a brief summary of the qualitative effects of these processes as predicted by simulations.

1. **Mergers:** Mergers have the ability to disrupt existing angular momentum distributions, resulting in more spheroidal structures, especially in gas-poor (or "dry") mergers where the subsequent dissipation of energy is least efficient. (Naab et al. 2006, van Dokkum 2005)

    (a) Minor mergers are typically defined to be mergers between two galaxies whose mass ratio is smaller than 1:10 and tend to build thick disks and bulges, with the effect being stronger in dry mergers. (Walker et al. 1996, Moster et al. 2010)
    (b) Major mergers are those involving two galaxies of similar masses. These tend to more thoroughly disrupt existing structures than minor mergers do. Such events can redistribute angular momentum, thereby destroying disks and creating more spheroidal morphologies (Toomre 1977).

2. **Gas accretion:** The accretion of gas, and the dissipation of the angular momentum of that gas due to interactions with gas in the interstellar medium of a galaxy builds galactic disks. The thickness of these disks and the star formation patterns within them are influenced by whether the infalling gas reaches the virial temperature (hot mode accretion) or not (cold mode accretion) according to Dekel & Birnboim 2006.
    (a) Hot mode accretion is the dominant mode of gas accretion for high mass galaxies (Kereš et al. 2005). If this mode dominates in the early universe, then it means that galaxies were assembled rapidly early on, but that star formation within these more massive galaxies would have to be relatively slow because the long cooling times for such hot gas delay the collapse of that gas into stars. This would result in gas rich disks at early times—disks that are not disturbed as easily by major mergers or thickened by minor ones. The fraction of spheroidal galaxies and galaxies with thick disks or bulges would therefore be expected to be smaller at low redshifts than in a scenario where galaxies form predominantly by cold mode accretion in the early universe.
    (b) Cold mode accretion is the dominant mode of gas accretion for low mass galaxies (Kereš et al. 2005). Because the accreting gas never reaches a high temperature, it can collapse into a disk and form stars more quickly than in hot mode accretion (Brooks et al. 2009). If this mode of accretion occurs in the early universe, which

> would be reasonable because it is predicted to dominate at low masses, then we would expect to see a large fraction of galaxies with disk-like morphologies and rapid star formation at early times. Furthermore, rapid star formation means rapid depletion of gas, and therefore disks that are easily destroyed by major mergers and thickened by minor ones. We should therefore expect to see more thick disks, bulgy galaxies, and spheroidals at lower redshifts, when mergers have had time to occur, and the change in the relative numbers between these galaxies and galaxies with thin disks over time provides insight into the time periods during which this method of accretion dominates and those during which mergers of various sizes become significant contributors to galaxy evolution.

In addition to mass assembly processes like these, which deal with interactions between a galaxy and its environment, there are other galaxy-level or smaller effects that can contribute to the creation of structures in galaxies. For example, feedback from star formation, supernovae, or active galactic nuclei can slow star formation rates even in cold accretion scenarios where the star formation rates are otherwise expected to be quite high (Brooks et al. 2009). Mergers may be able to "quench" star formation, possibly contributing to the correlation between bulgy/spheroidal morphologies and older stellar populations. Furthermore, violent disk instabilities, possibly triggered by cold gas accreting onto a disk may cause giant star-forming clumps of masses up to $10^9 M_\odot$ to form as explained theoretically by Noguchi 1999 and observed by Genzel et al. 2011. Through processes like dynamical friction, these clumps may migrate inward, possibly being disrupted and dissipating into the disk or otherwise contributing to the formation of a bulge. Such an instability should be observable both through the existence of giant clumps at high redshift and an increase in the fraction of galaxies with bulges or thick disks at lower redshifts.

Because mergers of both varieties lead to more spheroidal morphologies, and because hot and cold mode accretion lead to more disk-like morphologies, a seemingly simple way to distinguish between these mass assembly histories for a given galaxy is to attempt to measure how "bulgy" or "disky" it is. The bulge-to-disk ratio (B/D), the ratio of a galaxy's light attributable to a bulge or a spheroidal component to the light attributable to a disk component turned out to be possible to predict in simulations (e.g., Hatton et al. 2003), making it appear to be an ideal observable to use to begin to probe the mass assembly histories of galaxies in the local universe where galaxy light profiles can be well measured. Several methods for doing this have been developed and employed (Peng et al. 2002, Simard et al. 2002) to varying degrees of success. However, due to projection effects and the ease with which structures like bars can be confused with bulges or thick disks, B/D fell out of favor among observers (Lotz, Primack, & Madau 2004, for example, describe several ways in which B/D is difficult to measure robustly in real data).

More robust measurements of galaxy morphology have since been developed by observational astronomers to help increase the suite of metrics that may be used in the pursuit of a better understanding of galaxy formation histories. These include the concentration, C, (Bershady et al. 2000), asymmetry, A, (Conselice et al. 2000), clumpiness parameter, S (Conselice 2003), the Gini coefficient, G, (Abraham et al. 2003, Lotz et al. 2004), the second moment of the brightest 20% of a galaxy's light, M_{20} (Lotz, et al. 2004), the internal color dispersion, ICD (Papovich et al. 2003), and the multi-mode, M, intensity, I, and deviation, D, statistics (Freeman et al. 2013). While many of these are necessarily sensitive to baryonic effects and are therefore still difficult to predict in simulations, they have already begun to be used observationally to classify galaxy types and provide insight into their mass assembly histories. For example, the Gini and M_{20} (and, at low redshifts, Gini and S) are known to be sensitive to recent mergers thanks to the work of Lotz et al. 2004.

These authors also showed that many of these metrics, with the Gini and M_{20} being notable exceptions, are not robust at high redshifts due to the limited spatial resolution currently achievable even with the Hubble Space Telescope (HST). However, because understanding galaxy formation necessarily requires understanding the changes in galaxy morphology *as a function of time*, any metric or statistic that we wish to use as a diagnostic of mass assembly processes must be applicable at high redshift.

But if even HST cannot provide imaging data with high enough spatial resolutions for these statistics to be reliable at high redshifts, what are we to do? Even the James Webb Space Telescope (JWST), which is optimized for infrared observations and is therefore well-suited for studying high redshift galaxies, will only provide resolution that is about the same as that of HST. Fortunately, there is one natural way to get around these resolution limits. Strong gravitational lensing can magnify high redshift sources by factors of tens to hundreds. So even though HST and JWST are only capable of resolving physical scales larger than about 500 parsecs at redshifts 1-3, by taking advantage of gravitational lensing it is possible to access scales at least as small as tens of parsecs. In fact, this effect has been exploited successfully many times already using both broadband imaging data and spectroscopic data to measure features on the scale of 10-100pc in galaxies at these redshifts (e.g., Rigby et al. 2017, Swinbank et al. 2009, Jones et al. 2010, Stark et al. 2008).

Now that the sample of observed strong-lensing systems has grown into the hundreds, it is finally possible to make measurements of small-scale structure in high redshift galaxies in a statistically meaningful number of sources. Or rather, it would be, if the magnification gains of strong gravitational lensing did not come at the cost of severe spatial distortions due to shearing and differential magnification of the images. It is often possible to work around these limitations by modeling the lensing masses (e.g., Hoag et al. 2016, Jauzac et al. 2014, Richard et al. 2014, Ishigaki et al. 2015, Johnson et al. 2014, Grillo et al. 2015) and reconstructing what the lensed sources looked like, unperturbed, in the source plane (e.g., Sharon et al. 2012), but it requires a significant investment of researcher time and computing resources in addition to often requiring several rounds of follow-up observations to obtain redshifts of cluster members and faint arcs that provide necessary constraints for constructing lens models. And, in some cases, there are not enough constraints to make lens models that are precise enough to be used to reconstruct the unperturbed sources. It is therefore infeasible, currently, to do source reconstructions for a statistically large sample of high redshift lensed galaxies like the kind that would be needed in order to draw generalized conclusions about the nature of galaxy formation and evolution.

If morphological metrics exist that can be measured in the image plane, thereby circumventing the bottleneck of lens modeling and source reconstruction, they will be in high demand as the number of strongly lensed high redshift sources continues to grow, affording us the opportunity to investigate theories about different mass assembly scenarios over cosmic time. In chapter 2 of this thesis, I show that the Gini coefficient is one such metric. While the Gini coefficient is also fairly robustly measured in high redshift unlensed galaxies, the fact that the lensed and unlensed samples represent differently selected—and therefore potentially different—populations means it will be useful to be able to tie the lensed sample to the unlensed sample. Additionally, the lensing may provide access to fainter objects than are contained in the unlensed sample, and measuring the Gini coefficient in these would allow astronomers to collect morphological information about this potentially different class of objects.

The Gini coefficient is of particular interest to the question of galaxy mass assembly because of its sensitivity to recent mergers. But in addition to the direct benefits, it helps facilitate other measurements by providing additional constraints to lens models, thereby enabling better source reconstruction so that morphological statistics that depend on having access to small physical scales (i.e., having many sightlines into the galaxy) can finally be measured in high redshift objects. In chapter 3, I will show that the Gini coefficient, when paired with galaxy colors, can improve the accuracy of image family identifications in complex lensing systems with several multiply imaged sources while also reducing the amount of follow-up observations, telescope time, and researcher time required to make these identifications.

To further take advantage of the small spatial scales made accessible by gravitational lensing, I will present a first attempt to quantify the fraction of light contained in star-forming clumps in high redshift lensed galaxies in the image plane in chapter 4 and refinements to the method in chapter 5. Due to the conservation of surface brightness in gravitational lensing, it may be possible to measure properties like this one in the image plane. I present a method of identifying clumps and modeling the clump and diffuse components in a simulated sample of both clumpy and smooth galaxies spanning a range of morphologies similar to those that have been previously observed. The results are promising, but not yet adequate to make a high quality measurement of these properties in the image plane. I will discuss the dominant sources of uncertainty in this method and potential steps that can be taken to improve it. Chapter 6 summarizes the work presented in this thesis in the context of the existing literature and discusses possible future applications of the techniques that I have developed.

CHAPTER 2

THE GINI COEFFICIENT

The highly magnified sample of high redshift galaxies made available by strong gravitational lensing provides access to both smaller spatial scales than would typically be possible to probe otherwise even with the Hubble Space Telescope (or with the soon-to-be-launched James Webb Space Telescope). It also represents a sample selected differently than the unlensed sample, which is essentially selected by apparent magnitude only. If we wish to understand the morphologies of these types of objects, and through that their formation and evolutionary histories, then we need to take advantage of the strongly lensed sample. As the sample of gravitationally lensed sources continues to grow—it is expected to reach the thousands in the era of the Wide-Field Infrared Survey Telescope (WFIRST)—any morphological information that can be extracted from the image plane will be of particular value because it allows the tedious, resource- and researcher-intensive, and uncertain process of lens modeling and source reconstruction to be bypassed entirely for certain applications. In this chapter, I show through simulations that the Gini coefficient is one of these morphological metrics that can be measured, meaningfully, in the image plane—a metric, in fact, that is also applicable in unlensed high redshift sources with HST-like resolutions, making it potentially very interesting as a bridge between the lensed and unlensed samples. Much of this chapter is reproduced, with only minor edits, from my paper, Florian et al. 2016b.

2.1 Definition of the Gini coefficient

The Gini coefficient, as applied to galaxy morphologies, is a measurement of the inequality of the distribution of light in a galaxy. It was originally developed in the field of economics as a way of quantifying wealth and income inequality, but was introduced to astronomy by Abraham et al. 2003. Conceptually, this measurement is made by ordering the pixels that make up the image of a galaxy in ascending order by flux and then comparing the resulting cumulative distribution function to what would be expected from a perfectly even flux distribution. A low Gini coefficient (close to zero) means that the distribution of light is fairly uniform, while a high Gini coefficient (close to one) means that most of the galaxy's light is contained in only a small fraction of the pixels. In practice, the Gini coefficient is calculated using the following formula from Lotz et al. 2004:

$$G = \frac{1}{|\overline{X}|n(n-1)} \sum_{i=1}^{n} (2i - n - 1)|X_i| \tag{2.1}$$

where X_i is the value of the i^{th} pixel when ordered by flux, $\overline{X}$ is the mean of the pixel values, and n is the total number of pixels.

2.2 Testing the stability of the Gini coefficient under strong gravitational lensing using simulations

Gravitational lensing preserves surface brightness. This means that, compared to an unlensed image of a source galaxy, any additional area on the sky (or additional pixels on a detector) occupied by a magnified and sheered lensed image must come with an equal amount of additional flux. This means that the cumulative distribution of the flux in the pixels that constitute the detected image should be unchanged and therefore the Gini coefficient would also be unchanged. There should be minor effects due to the convolution of the image with the telescope point spread function (PSF) and due to the exact pixelization of the resulting convolved image by the detector, but if these really are only minor effects, the Gini coefficient should still be measurable in the image plane.

This hypothesis was tested by simulating the lensing effect on actual images of low redshift galaxies treated as if they were galaxies at higher redshifts using steps which will be explained in further detail in the following sections. Briefly stated, a set of 33 detailed images of low redshift galaxies ($z \lesssim 0.45$) were chosen from the mosaics produced by the CANDELS team (Grogin et al. 2011, Koekemoer et al. 2011) and selected from the CANDELS UDS catalog (Galametz et al. 2013). Those galaxies were placed at higher redshifts and on or near caustics of a galaxy cluster scale lensing mass and run through a gravitational lensing simulation code to produce arcs of a similar size to those seen in observations. These final images of the arcs were then convolved with an HST-like PSF and rebinned to a pixel scale of 0.03 arcseconds/pixel. Varying amounts of noise were added to the images, and masks were made based on a generalization of the Petrosian radius defined in such a way as to be applicable to objects of arbitrary shape. Within these apertures, Gini coefficients were calculated and compared in a variety of ways which will be detailed in section 3.

2.2.1 Galaxy selection

The 33 galaxies that were used as sources in the lensing simulation were selected from the CANDELS UDS field (Galametz et al. 2013). They were chosen to span a range of morphologies (11 ellipticals, 20 spirals, and 2 irregular galaxies). It is difficult to know whether these galaxies have morphologies similar to galaxies at z $\approx$ 2 since small spatial scales are difficult enough to access in such galaxies that significant debate remains over their typical morphologies. However, these galaxies do span a range of morphologies and a range of Gini coefficients ($\approx$ 0.2 to $\approx$ 0.6) similar to those observed in previous studies at low to moderate redshifts (see, for example, Abraham et al. 2003 for low redshift galaxies and Lotz et al. 2006 for galaxies at z $\approx$ 1.5 and z $\approx$ 4). We chose galaxies that are low redshift, and large on the sky, so that detailed images of small scale structures were available. Objects chosen were typically between redshifts 0.2 and 0.4, corresponding to spatial scales of approximately 100–150 parsecs per pixel. Before lensing, an aperture was made based on a 12σ threshold on the stack of the F160W, F814W and F606W images, which, given the high signal-to-noise ratios in the CANDELS UDS images, still included most of the light associated with each galaxy. Pixels outside of these apertures were set to zero, minimizing the number of pixels that needed to be treated by the gravitational lensing ray-tracing code while also isolating the target galaxies from nearby objects.

2.2.2 Simulating gravitational lensing by ray-tracing

In our first gravitational lensing image simulation, the lensing mass had a spherical NFW profile with virial mass $\mathrm{M}_{200} = 10^{15}\ \mathrm{M}_{\odot}/h$ and concentration parameter c = 5. The lensing mass was placed at redshift $z_l = 0.2$. The source plane was located at $z_s = 1.0$. Because they were actually observed at a redshift much closer to 0, the 33 galaxies selected from the CANDELS UDS field were treated as though each pixel was 0.0075 arcseconds in the source plane even though the original images were drizzled to a 0.03 arcsecond/pixel resolution. For each galaxy we picked 50 random positions inside an 8 arcsecond square grid, centered on the lens so that the image positions would be close to caustics. For images of each galaxy in each of 3 filters (F160W, F814W, and F606W), the ray-tracing was performed using the code described by Li et al. 2016 at each of the 50 source positions, resulting in 4950 total lensed images (33 galaxies × 50 positions × 3 filters). At this stage, the images were sampled to a 0.01 arcsecond pixel grid. It is important to note that no SED shifting was done to convert observed low-z SEDs to $z \approx 1$ SEDs. Simulations using the "F160W" filter are therefore intended to portray rest frame NIR emission rather than optical or UV emission that was redshifted. For the purpose of testing the preservation of the Gini coefficient, having a realistic spatial light distribution is all that matters. For artificially redshifted photometry to add to our analysis, it would also have to account for morphological differences as a function of rest-frame wavelength, which is still highly unconstrained at all but the lowest redshifts.

The images produced by the ray-tracing code have a much finer pixel scale than could actually be observed with the Hubble Space Telescope. To create images similar to what would actually be observed, these images were convolved with a Gaussian PSF with the same FWHM as the HST PSF and rebinned to a final scale of 0.03 arcseconds per pixel. To test the effects of performing observations with different S/N per pixel, noise was added to each of the resultant arc images. For each arc, an image was produced with S/N per pixel of $10^{-3.5}$ through 10^{1} in logarithmic steps of $10^{0.5}$, yielding 10 images with different average

S/N per pixel for each arc. These logarithmic bins were chosen to correspond encompass the range of S/N levels typical of observations performed with HST. Thus, in summary, our database of simulated strongly lensed images includes nearly 50,000 individual frames (33 galaxies × 50 positions × 3 filters × 10 S/N levels) and many more individual arcs due to the high occurrence of systems with multiple images.

2.2.3 Generalizing the Petrosian radius to isolate arcs

Lisker 2008 found that the size of the aperture inside which the Gini coefficient is calculated can have a significant effect on the measured value of the Gini coefficient. Use too small of an aperture and only part of the galaxy is used for the measurement. Use too large of an aperture and so much sky is included that the Gini coefficient will be biased toward values near 1. Lisker 2008 measured Gini coefficients using elliptical Petrosian apertures of various sizes (i.e., using apertures defined using multiples of the Petrosian semimajor axis, which is used in place of a Petrosian radius). Apertures that best balanced inclusion of galactic light with exclusion of sky and therefore maximized the differences between Gini coefficients of different objects were found to be constructed from semimajor axes that fell between 2/3 and 1 times the Petrosian semimajor axis.

In light of these results, it is important to define a similarly inclusive aperture for arcs that also avoids including too many sky pixels. To do this, the definition of the Petrosian radius used by the Sloan Digital Sky Survey (Blanton et al. 2001, Yasuda et al. 2001) was adopted and then reformulated in terms of areas (rather than radii) in order to find apertures for galaxies whose shapes were severely distorted by strong gravitational lensing.

The Petrosian radius as used by SDSS is defined implicitly by the following equation from Yasuda et al. 2001:

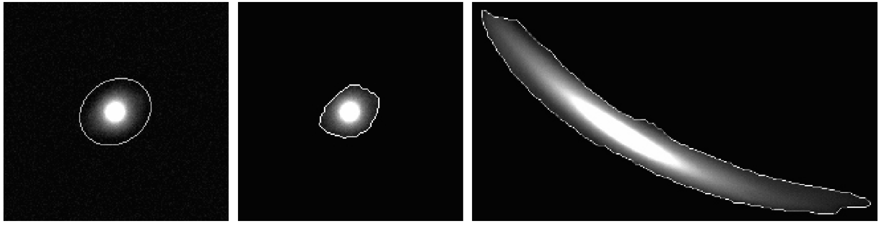
$$\eta = \frac{2\pi \int_{0.8r_p}^{1.25r_p} I(r) r dr / \{\pi[(1.25r_p)^2 - (0.8r_p)^2]\}}{2\pi \int_0^{r_p} I(r) r dr / (\pi r_p^2)} \tag{2.2}$$

where $\eta = 0.2$, $I(r)$ is the surface brightness profile of the galaxy, r is the radius over which we are integrating, and r_p is the Petrosian radius. Essentially, this compares the average surface brightness within an annulus with an inner radius of $0.8r$ and outer radius of $1.25r$ to the average surface brightness in the circle of radius r. When the ratio of these two values (the Petrosian ratio) is 0.2, r is the Petrosian radius.

This understanding of the Petrosian ratio leads very naturally to a redefinition in terms of areas. We look for an annulus with an area equal to $[2\pi(1.25r)^2 - 2\pi(0.8r)^2]/[2\pi r^2]$ or simply (1.25^2 - 0.8^2) times the area of the circle of radius r. The average surface brightness within the annulus is then compared to the average surface brightness within the full circle. When the ratio is 0.2, r is the Petrosian radius. With this reformulation in terms of areas, we extend the definition of a Petrosian aperture to include any arbitrary (connected) shape. To do this, we take a shape, calculate its area in pixels, and build contours either inward or outward until the area of the contours contains (1.25^2 - 0.8^2) times the number of pixels that the original shape did. When the ratio of the average surface brightness of the outer contour is 0.2 times that of the inner shape, we have found an aperture analogous to the one defined by the Petrosian radius. Defining the initial shape of the aperture, from which we build inward or outward (usually outward), requires nothing more than thresholding. For these simulations, we used a threshold of 2.5σ above the background. Fig. 2.1 shows the aperture produced by this method for an unlensed elliptical galaxy, compared to the elliptical Petrosian aperture produced for the same galaxy by Source Extractor (Bertin & Arnouts 1996). It should be noted that while the two images with apertures defined by this method have similar Gini coefficients (0.500 ± 0.008 for the unlensed one and 0.512 ± 0.003 for the lensed one), these differ substantially from the Gini coefficient inside the elliptical Petrosian aperture (0.649 ± 0.007). Comparisons of Gini coefficients between lensed and unlensed samples must therefore both use our aperture definition.

Each simulated lensed image was masked using the method described in this section. In some cases, especially those for which the S/N per pixel was very low, defining apertures in

Figure 2.1: Comparison of aperture types



Left: An elliptical galaxy and its Petrosian ellipse as determined by Source Extractor. Center: The same galaxy with its aperture determined using the method from this paper. Right: The same galaxy after being lensed and masked using the method from this paper.

this manner was not possible. At extremely low S/N, the initial aperture shape determined by thresholding does not necessarily follow the light distribution of the galaxy since it is more easily influenced by the sky noise. As a result, these starting apertures can be quite large and quickly grow past the size of the simulated image before the Petrosian ratio falls to 0.2. Alternatively, the starting aperture could be too small (if, for example, part of the arc is very bright, but the rest is faint) and as a result, the shape of the aperture will not accurately reflect the shape of the arc and parts of the arc may not be included. Instances like this were rare and typically occurred only at the lowest two S/N levels that we tested and for the vast majority of lensed images, there were no such problems.

2.2.4 Measuring the Gini coefficients

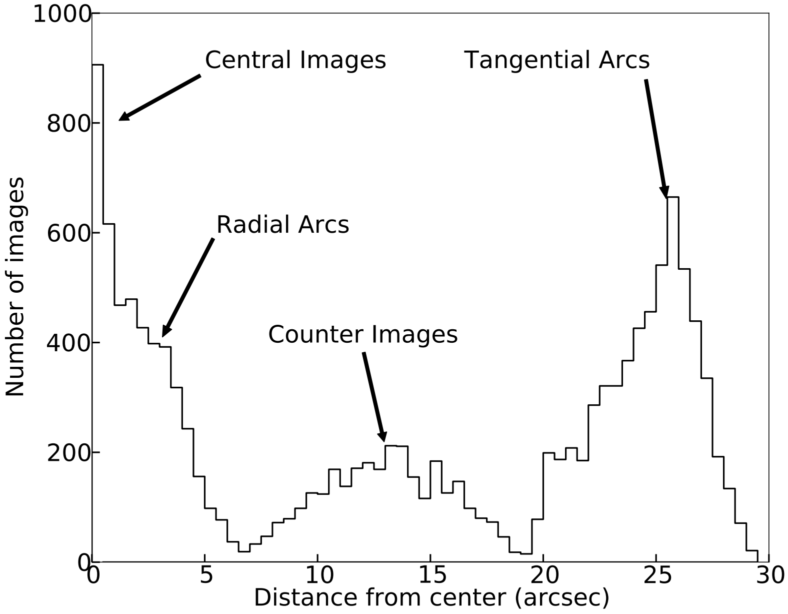
Finally, for each image of each galaxy in each filter and S/N per pixel bin, the Gini coefficient and its associated uncertainty was measured according to the prescription in Abraham et al. 2003 within apertures defined as in Section 2.3. For moderate to high S/N levels, changing the threshold of the cut used to determine the shape of the aperture from 2.5σ to 1.5σ or to 3.5σ resulted in changes to the measured Gini coefficient of only about 2–3%. Changing the threshold level slightly changes the initial shape of the aperture, but it will not have a significant effect on the final aperture size. This is why the Gini coefficient is hardly changed. If the Petrosian ratio is chosen to be something other than 0.2, though, the size of the final aperture could change significantly regardless of the chosen threshold level, which would have a strong effect on the measured Gini coefficient. As we will show in Section 3.1, such a change due to adjusting the threshold by 1σ is still small compared to the slight variation in the measured Gini coefficient between different lensing model realizations of the same source galaxy and is therefore not a significant source of uncertainty. However, at the very lowest S/N levels, the aperture shape can vary greatly depending on the threshold chosen, but at such low S/N levels, the sky dominates the measurement of the Gini coefficient anyway and renders it useless regardless of aperture shape.

2.3 Results of the simulation

2.3.1 Defining the sample of lensed images

Before the analysis was carried out, two cuts were made on the simulated images in order to ensure that the sample was as close as possible to what would exist in an observed sample. First, any images that were magnified by less than a factor of 4 were removed. This mostly removed central, demagnified images that would not be observed in real data. It also removed images where a piece of sky noise in the original unlensed CANDELS image fell on a caustic and resulted in a large image of something that wasn't actually the galaxy. The second cut was to remove all central images and radial images so that the analysis could focus entirely on tangential arcs and counter images, which are far more likely to be observed without significant contamination from light from the intervening lens in practice. We plot, in Fig. 2.2, the distance (in arcseconds) of the point in each of the arcs in our sample that is closest to the center of the lensing halo. We find 4 distinct regions. The innermost images are central images and have distances of nearly zero, which are often demagnified and unobservable. There is another bump in the histogram centered near 3", where the radial arcs lie. Then there are two more clearly defined peaks at around 6-18 and 20-30 arcseconds, corresponding to the counterimages and the tangential arcs, respectively. A cut was made in this space at 6.5", removing anything closer so that only tangential arcs and counterimages

Figure 2.2: Distances from Simulated Lensed Images to Halo Centers



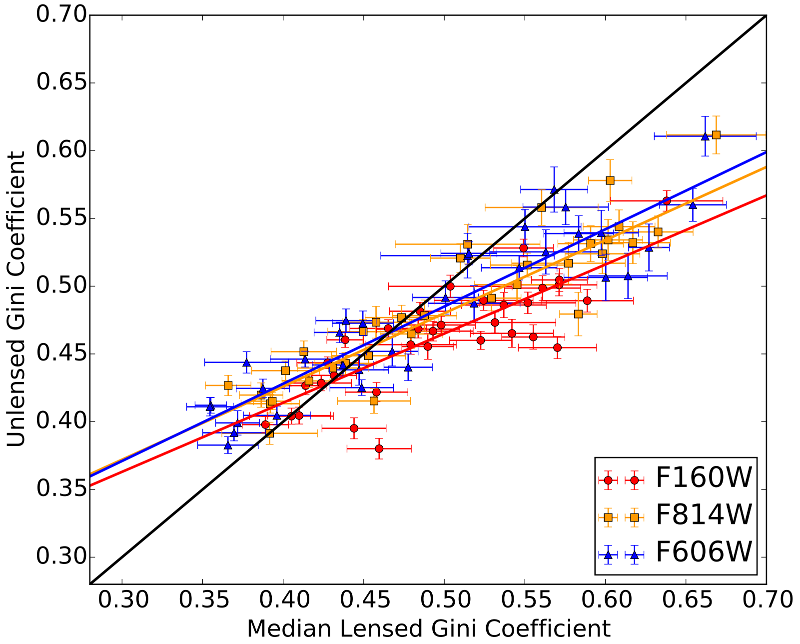
Distribution of the distances, in arcseconds, of the nearest point on each lensed image to the center of the lensing halo. We see four nearly distinct distributions consisting of central images, radial arcs, counter images, and tangential arcs, which allows tangential arcs and counter images to be identified using a cut based on this distance.

were included in the sample. All of the following results in section 3 are drawn from the sample of strongly lensed galaxy images that remained after these two cuts.

2.3.2 Relationship between unlensed Gini coefficient and lensed Gini coefficient

We consider first the relationship between the Gini coefficient of the unlensed source galaxy and the Gini coefficient of the corresponding simulated lensed galaxy images. We used, for the unlensed Gini coefficient, the Gini coefficient of the source galaxy as if it were placed at the same redshift as the source galaxies used in the simulation. This required resampling the image to reduce the pixel scale by about a factor of 3 (reducing the total image size by a factor of 9). In other words we convolved the images with a Gaussian PSF with FWHM equal to the width of the PSF of HST WFC3 or ACS images taken in the corresponding

Figure 2.3: Relationship between unlensed and lensed Gini coefficients

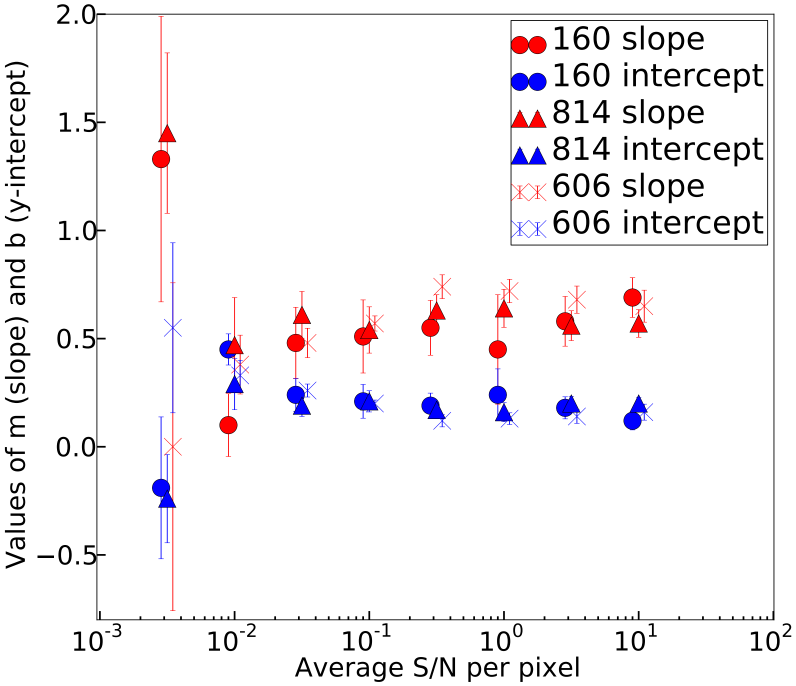


Unlensed Gini coefficients plotted against the median lensed Gini coefficients for all filters, using arcs with average S/N per pixel of 0.1. A relationship between unlensed and lensed Gini coefficients is evident, though the exact relationship may depend on the filter (likely due to differences in PSF). To guide the eye, the 1-to-1 line has been plotted in black.

filter (either F160W, F814W, or F606W), then rebinned the resulting images from a 0.01 arcseconds/pixel effective resolution to 0.03 arcseconds/pixel and added noise as we did with the lensed images (see Section 2.2). These unlensed images were then masked using the same method as was used for the lensed images.

For each source galaxy, in each S/N ratio bin and in each filter, the median Gini coefficient of all of the lensed images of that galaxy was plotted against its unlensed Gini coefficient. The resulting plot for model realizations in the 0.1 S/N per pixel bin is shown in Fig. 2.3. Uncertainties in the median lensed Gini coefficients are simply the standard deviations in the distribution of Gini coefficients for lensed images of each source galaxy in the relevant filter. The uncertainty in the unlensed Gini coefficient for each galaxy is calculated using the bootstrapping method described in Abraham et al. 2003. Best fit lines for each filter were calculated using a maximum likelihood technique with weighting for each point deter-

Figure 2.4: Slope and y-Intercept of Best Fit Lines to UGC vs. LGC Relationship



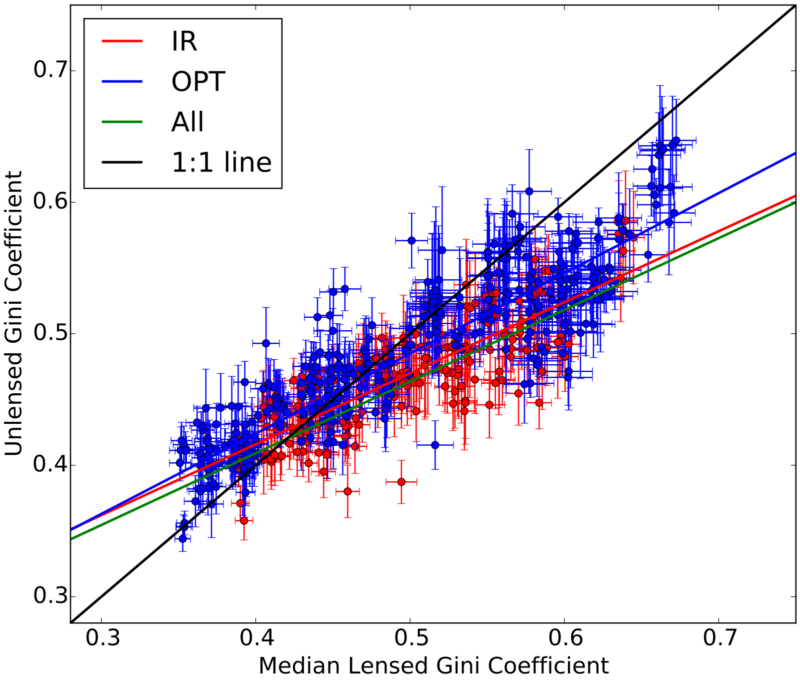
Values of the slope and y intercepts of the best fitting lines for unlensed vs. lensed Gini coefficient plots like the one in Fig. 2.3, but for values of average S/N per pixel between $10^{-2.5}$ and 10. Note that at each S/N level, a small offset has been artificially introduced for each filter, to aid in clarity by preventing points from overlapping with each other. There is little change in these parameters until S/N per pixel drops below $10^{-1.5}$

mined by uncertainties. While each filter has a slightly different slope and y-intercept, these parameters are relatively unchanged by the S/N per pixel values in the high S/N bins. This relationship is shown in Fig. 2.4 (note that a slight displacement along the S/N per pixel axis has been introduced for clarity). Uncertainties were again determined by bootstrapping the selection of source galaxies used in the trendline analysis. At most S/N levels higher than about 0.01/pixel, the best fitting parameters do not vary significantly with S/N. Furthermore, the two optical ACS filters (which had very similar PSFs and whose PSFs were sharper than for the WFC3 IR filter) are indistinguishable from each other. The slope of the best fit line for the F160W filter images tends to be shallower than for the other two filters regardless of S/N level until the noise becomes very high. Therefore to characterize the relationship between the lensed and unlensed Gini coefficients, we fit one line to the optical (sharper PSF) data and another line to the IR (broader PSF) data. In each case, the lines were fit to a dataset consisting of all points from all 6 S/N levels greater than 0.01 using a maximum likelihood technique with the significance of each point weighted according to the uncertainties in the lensed and unlensed Gini coefficients. All points used in this analysis are shown in Fig. 2.5 along with the best fit lines for the IR sample, the optical sample, and the entire sample. The following are the equations describing the best fit lines where UGC is the unlensed Gini coefficient and LGC is the lensed Gini coefficient:

IR: $\mathrm{UGC} = (0.54\pm0.04)\times\mathrm{LGC} + 0.20\pm0.02$
OPTICAL: $\mathrm{UGC} = (0.61\pm0.02)\times\mathrm{LGC} + 0.18\pm0.01$
ALL: $\mathrm{UGC} = (0.60\pm0.02)\times\mathrm{LGC} + 0.18\pm0.01$

It is interesting to note that the increased spread in the lensed Gini coefficient relative to the unlensed Gini coefficients suggests that there is more spatial information available in the lensed images, as one might expect because of the additional magnification. However, the fact that they are so similar to the unlensed Gini coefficients also suggests that most of

Figure 2.5: Relationship Between Unlensed and Lensed Gini Coefficients at all Signal-to-Noise Ratios



Median lensed Gini coefficients plotted against unlensed Gini coefficients for all filters, with optical filters in blue and the IR filter in red. Maximum likelihood lines are plotted for the optical filters (in blue) and the IR filters (in red). All values of average S/N per pixel of $10^{-1.5}$ and higher are used. The parameters for these best fit lines are given in section 3.1.

the structure that determines the Gini coefficient is still visible at the resolution used for the unlensed galaxies. This should not be surprising, however, since previous studies have shown that HST-like resolutions are sufficient to extract meaningful morphological information from the Gini coefficient of galaxies even up to $z \approx 4$ (Lotz et al. 2006). This means that it should be possible to use the Gini coefficient to compare samples of strongly lensed galaxies to their unlensed counterparts at similar redshifts, which could allow questions of selection effects in strongly lensed samples to be better addressed.

It is also worth noting that both the lensed and unlensed values of the Gini coefficients typically extended to higher values in the bluer filters even though the Gini coefficients behaved similarly near the lower end of the range in all filters. This relationship persisted even when the images originally taken in the F606W filter were convolved with a wider Gaussian to achieve a PSF similar to that of the F160W filter. This suggests that the effect actually has an astrophysical interpretation and is not just the result of PSF convolution and pixelization. If, for example, young and old stellar populations are spatially distributed with different uniformity in a particular galaxy, a mismatch between the Gini coefficients across these two filters would be expected. If this were the case, then it indicates that comparing Gini coefficients measured in different filters may yield further information about the morphology and stellar structure of a galaxy.

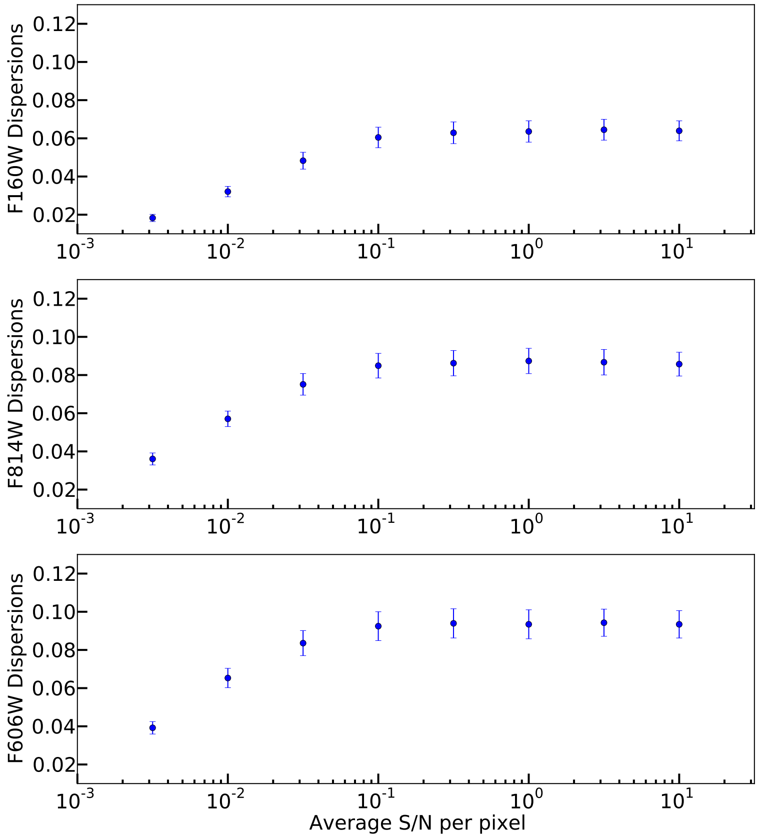
2.3.3 Gini coefficients and the effect of signal-to-noise ratio

We have seen in the previous section that the standard deviation of the Gini coefficient in different model realizations of the same lensed source galaxy is small relative to the overall dispersion of the median lensed Gini coefficients of the 33 different source galaxies. This means that the Gini coefficient can be used to distinguish between the lensed images of different source galaxies (Florian et al. 2016a). However, it is easy to see that in the extreme case where the average S/N per pixel is very low, the noise will dominate the Gini coefficient measurement and the Gini coefficients of the 33 different galaxies will begin to converge. The natural question to ask is: what is the minimum average S/N per pixel required for the conclusions of section 3.1 to hold? To test this, we plot the dispersion of the 33 median Gini coefficients against average S/N per pixel. That is, from all of the measured strongly-lensed images of each galaxy, we calculate the median Gini coefficient, and take the standard deviation of the 33 medians (one for each source galaxy). When noise is not the dominant source of flux (i.e., when the galaxies' Gini coefficients are discernibly different), we should expect a high dispersion, but the dispersion should tend toward zero as the S/N level decreases. This is borne out in Fig. 2.6 (where uncertainties in the dispersion are calculated by bootstrapping). It appears that the Gini coefficient is most informative at average S/N per pixel greater than or equal to about 0.1.

2.3.4 Lensed Gini coefficients of multiply-imaged sources

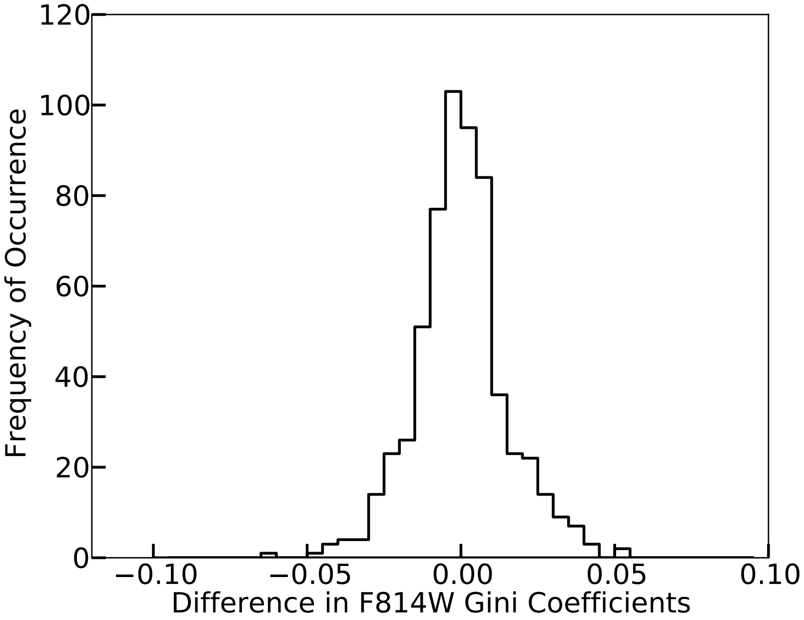
One way to check the stability of the Gini coefficient under gravitational lensing using real observational data would be to compare the Gini coefficients of galaxies that have been lensed in such a way as to produce multiple images. This can also be done with the simulated images. We have taken pairs from multiple-image configurations and subtracted their Gini coefficients. The histogram of the resulting distribution for the S/N = 0.01 per pixel subsample is shown in Fig. 2.7. Because the Gini coefficients were picked in random order, the distribution contains both negative and positive values and is, predictably, centered at zero. However, examining the spread of these differences in Gini coefficient, we find that the standard deviation is only about 0.015 while the total range in Gini coefficients as seen in Fig. 2.3 runs from just under 0.4 to just over 0.6 (though these are medians–some individual images have Gini coefficients that range closer to 0.3 or 0.7). This shows that the Gini coefficients of different images of the same galaxy should be expected to be consistent with each other, and that any differences are small compared to the differences possible based on actual structural differences between two different galaxies. Furthermore, it seems that differences in Gini coefficient due to changes in differential magnification from one realization

Figure 2.6: Effect of Signal-to-Noise Ratio on Stability of Gini Coefficient



The dispersion in the 33 median Gini coefficients at each S/N level. As S/N drops, the Gini coefficients for each galaxy approach the same value, but at high S/N levels, the dispersion in the Gini coefficients of the 33 distinct galaxies is much higher.

Figure 2.7: Gini Coefficient Stability in Multiple Image Systems



The distribution of differences in F814W Gini coefficients between pairs of images from all lensing configurations in the sample for which there are multiple images produced. The order of the pairs is selected randomly, leading to both negative and positive values. The average S/N per pixel for arcs used in this figure was 0.1. The narrowness of this distribution confirms that different images of the same galaxy should have very similar Gini coefficients even when strongly gravitationally lensed.

to another are small. However, it is worthwhile to more thoroughly investigate the effects of differential magnification, which we do in the following section.

2.4 Elliptical halos and the effects of higher magnifications and merging images

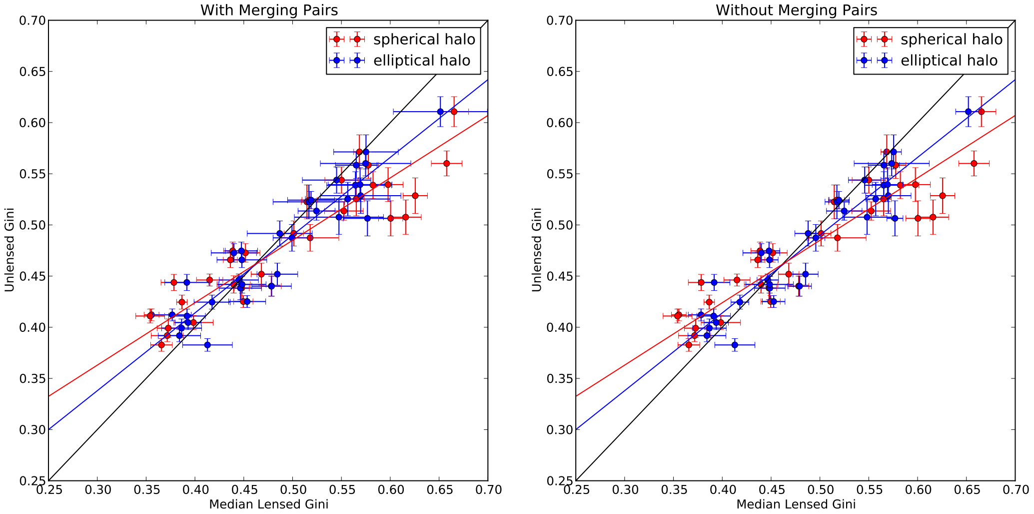
The case of the spherical halo, while allowing some shear and some magnification, does not permit certain classes of image configurations to form, including those with merging tangential arcs. Furthermore, even in cases of complete images, differential magnifications can be much higher in the case of an elliptical halo, for example, than for a spherical one. While we are encouraged by the results obtained in the spherical halo simulation, it would be prudent to further investigate the effects of these more exotic image configurations on the Gini coefficient. There are a few types of halos that could be used to do this. We could use a halo from a numerical simulation or a model from a Frontier Fields cluster (Koekemoer et al. 2014, Lotz et al. 2014), for instance. The downsides of these options are that they introduce additional uncertainties due to, for example, density estimators or uncertain empirical deflection matrices. Instead, since elliptical halos can produce merging images and high differential magnifications but are analytically defined, they allow better isolation of these particular effects. Therefore, we have chosen to investigate these effects using an elliptical NFW profile as the lensing mass.

The same 33 source galaxies as before were used as input and were lensed by an elliptical NFW profile with $\mathrm{M}_{200} = 10^{15}$ $\mathrm{M}_{\odot}/h$, c = 5.0, $z_l = 0.2$, $z_s = 1.0$, b/a = 0.8 (implying $\theta_E = 17.375$ arcseconds). Since sky noise had little effect when the S/N was at or above 0.01/pixel, only one S/N bin was used (S/N=0.1). Images were visually categorized by whether they were tangential/counter images, radial images, or central images, as well as by whether they were merging images. Merging images come, in general, from configurations where the source image crosses a caustic such that only a portion of the source is multiply-imaged to locations in the image plane that merge at the related critical curve, producing an apparent single image made up of two or more partial images. Apertures were created using the same prescription as in the spherical halo case. When images were merging, there was no attempt to break the arc into component partial images since in real observational data, it would be unclear where the breaks should be until a full lens-model has been completed. The relationship between lensed and unlensed Gini coefficients is shown in Fig. 2.8, the analogue of Fig. 2.3 except, for simplicity, only the F606W images are included. For comparison, the same values from the spherical halo case are also plotted. It is clear that there is increased scatter in the elliptical case. While the Gini coefficient appears to be well-preserved in most cases, when a source galaxy partially crosses a fold caustic, the interpretation of the Gini coefficient is more complicated. In such configurations, it is possible that four images (plus, in principle, a central image) can appear, but two of these images will be merging. In this merging pair of images, each of the two component images are not completely mapped into the image plane. When this happens, the merging images that are produced will together contain two highly magnified images of some portion of the source galaxy, but no images of the rest of the source galaxy (i.e., the portion of the galaxy that is inside the fold caustic will be imaged twice, but the portion that is outside the caustic will only be visible in the other three well-separated images—or two if the central image is not observable—but in neither of the component images of the merging pair). In configurations like this, the lensed images that are not merging will have Gini coefficients that are similar to the unlensed source image. However, the incompletely imaged merging pair may have a wildly different Gini coefficient. Whether the lensed Gini coefficient can be trusted in such a case depends on how much of the source lies on either side of the caustic—put differently, it depends on how much of the source galaxy is actually imaged in one or both components of the merging pair. When at least 50% of the source image is visible in the component images of the merging pair, the Gini coefficient tends to be trustworthy, but when less than 50% is visible, the lensed Gini coefficient tends to be biased very low. In Fig. 2.8, the left panel includes points that come from merging pairs that contain less than 50% of the source image while the right panel has omitted them. The preservation of the Gini coefficient is noticeably better when these cases are excluded.

$\mathrm{UGC} = (0.76\pm0.05)\times\mathrm{LGC} + 0.11\pm0.02$

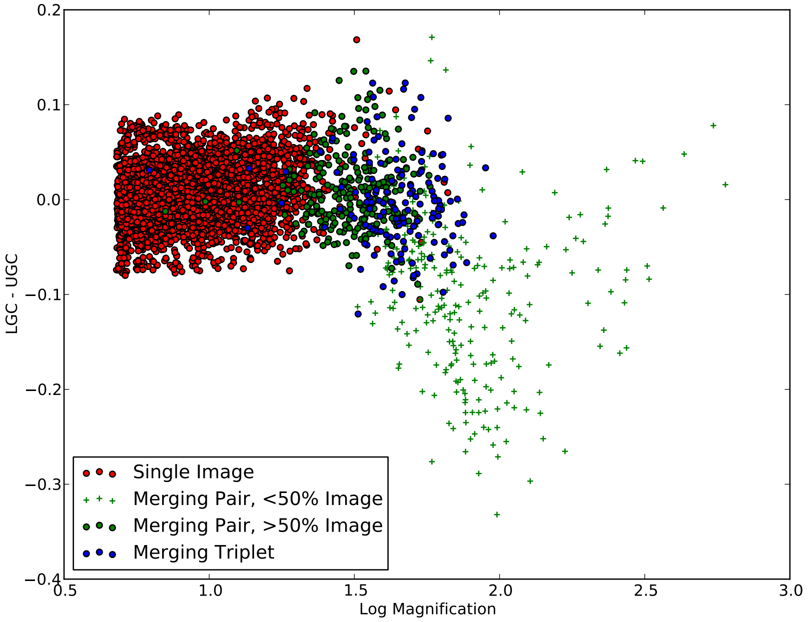
To further illustrate this effect, we have plotted, in Fig. 2.9, the difference between the lensed and unlensed Gini coefficients as a function of magnification for all tangential and counter images. The points are colored based on whether they represent values from individual isolated images, merging pairs, or merging triplets. No aperture was applied in the lensed image when calculating magnifications (i.e., all nonzero points in the noise-free

Figure 2.8: Relationship Between Unlensed and Lensed Gini Coefficients for Elliptical Halo Simulation



The unlensed Gini coefficient as a function of the median lensed Gini coefficient of tangential arcs and counter images in the F606W filter for sources lensed by an elliptical halo (in blue) compared to sources lensed by a spherical halo (in red). Uncertainties are calculated the same way as in Fig. 2.3. On the left, merging pairs are included in the sample. On the right, they are excluded. The best fit line is for the spherical case is plotted in red. The best fit line for the elliptical case without merging pairs is plotted in blue.

Figure 2.9: Stability of Gini Coefficient as a Function of Magnification and Image Configuration



The difference between the lensed and unlensed Gini coefficients of tangential arcs and counter images plotted against the log of the magnification. Different colors correspond to isolated single images, merging pairs, and merging triplets, while circular points are for images where at least 50% of the source image is visible in the lensed images and crosses when there less than 50% is visible. Only merging pairs have less than 50% of the source image visible in the lensed image. Lensed Gini coefficients deviate from the unlensed Gini coefficients by less than about 0.08 across a wide range of magnifications for the single images and the merging triplets. For merging pairs where only a small fraction of the source image is visible, however, the lensed Gini coefficient can deviate significantly from the unlensed Gini coefficient. Care must be taken in the interpretation of Gini coefficients for merging pairs of images.

simulated image were included). The figure demonstrates several important points. One is that except for some cases of merging pairs, where the lensed Gini coefficient is typically lower than the unlensed one, there is no noticeable bias in the Gini coefficient with increasing magnification, nor is there a clear difference in scatter across magnifications that are well-represented in our simulated data. Another is that high magnification merging pairs tend to have their Gini coefficients lowered by lensing, but isolated images and merging triplets do not show this bias. It is important to note that these are average magnifications. Merging pairs that are only partially imaged will have extremely high magnifications for the section that is imaged, but they may still have low surface brightnesses. Merging pairs that are almost fully imaged, while having a slightly lower average magnification, will still be much brighter due to having their brightest regions imaged while also having their brighter regions lie closest to the caustics (whereas the partial images typically have their faintest regions overlapping the caustics). For an illustration of this see Fig. 2.10, produced using an open source toolkit named "Bending Light" [1], which is designed for visually demonstrating configurations of strong gravitational lensing. As is evident in Fig. 2.9, when the fraction of the source that is visible in the component images of the merging pair is low, the Gini coefficient is untrustworthy, but when it is high, the lensed Gini coefficient is a reliable estimate of the unlensed Gini coefficient. This is easy to understand in light of the example images in Fig. 2.11. On the left, the circled image is of a merging pair where only a small, faint region at the outskirts of the source galaxy is visible in the lensed image. This small, and fairly uniformly faint portion of the galaxy is imaged twice (albeit at high magnification), but the brighter central region (the region that will contribute the most to increasing both the Gini coefficient and the observed surface brightness) is not visible. In this case, the observed Gini coefficient is very low. However, the arc would also have a very low surface brightness despite its high magnification and may not be observable in real data. And even if it was, no other photometric measurement would be able to tie it definitively to other members of

---

[1] http://linan7788626.github.io/bending_light_cython/

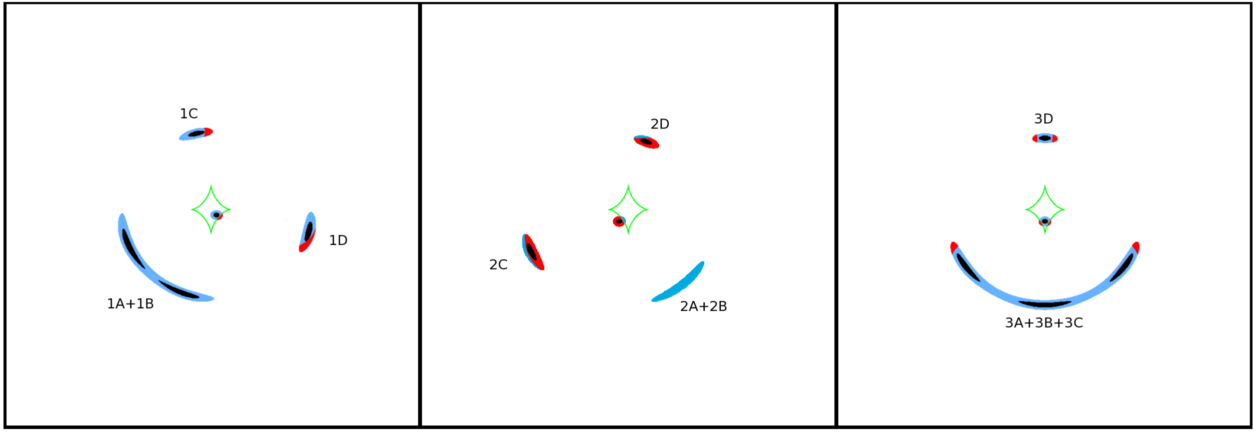
its image family. On the right, however, the circled image is of a merging pair where the central region of the source galaxy appears in the observed image. As a result, it is bright and easily observable, and its Gini coefficient hardly differs from its unlensed counterpart.

The results of this section reaffirm the consistency of the Gini coefficient as a morphological metric in lensed galaxies across a wide range of magnifications and image classes. They do, however, also draw attention to an important caveat. Care must be taken when interpreting the Gini coefficient of merging pairs of images. If a large portion ($> 50\%$) of the source galaxy is visible in each of the component images of a merging pair, it is more likely that the Gini coefficient will be consistent with the unlensed source and with other lensed images of the same source. If less than 50% the image is visible, the Gini coefficient is not a reliable metric. In the case of merging triplets, however, enough of the source galaxy is almost always observable in the lensed images that the Gini coefficient does not change significantly from its unlensed value. To summarize, the lensed Gini coefficient is a reliable estimate of the unlensed Gini coefficient in single images and merging triplets, as well as in merging pairs where a majority of the source image appears in at least one of the component images, but it is not reliable in the case of merging pairs where the component images do not, collectively, contain at least half of the source galaxy.

2.5 Discussion of these results in the context of galaxy formation and evolution

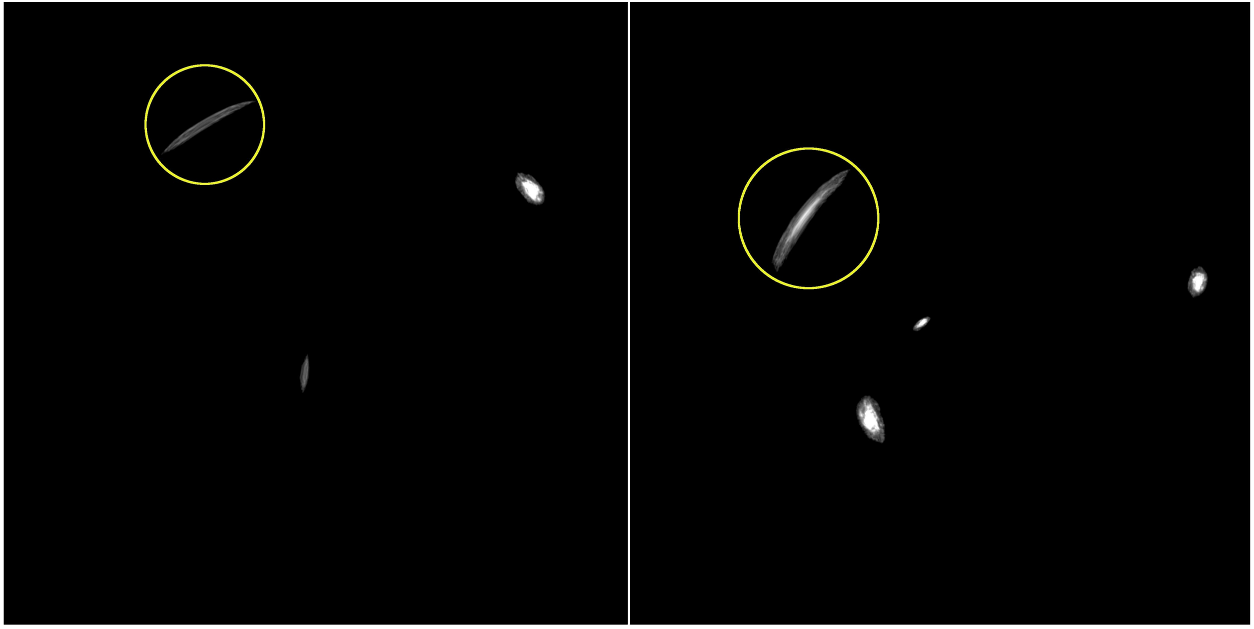
In this chapter, I have shown that the Gini coefficient is well-preserved by gravitational lensing and can therefore be measured in the image plane. There are some mild effects from PSF convolution and pixelization, but much of this can be modeled out, provided the aperture is defined carefully, via the prescription in section 3.2.3. This provides us with a morphological metric—one that is well-studied in both the low redshift and the high redshift unlensed samples—that can be easily measured in gravitationally lensed sources.

Figure 2.10: Illustration of Merging Image Configurations



Three different lensing configurations that include merging images. Each illustration depicts the tangential arcs and counter images in the image plane, as well as the tangential caustic and the placement of the source galaxy in the source plane. Central images, which are typically demagnified and not observed, are omitted to avoid confusion with the illustration of the source plane. The left panel shows a lensing configuration formed by a source that straddles the tangential caustic (thin green line) of an elliptical mass profile such that four images and a central image are produced. In images 1C and 1D, the entirety of the source galaxy is imaged. However, in the merging pair 1A+1B, the portion of the source image that has been colored red (i.e., the portion that lies outside the tangential caustic) is not imaged. In this case, however, the majority of the source is imaged—including the bright central region, colored black—and the Gini coefficient is measured robustly. In the middle panel, the source galaxy straddles the tangential caustic differently, with only a small fraction (colored blue) inside the caustic. This results, again, in two images (2C and 2D) and a central image that are fully imaged. However, in this case, the merging pair (2A+2B) contains two merging images of the blue portion of the source, but no images of the red portion. Notably, the bright central (black) region of the source, which would greatly influence the Gini coefficient, is also not imaged. In this case, the Gini coefficient is not measured robustly (and is typically biased low). While the merging pair of images 1A+1B would likely appear brighter on the sky than the pair 2A+2B (similar to the merging pairs in Fig. 2.11, where the pair on the left is more similar to 2A+2B and the pair on the right is more similar to 1A+1B), the average magnification of the 2A+2B pair is higher than that of the 1A+1B pair. On the right is a lens configuration that results in a merging triplet and a counter image. The entire source is imaged in both the counter image, 3D, and the merging triplet 3A+3B+3C, with most of the source being imaged three times. In this case, and indeed with most merging triplets, the Gini coefficient can be measured robustly.

Figure 2.11: Examples of Merging-Pair Image Configurations



On the left: A lensing configuration in which the apparent tangential arc is formed by a merging pair of images (circled) where the resulting combined image does not include the center of the source galaxy. The Gini coefficient of this merging pair is less than the unlensed Gini coefficient for the source galaxy by more than 0.1. On the right: A lensing configuration that is similar except that the merging pair of images (circled) does not contain the center of the source galaxy. The Gini coefficient of this merging pair agrees with the unlensed Gini coefficient within about 0.01. For both the left and right image, the same source galaxy was used.

This is particularly interesting in the context of galaxy formation and mass assembly histories for several reasons. First, the Gini coefficient is known to be sensitive to merger histories (Lotz et al. 2004). It has also been used, when paired with the clumpiness parameter or M_{20} (usually M_{20} because it is more robust at lower resolutions) to further distinguish between galaxy morphological types (Conselice et al. 2008). This opens the lensed sample up to the same type of analysis, especially if it becomes possible to make a measurement of clumpiness in the image plane as well, like I attempt to do in chapters 4 and 5.

The fact that these values can be measured in the image plane will likely prove key to creating statistically large samples of morphological information drawn from high redshift galaxies. While the Gini coefficient can be measured in both lensed and unlensed objects at high redshift, other important information, like the clump luminosity fraction can only be measured in high redshift galaxies using lensed sources with current technology. With new telescopes like WFIRST and the Large Synoptic Survey Telescope (LSST) planned for the not-too-distant future, the size of the sample of lensed objects is expected to balloon drastically. Measurements that can be made in the image plane will be the only ones that can be made on a reasonable timescale for samples of lensed objects in the thousands unless the lens modeling process can be automated.

A further advantage of the Gini coefficient is that due to its stability at lower resolutions, we have a potential bridge between the unlensed and lensed samples at high redshift for the first time. As long as care is taken in the aperture definitions, the Gini coefficients of high redshift lensed and unlensed objects can be directly compared. The importance of this is subtle, but notable. The lensed and unlensed samples are selected differently and therefore are likely biased differently. For example, the lensed sample may be able to access intrinsically fainter objects. After all, searching near caustics of galaxy- and cluster-scale lenses is one method currently being used to find high redshift objects (e.g., Laporte et al. 2014, Coe et al. 2015) that would not be detectable unlensed. If morphology is correlated with luminosity—and there are many reasons to expect that it is, since for example, luminosity tends to scale with mass and cold mode accretion dominates over hot mode accretion in low mass objects and hot mode accretion dominates over cold mode accretion in high mass objects—then the morphologies of the lensed and unlensed samples could be different reasons as simple as the sampling biases. A statistic like the Gini coefficient, which can be robustly measured in both samples, may be able to shed light on the types and severities of the biases in the selection of these two samples. Understanding the selection biases of high redshift samples will be necessary if we ever wish to draw conclusions about the evolution of morphology with redshift.

Incidentally, the stability of the Gini coefficient under gravitational lensing provides some auxiliary benefits—ones that may ultimately make it easier to automate lens modeling and source reconstruction, thereby allowing statistics like C, A, S, M_{20}, M, I, and D, which I briefly mentioned in the introductory chapter, to be measured in these objects. The Gini coefficient accomplishes this by improving image family identification for multiple-image lens configurations, which provide more robust constraints for lens models and therefore improves the quality of source reconstructions, all while reducing the amount of researcher and telescope time required for lens models to be made, and potentially bringing us a step closer to automating the entire process. In the next chapter, I will show how this is all possible and further elaborate on the benefits. Much of that chapter will be reproduced from my paper, Florian et al. 2016a.

CHAPTER 3

USING THE GINI COEFFICIENT FOR IMAGE FAMILY IDENTIFICATION

3.1 Challenges in lens modeling

Lens modeling is typically an iterative process, where an approximation of the lensing mass distribution is made, then repeatedly refined until it produces a model consistent with the locations of the observed lensed sources using a Markov Chain Monte Carlo method or something similar (for a review of lens modeling codes, see Lefor et al. 2013). Some codes use a mass-follows-light approach, in which the mass of the lens is proportional to the observed light distribution of the lensing galaxy or cluster. In such a case, it is critical to know the locations and fluxes of as many of the lensing galaxies as possible (if more than one), and the redshift of the cluster. But even when using a non-parametric approach—one where the mass distribution is allowed to deviate from the light distribution—knowing the locations of lensed images, and whether they are images of the same or different sources (and at what redshifts) is critical for providing the best constraints on the mass distribution and thus to create the best possible lens model.

This is where the Gini coefficient comes into play. Currently, there are very few metrics available to help a researcher determine which images in a complex lensing system are images of the same sources and which are images of different sources (in fact, even an unlensed, edge-on thin disk, if at the right orientation, can appear incorrectly to be a lensed image, so identifying which objects are even lensed at all can be a challenge). One extremely resource intensive and time consuming way—but the most accurate way—is to do follow-up spectroscopy in order to obtain redshifts for all of the images of all of the objects in question. Another way simply involves looking at the various images by eye and trying to divide them into image families subjectively, based on appearance. Colors can be measured fairly reliably in lensed objects because gravitational lensing is achromatic, though if there are color gradients, differential magnification can introduce some scatter. Some basic morphological properties can occasionally be distinguished by eye as well, which can help identify corresponding galaxy images. Over the last few years, lens models for the Hubble Frontier Fields have been published in which image family identification was carried out using exactly these sorts of methods (Jauzac et al. 2014, 2015, Johnson et al. 2014, Richard et al. 2014, Grillo et al. 2015, Ishigaki et al. 2015, Treu et al. 2016, Kawamata et al. 2016). Regardless of the method employed, it is important to get these image family identifications correct because changing which image family a single image belongs to can result in significant changes to the derived lens model (e.g., Jauzac et al. 2014, Sharon et al. 2012).

Unfortunately, these approaches are somewhat impractical even for small samples of lensing clusters. The next generation of surveys, using telescopes like the Wide-Field Infrared Survey Telescope (WFIRST), will reveal thousands of strong-lensing galaxy clusters. The ability to create models of these lensing masses will be critical to fully leveraging these datasets for the study of dark matter, structure formation, cluster physics, early galaxy and star formation, cosmology, and more. Lensing has already provided insights into all of these subfields. For example, Barnabè et al. 2011 and Umetsu et al. 2012 used lensing data to measure the mass distribution and dark matter content galaxy clusters and infer information about their internal dynamics. Others like Mahdi et al. 2014 and Jullo et al. 2010 have used arc statistics (for which determining the difference between multiple lensed galaxies and multiple images of one source is particularly important) to constrain cosmologies including both warm dark matter and λCDM cosmologies, and even to derive constraints on the equation of state of dark energy. Taking this type of data in another direction, Bouwens et al. 2014 used lens models to help their search for $z \geq 9$ galaxies, potentially shedding light on both early galaxy formation and the luminosity function of galaxies over cosmic time. Needless to say, the applications of strong lensing data—particularly when robust lens models are possible—are widespread. Now, as the sample of strong-lensing clusters grows to into the thousands, a faster, less hands-on approach than current techniques allow, with the least possible amount of follow-up imaging and spectroscopy will be necessary to take full advantage of these datasets to further explore these areas efficiently.

The Gini coefficient, as I showed in the previous chapter and in Florian et al. 2016b, can be measured accurately in the image plane. This provides a quantitative objective measurement of the morphological information that could otherwise be assessed only by eye. One might expect, then, that image families should cluster in this Gini-color space, allowing for the potential resolution of ambiguous image family identifications. In this chapter, as in Florian et al. 2016a, I show, using the same simulated data that I used for the work described in the previous chapter, that this is indeed the case and that the Gini coefficient can indeed be used as a tool for image family identification and that, because it does not correlate as strongly with color as other colors do, using one Gini coefficient and one color is a notable improvement over using two colors, and that it reduces the required observing time by a third because using two colors requires observations in at least three bands, but using a color and a Gini requires only two.

3.2 The simulated lensed sample

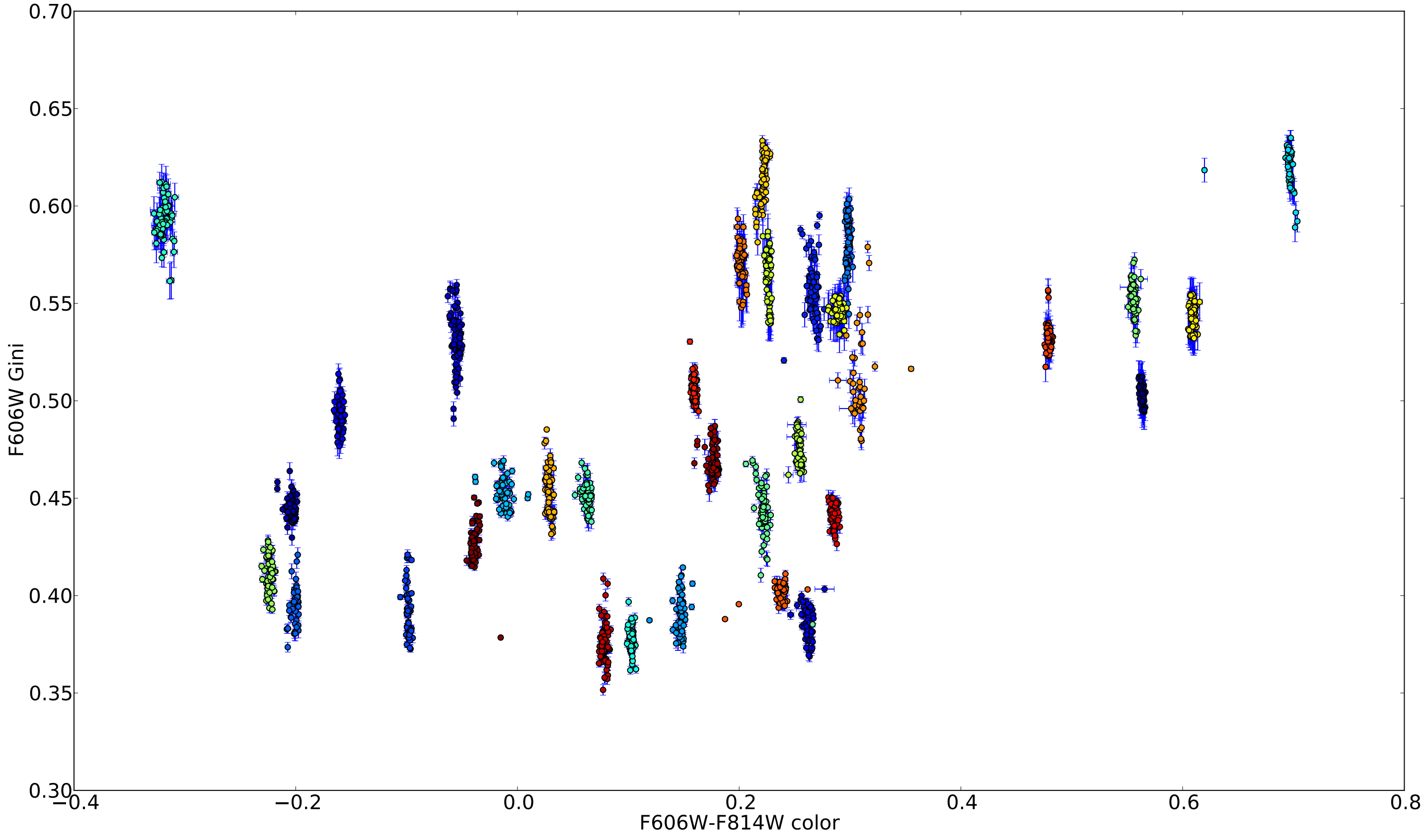
The strong lensing simulations used here are described in Florian et al. 2016b and the previous chapter. To briefly restate the basics: we selected low redshift galaxies (11 elliptical, 20 spiral, and 2 irregular) from the CANDELS UDS field to be used as source galaxies for a gravitational lensing ray-tracing code. Low redshift galaxies were chosen because direct HST observations of higher redshift galaxies do not contain as much small scale complexity due to the finite resolution of the telescope. These source galaxies were placed at redshift $z = 1$ and lensed by an analytical spherical NFW-profile with $M_{200} = 10^{15} M_{\odot} h^{-1}$ and concentration parameter $c = 5$ placed at $z = 0.2$. For each of the 33 sources, images were produced for 50 unique positions in the source plane near caustics, in each of 3 filters (HST ACS/WFC F814W and F606W, and WFC3/IR F160W). A gravitational lensing ray-tracing code (Li et al. 2015) was run for each of these configurations, and images of the resulting image plane configurations were produced. These were convolved with appropriate HST-like PSFs and resampled to HST-like pixel scales (0.03 arcseconds per pixel). Finally, Gaussian noise was added to degrade the average S/N per pixel of each arc to 0.1. Gini coefficients and colors were measured only for the arcs that were tangential arcs or counterimages (we did not consider typically demagnified central images or radial arcs because they are often either not seen, or significantly contaminated by cluster galaxy or intracluster light).

For purposes of measuring the Gini coefficient and associated uncertainties, apertures were defined as in Florian et al. 2016b, one for each filter. Gini coefficients were calculated according to Eq. 1 and uncertainties were bootstrapped as in Abraham et al. 2013. However, for colors, apertures were obtained from the stack of the images in the F160W and F814W filters. The resulting aperture was then applied to the F160W image and the F814W image separately. The total flux within this aperture was measured in each filter and the results were converted to instrumental magnitudes. Colors were defined as the difference between the magnitudes in each filter (F814W - F160W and F606W - F814W in this paper). Uncertainties in the colors were determined by making many separate noisy realizations of each simulated arc and finding the standard deviations of the resulting distributions. For each arc, an aperture was created. 100 different noise fields were then applied to the original image (with the average SNR remaining 0.1 per pixel), in each of the two filters. Finally, the aperture was applied to each of these 100 realizations and the flux and magnitude were calculated.

3.3 Separation of source galaxies in Gini-color space

To determine whether the Gini coefficient can be used, along with color, to help identify the image family to which a given image belongs, we plotted, for each source, the Gini coefficient in the F606W filter against the F606W-F814W color for every lensed tangential arc and counterimage from every model realization of that galaxy. The result is shown in Fig. 3.1. In this figure, the color of each point corresponds to the source galaxy (i.e., all points of a given color are different lensed images of the same source galaxy).

Figure 3.1: Clumping in Gini-Color Space

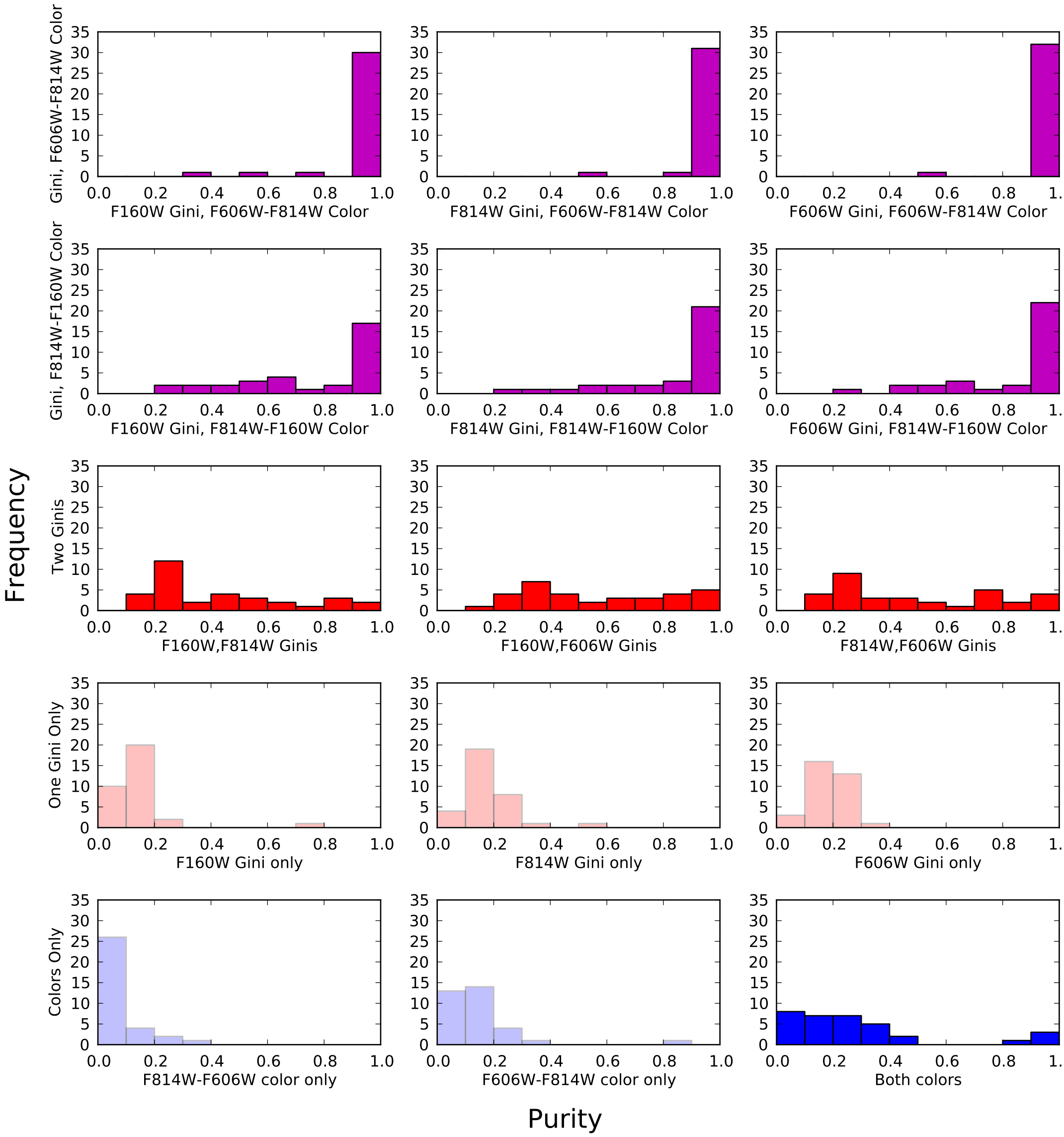


The Gini coefficient in the F606W filter plotted against the F606W-F814W color. The points are color-coded by source the source galaxy that was used to make each simulated arc (i.e., all points of the same color are different arcs created by lensing the same source galaxy, but at different positions relative to the caustics). Arcs from the same source galaxy tend to clump together in this space.

From Fig. 3.1, it is clear that different lensed images of the same source galaxy clump together in Gini-color space, suggesting that the combination of these measurements can indeed be used to distinguish between members of different image families. It is, of course, possible that two galaxies can have the same color and the same Gini coefficient. However, in cases where the two had the same color, they would not have been easily distinguished by the conventional means of looking at colors only. The strength of the Gini coefficient, therefore, arises from its ability to break that degeneracy and in that one can obtain a color and two Gini coefficients from only two observations whereas obtaining two colors would require at least three observations (and colors tend to be correlated more strongly than the Gini coefficient appears to correlate with color in r̃efGini814vColor). The significance of resolving the uncertainty in image families that arises from using colors only is explored in Fig. 3.2.

In Fig. 3.2, we compare the relative abilities of different pairs of color and Gini coefficient information to distinguish between the 33 source galaxies in our sample. For each combination of Gini coefficients and/or colors, a plot like Fig. 3.1. was constructed. We defined regions of the Gini-color, Gini-Gini, or color-color space based on the outlines of each of the 33 clumps. We then calculated the purity of each clump using the following process. For the i^{th} clump, we counted the number of points corresponding to galaxy i inside the clump and divided by the sum of that number and the number of points from any galaxy $j \neq i$ inside that clump to determine a purity, where 1 is a clump that is perfectly separated from all of the others and lower values indicate more contamination from images of other galaxies. We also compared methods using only a single Gini coefficient or only a single color. In these cases, the regions are 1-dimensional and defined by the two images with the least and greatest Gini or color values. Purities were calculated similarly for these 1-dimensional clumps. The distribution purities arising from each combination of Gini-color, Gini-Gini, color-color, single Gini, or single color are shown in Fig. 3.2. Red histograms correspond to methods of defining clumps that involve only Gini coefficient information, while blue histograms come

Figure 3.2: Effectiveness of Various Image Family Identification Methods



Histograms of purities of "clumps" like those in Fig. 1 determined from different combinations of data. Red histograms included only Gini coefficients. Blue histograms included only colors. Purple histograms use a Gini coefficient and color pair. Light red and light blue histograms use only a single color or a single Gini coefficient, while the darker ones use pairs of Gini coefficients or pairs of colors. Purities are the highest in general for clumps defined using one Gini coefficient and one color.

from methods that use only color information, and purple histograms include both types of information. The light red and light blue histograms denote methods that use only one piece of data (i.e., a single Gini or a single color) while the darker histograms use two pieces of data (two Gini coefficients, two colors, or in the case of the purple histograms, one of each).

We find that methods of separating images into image families that include both a Gini coefficient and a color consistently yield noticeably higher purities than all other methods. This means that the inclusion of morphological information from the Gini coefficient adds considerable power to image family identification methods above that would not be available from colors alone. While we find that using the F606W Gini coefficient paired with the F606W-F814W color gave the highest purities the most often, it is unclear why these particular filters yielded the best results in this study. It may be entirely due to the sharper in PSF in the F606W filter (and that the PSF is sharper for F814W images than for F160W images). But it may also be because of some sort of morphological quality that is more present in the F606W band at low redshift or because the F606W and F814W filters capture the 4000Å break in these low redshift galaxies (typically z∼0.2-0.4), causing the F606W-F814W color to hold significant morphological information that would not be available in the F814W-F160W filter pair at these redshifts, but could be at higher redshifts. To fully investigate this aspect of our result would require having a sample of highly spatially detailed images of galaxies with SEDs similar to those seen in the existing strong-lensing sample (i.e., with redshifts in the 1-3 range, or greater) which currently does not exist. However, it may be possible to simulate such a sample using a code like GAMER (Groeneboom & Dahle 2014) and a sample of SEDs from known moderate to high redshift galaxies that are more directly representative of typical lensed sources. Gini analysis of well-studied strong-lensing clusters with robustly identified multiply-imaged families would allow an in-situ test of the applicability of this method. I plan to do such a test using a cluster from the Frontier Fields for a future project.

Regardless of the reason for better results with some combinations of filters, it is clear that including the Gini coefficient from *any* filter in attempts to identify image families is a substantial improvement over using colors alone. And while degeneracies still remain even in the best filter combinations tested, it is possible that these degeneracies could be further broken by the inclusion of more Gini or color information (i.e., by using higher dimensional “clumps”) or by inclusion of some other measurement aside from these that is also preserved by gravitational lensing. This is exciting in the context of lens modeling, where accurate image family identification is required in order to make the best possible models of complex clusters like those in the Frontier Fields (e.g., Jauzac et al. 2014, Johnson et al. 2014, Richard et al. 2014, Grillo et al. 2015, Ishigaki et al. 2015).

3.4 Conclusions

We found, in Florian et al. 2016a, and I have shown in this chapter, that the Gini coefficient is likely to be an effective tool for the identification of image families in strong lensing systems with many images of an unknown number of source galaxies. We have shown that the Gini coefficient, combined with a color, is capable of distinguishing between image families with effectiveness substantially greater than using one or two colors only. This provides the additional benefit of minimizing the total number of observations that one needs to make in order to accurately sort lensed images into their respective image families. Using two colors requires making observations in at least three filters, but using a Gini coefficient and a color requires only two. This reduces the amount of observing time required to obtain reliable image family identifications by about a third, which is significant given the tremendous demand for telescope resources.

Moreover, these results show that it is likely possible to automate a process—namely image family identification by image morphology—that is at present mostly a by-eye process requiring a considerable amount of researcher effort. As we move into an era in which thousands of strong lenses will have imaging from space telescopes, such automation will be of great benefit. And as lens models become more accurate and easier to create, source reconstructions will improve as well, meaning that the morphological metrics that we wish to measure in the source plane—those that are now only commonly used at low redshift—will finally be possible to measure in a statistically meaningful number of high redshift galaxies, providing an unprecedented view of the effects of the mass assembly histories of galaxies in the early universe.

CHAPTER 4

STAR-FORMING CLUMPS IN LENSED GALAXIES

Motivated by the work on the Gini coefficient presented in the previous two chapters, one might reasonably ask what other morphological information can be extracted from the image plane without lens modeling or source reconstruction. A particularly interesting statistic to pair with the Gini coefficient is M_{20}, since a galaxy's location in the Gini-M_{20} plane is strongly influenced by its recent merger history (Lotz et al. 2004) and can be used to distinguish between several different morphological types, at least at low redshift (Conselice et al. 2008). This metric depends on relative physical locations within the observed galaxy, so it will not be preserved by gravitational lensing. The clumpiness parameter, an alternative to M_{20} which can be similarly paired with the Gini coefficient (Lotz et al. 2004), requires high resolutions that are typically beyond the diffraction limit of HST and JWST in high redshift objects. But because of the magnification effects of gravitational lensing, these structures can often be resolved in the lensed sample. Unfortunately, the clumpiness parameter as defined in Conselice 2003 will not be preserved by lensing because of differential magnifications (the smoothing kernel used will not correspond to the same spatial scale across the entire source galaxy), but it might be possible to find some other quantity related to clumpiness, that *can* be measured in the image plane, especially if gravitational lensing's conservation of surface brightness can be exploited. The fraction of a galaxy's light that is contained in clumps, for example, might be such a quantity. In this chapter, I explore a possible method for identifying, modeling, and measuring the luminosity of clumps in lensed objects relative to the total luminosity of the object using a process similar to the one used in a series of three papers by our collaboration (Johnson et al. 2017a, Rigby et al. 2017, Johnson et al. 2017b) to estimate similar properties for an unusually clumpy lensed source, SGAS1110+6459. The process involves decomposing the source into clumpy and smooth components using GALFIT, a code developed by Peng et al. 2002 which fits galaxy light profiles parameterically using a user-selected number of model profiles such as Sersic profiles (as well as the special cases of Sersic profiles: exponential, de Vaucouleurs and Gaussian profiles), among others. However, where the analysis for SGAS1110+6459 involved lens modeling and source reconstruction, I intend to develop a method to do the modeling and to measure the fraction of light contained in clumps directly in the image plane. As I have argued in chapters 2 and 3, if such a process is possible, it could dramatically reduce the amount of researcher time and effort required to make such measurements and increase the size of the sample on which such measurements are possible. After all, our SGAS collaboration wishes to measure this value and other clump properties in a sample of over 70 lensed galaxies from the HST program GO13003 (PI: Gladders), even though robust source reconstruction is not possible for a number of these objects.

4.1 The relationship between galaxy formation and clumpiness

The presence or absence of clumps in a galaxy, the spatial distribution of those clumps and of their sizes, and the fraction of the galaxy's light contained in those clumps can all tell us interesting things about the formation and evolution of those galaxies or of populations of galaxies over cosmic time. So what, specifically, can we learn from such properties?

Even a purely qualitative, binary measurement—whether a galaxy contains clumps or not—has already proven to be useful for studying galaxy evolution over time. Guo et al. 2015, Shibuya et al. 2016, and others (including several cited in Shibuya et al. 2016) have studied the fraction of galaxies with giant clumps as a function of redshift. They find that this fraction is relatively low until about redshift 4, where it begins to rise quickly, peaking at or slightly below redshift 2 and declining to the present day (not surprisingly, since these clumps host star formation and the star formation history of the universe is known to peak at a redshift around or slightly below 2 according to Madau et al. 1998). It is important to note that these efforts have focused on so-called "giant clumps," those that extend to a kiloparsec or larger in size. Knowing the cosmic abundance histories of galaxies with these giant clumps also provides insight into the times at which the processes that form these clumps become most relevant. It is thought that gas accretion onto disks feeds gravitational instabilities (violent disk instabilities, or VDIs) and causes the formation of giant clumps, which may then migrate inwards, break up, and thicken disks or form galactic bulges (Noguchi 1999, Dekel et al. 2009, etc.). However, it is also possible that these clumps come from major or minor mergers. Indeed, Shibuya et al. 2016 compared the fraction of clumpy galaxies over time to the major merger rates from Lotz et al. 2011 and found some degree of similarity. Regardless of formation mechanism, such giant clumps are commonly observed in high mass galaxies z$\sim$2 (e.g., Förster Schreiber et al. 2009).

The ability of gravitational lensing to provide access to objects with lower intrinsic luminosities means that the lensed sample may represent a sample containing more lower mass objects than in unlensed samples, which could influence the inferred fraction of clumpy galaxies over cosmic time. And, because the magnification provides access to smaller spatial scales, it can provide a look not just at giant clumps, but at star forming clumps with sizes in the range of tens to a few hundreds of parsecs, which is more typical of the ones in the Milky Way and nearby galaxies (Kennicutt 1984). This could shed light on the question of whether the giant clumps do indeed migrate inward and break into smaller ones. Alternatively, it is also hypothesized that the smaller clumps can form in-situ and still provide the clumpy, but mostly exponentially decreasing light profiles observed in nearby star-forming disk galaxies (Elmegreen et al. 2005, Elmegreen & Struck 2013). Determining which of these processes dominate when, and for which types of galaxies, will require being able to observationally determine the sizes, luminosities, spatial distributions, and abundance of these smaller clumps in high redshift objects—something that can only be done, in principle, with the lensed sample given current technology.

The abundance of star-forming clumps and the luminosity of those clumps relative to the overall galaxy luminosity are indicators of the cold gas fraction of those galaxies, as pointed out by both Shibuya et al. 2016 and Tacconi et al. 2013. So in addition to gaining insight into whether clumps are formed through mergers or violent disk instabilities, we can also use the clump luminosity fraction to make inferences about the fraction of cold gas in these galaxies, which speaks to the relative importance of hot and cold mode accretion over cosmic time. Since the properties of clumps in galaxies at high redshift inform our understanding of so many aspects of galaxy formation and evolution, being able to measure them in the largest possible sample of high redshift objects is, scientifically, highly valuable. To access the spatial scales necessary requires using a lensed sample for the foreseeable future and to make these measurements in the largest possible sample requires either being able to automate the lens modeling and source reconstruction processes (which, as I discussed in the previous chapter, is extremely challenging) or making measurements in the image plane. I will take the latter approach in the rest of this chapter, but the first hurdle to achieving this goal is to settle on a definition of a clump.

4.2 Previous efforts to define and measure clumps

In order to measure the clumpiness of a galaxy, we must first come to a consensus on what constitutes a clump. Intuitively, this is simple. When we look at images of most elliptical galaxies, for example, we see a light distribution (typically a de Vaucouleur profile or something similar) that peaks in the center and drops off smoothly toward the edges with few or no local maxima in between. Intuitively, we would call this a smooth galaxy. In contrast, when we see a galaxy with large bursts of star formation throughout its disk, or a recent merger, any smoothly changing radial light profile that might exist is disrupted by large local maxima—clumps of emission. We would probably call these types of galaxies clumpy. It is important to keep this intuitive picture in mind when formulating a definition, but we shall soon see that when it comes time to measure properties of clumps and how those properties relate to each other within a galaxy, and to other clumps in other galaxies at the same redshift or across cosmic time, we need a much more rigorous definition. For instance, is a galaxy with one or two very large clumps as "clumpy" as a galaxy with many small clumps? How big do these local maxima need to be, in size or in brightness, to qualify as a clump?

Early studies on the clumpiness of galaxies over time, like Elmegreen et al. 2007, depended on a visual classification of clumps, using something akin to the intuitive picture that I have just discussed. The authors went further and classified the morphologies into several categories based largely on their distribution of clumps. These categories were: chains, doubles, tadpoles, clump clusters, spirals, and ellipticals. "Chains" were objects in which several distinct clumps were observed in a single line through the galaxy. "Doubles" were similar to chains, but showed only two distinct clumps. "Tadpoles" were objects that has a major off-center clump, and a thin "tail" protruding mostly off to one side of the clump. "Clump clusters" were more randomly arranged clumps, while spirals and ellipticals are self-explanatory. This approach had the benefit of providing a classification scheme for distant galaxy morphology that could help subdivide the clumpy galaxy sample and allow researchers to begin thinking about how these different morphologies—even within the clumpy subsample—came to be. However, it is unclear whether these categories are truly distinguishable by the physical processes that went into creating them. For example, it is reasonable to wonder whether the so-called "chains" (and perhaps doubles) were simply "clump clusters" viewed edge-on. This sort of approach, then, is good for small samples where images can be evaluated by eye, and for applications where a simple binary categorization of whether an object is clumpy (and if it is clumpy in what way it is clumpy) is all that is needed. Unfortunately, it does not provide a particularly quantitative measurement of the clumpiness or of the clumps themselves and is dependent on the judgment of the researcher, particularly when it comes to just how prominent a clump needs to be in order to be considered a clump.

More quantitative algorithmic methods have been used as well. For example, Lotz et al. 2006 and Conselice & Arnold 2009 both approached the problem of galaxies with clumpy morphologies using classification systems based on the C,A,G,M_{20}, and/or S parameters. The benefits of such an approach are that they are well defined algorithmically, so they are perfectly repeatable and they provide a quantitative measurement of the overall clumpiness of a galaxy. For example, the clumpiness parameter, S, will be close to 0 for very smooth objects, and higher for very clumpy objects. This allows the clumpiness of multiple galaxies to be compared in a well-defined and well-ordered way. Unfortunately, the values C,A,G,M_{20}, as used in Lotz et al. 2006 are indirect measures of clumpiness (i.e., they measure properties other than clumpiness, but combinations of them tend to correlate with our intuitive idea of clumpiness). And by including the clumpiness parameter, S, Conselice & Arnold add a direct measurement of clumpiness, but that measurement says nothing specific about the properties of the clumps and there is a strict lower limit on clump size set by the smoothing kernel that is used to calculate that parameter. Any attempt to study individual clumps or the distributions of the properties of those clumps still requires a more precise metric (or several) that can be applied locally within a galaxy rather than only globally (i.e., ones that can identify aspects of individual clumps rather than be measurements based on the total galaxy light distribution) and that do not remove interesting sections of the parameter space—for example, low-luminosity star-forming regions that while small, are clearly detected.

There have several efforts in the last few years to perform clump detection algorithmically. This type of method, by virtue of being able to identify single clumps, makes it possible to then go on to measure individual clump properties and the distributions of those properties within a single galaxy or across many galaxies. And best of all, this type of method can be entirely automated in a repeatable fashion, and it does not depend on the subjectivity of a researcher's judgment. A particularly popular example of this sort of algorithm is the one set forth in Guo et al. 2015 which was subsequently used to study the fraction of clumpy galaxies in the universe as a function of redshift by Shibuya et al. 2016. In this algorithm, the image of a galaxy is smoothed by a 0.6 arcsecond boxcar filter and the resulting image is subtracted from the original image. Using this difference image, Source Extractor (Bertin & Arnout 1996) is used to detect residual sources. These sources are identified as clumps. At this point, further cuts may be applied. For instance, Shibuya et al. 2016 focuses mainly on the giant clumps that are observed to be more prominent at high redshift than at low redshift (Genzel et al. 2011, Elmegreen et al. 2004, etc.) so they imposed a cut on UV luminosity contained in the clump relative to the total UV luminosity of the galaxy (it must be at least 8%) and a cut on being off-center to avoid confusing it with bulges (must be between 0 and 8 effective radii away from the center). Notably, the results that Shibuya et al. 2016 derived for the fraction of clumpy galaxies over time using this method was reasonably consistent with the results derived by other authors using visual clump identification methods or other algorithms, though the agreement was, of course, not perfect. This method succeeds in several ways that others have fallen short. For example, it is not subjective like visual classification is (though this subjectivity does not seem to be particularly problematic in this case), and it can be applied to a large sample very efficiently (i.e., it can be automated). It does a pretty good job of capturing what we intuitively think of when we say "clumps." However, the smoothing kernel again introduces a minimum spatial scale. The relationship between small star-forming regions and so-called "giant clumps" may prove to be important in understanding the development of each of those features in populations of galaxies over time, so the introduction of a minimum spatial sale other than the necessarily-imposed lower limit due to the telescope point spread function, is disfavored. And when we have access to very fine spatial information, like we do in the strong-lensing sample, setting a limit of clumps having to be more than 0.08 times the luminosity of the host galaxy forces us to ignore potentially morphologically interesting features. This is especially important because the size scale of star-formation in the local universe is much smaller than these giant clumps, typically in the range of a couple tens of parsecs rather than the range of about 500pc–1kpc or greater for giant clumps, so we might expect that similar scales of star formation could exist in more distant galaxies as well and are just unresolved. Furthermore, it is a bit suspicious that the size scale of giant clumps aligns so well with the resolution limit of HST. Could what we see as giant clumps really be smaller, but appear to be so large because of the PSF of the telescope? Might they even be several smaller clumps blended together?

A highly magnified sample is required to address these questions for the foreseeable future (JWST, for example, will have a similarly sized PSF as HST, just at longer wavelengths).

With these previous efforts and the intuitive definition of clumps in mind, there are a few criteria that need to be satisfied in any clump definition that we may want to apply to the full size and luminosity range of real clumps.

1. It must include, at least, what we intuitively think of as clumps
2. It should not include bulges, or should be able to flag potential bulges
3. It must be repeatable: The result should not vary from researcher to researcher
4. It must be applicable to large datasets: It should be mostly or fully automated
5. It should not artificially impose a size limit: Lower limits on sizes should be determined by the telescope PSF

And if we wish to use the algorithm on lensed data, in the image plane (which is currently still a necessity for automation in the lensed sample), it must not be dependent on the clumps being mostly round. I will now present a first attempt at such an algorithm and an extension to it that allows the measurement of the fraction of a lensed galaxy's light that is contained in clumps. These methods have been somewhat successful in simulated data, but they still need further refinement in several areas that I will describe in detail.

4.3 A new clump-finding algorithm and a method of modeling clumpy and diffuse components that can be used with lensed galaxies

I have been developing a clump-finding code that can take lensed images, return a list of clump locations that serve as input for a GALFIT code that models the lensed images using a set of Gaussians. I hope to show that if the GALFIT components can each be identified as clump-like or diffuse components reliably, then it will be possible to use the best fit parameters from GALFIT for these components to determine the fraction of a galaxy's light that is contained in clumpy components. I have tested a fully automated version of this code on various subsets of an extremely large set of model galaxies for which the luminosities of the clumpy and smooth components are known in order to gauge its efficacy. The results so far are promising, but not sufficient for this method to be implemented in real data in its current form. In the remainder of this chapter, I will discuss the sample of lensed model images, the clump identification algorithm, the GALFIT modeling of the lensed objects, and the successes, failures, and dominant sources of uncertainties in this method. Based on this information, I will propose potential improvements to this method, including the incorporation of machine learning approaches.

4.3.1 Model galaxies and artificial lensing

While I have already discussed several existing methods for categorizing or quantifying clumpiness, or identifying individual clumps, all of them have been developed in the context of real broadband imaging data. In other words, they have the advantage of being developed on the most realistic images possible (i.e., real images), but by nature of using real imaging data, there is no knowledge of the true underlying source information. Are these methods really identifying every clump? Are they at least identifying every clump larger than 1kpc? Larger than 500pc? What is the real completeness and purity of these clump selections? Without testing these algorithms on a sample of sources with known properties, it is impossible to tell.

Shibuya et al. 2016 takes an interesting approach. They artificially redshift an object from the local universe and use the sizes of apparent clumps in the resulting image to set a limit on the detectability of clumps. This is certainly a legitimate way of performing this calibration. Unfortunately, the goal of that paper was to detect giant clumps, so any clumps detected in an artificially redshifted low-z object were deliberately calibrated out of the detected clump sample (since they are already known not to be giant clumps). This makes for a method that is well-tuned to finding giant clumps, but is not as widely applicable as is necessary for the purpose of studying the scales and distributions of star formation within high redshift galaxies. To calibrate a clump identification code for such a project requires using well-known sources to test the completeness and purity of clump detection and put uncertainties on any other derived values (e.g., the clump luminosity fraction). To this end, I have created a set of realistic model galaxies with very high spatial resolution using the iraf package artdata for use as a testing ground for my clump identification and modeling code.

In order to be able to sample from the full range of possible galaxy morphologies, I need a very large sample of model sources so that properties like the underlying light profiles, the axis ratios, the number of clumps, and the number and luminosity distributions of clumps can all be varied over realistic ranges. I therefore constructed a sample of 5000 source galaxies spanning a wide variety of morphologies. Parameters determining the light profiles of each galaxy and its clumps were picked to fall within realistic ranges inspired by literature values which I will cite as I detail the parameters involved. Each galaxy was chosen to have an effective semimajor axis of 2.5kpc, which is roughly consistent with the findings of (Shibuya et al. 2015) for effective radii of moderate to high mass galaxies around redshift 2. Because Johnson et al. 2017b determined that in some strongly lensed objects, clumps as small as about 10-20pc across could be detected, the source plane pixel scale of the model galaxies was chosen to be smaller than this—5 parsecs/pixel. The underlying light profile was given the same magnitude for each of the 5000 sources, and that profile was randomly chosen to be either an exponential or a de Vaucouleur profile with equal probability, and to have an axis ratio randomly selected from a uniform distribution spanning from 0.1 to 1.0.

SGAS1110+6459, an extremely clumpy galaxy—so clumpy, in fact, that in an HST sample of over 70 lensed sources, it contains the most visually identifiable clumps—contained roughly 30 clumps at sizes greater than the detection limit of Johnson et al. 2017b. Because galaxies with more than 30 detectable clumps appear to be quite rare, the number of clumps in each model source was randomly chosen to be anywhere from zero to 30, but no more. The clumps were round and were given Sersic profiles with $n = 2.0$. Gaussian clumps and Sersic profiles were several other indices were tested here, but $n = 2.0$ resulted in the most realistic looking sources when examined by eye than Gaussians or Sersic profiles with significantly higher or lower n. The location of each clump was determined as follows. The clumps were forced to lie within 2 effective radii of the center of the galaxy. The 1.4 power of the flux distribution of the exponential or de Vaucouleur profile within this spatial boundary was used as a probability distribution function from which clump locations were randomly drawn. The reason for raising the flux to the 1.4 power rather than using the flux itself is to better approximate the prediction of the Kennicutt-Schmidt law (Kennicutt 1998). And indeed, when clumps locations were selected using this method, which forced them to concentrate a bit more strongly near the center than when the probability distribution was generated from the unaltered flux distribution, the resulting model galaxies produced more realistic looking arcs after simulated gravitational lensing. Specifically, the apparent widths of the arcs tended to be narrower and better approximated the widths of actual observed arcs. Of course, some observed arcs really are quite thick, but this modification still allows that to occur.

Clump luminosities were chosen from a luminosity function with parameters roughly interpolated from Fig. 8 of Guo et al. 2015., who measured the luminosity function of giant clumps in z$\sim$0.5-3.0. Specifically, L* was taken to be -1.1 and ϕ* was taken to be 0.9. α, the faint-end slope, which was not well-constrained by the Guo et al. paper due to its intended focus on higher luminosity objects was randomly chosen for each of the 5000 sources, and allowed to run between -1.0 and 1.0, thereby sampling a very broad range of realistic clump luminosity functions. It should be noted that the luminosity functions in Guo et al. 2015 were meant to model the luminosity of clumps as a fraction of the total galaxy luminosity and my code uses the same prescription to approximate the luminosity of clumps as a fraction of the luminosity of the underlying diffuse component. This may lead my sample to have some sources with higher clump luminosity fractions than the sample observed by Guo et al, but this should be a relatively minor problem for two reasons. First, my sources still sample the very lowest ends of the distribution of fractional clump luminosities (i.e., there exist objects with no clumps, or only one very small clump, etc.), so my source catalog should contain the entire range of diversity of objects seen in Guo et al. 2015 even if it does contain a few sources outside of that range. Secondly, the Guo et al. paper focuses on giant clumps, and find that there are typically only one or two such clumps (which would dominate the impact on fractional clump luminosity). My sources may contain many more clumps than the Guo et al. sources, but they are both realistic in number and size when compared with Johnson et al. 2017b and by being largely drawn from the faint end, do not contribute so much to the total luminosity of the galaxy that they would pull my clump luminosity fractions very far out of agreement with the Guo et al. values.

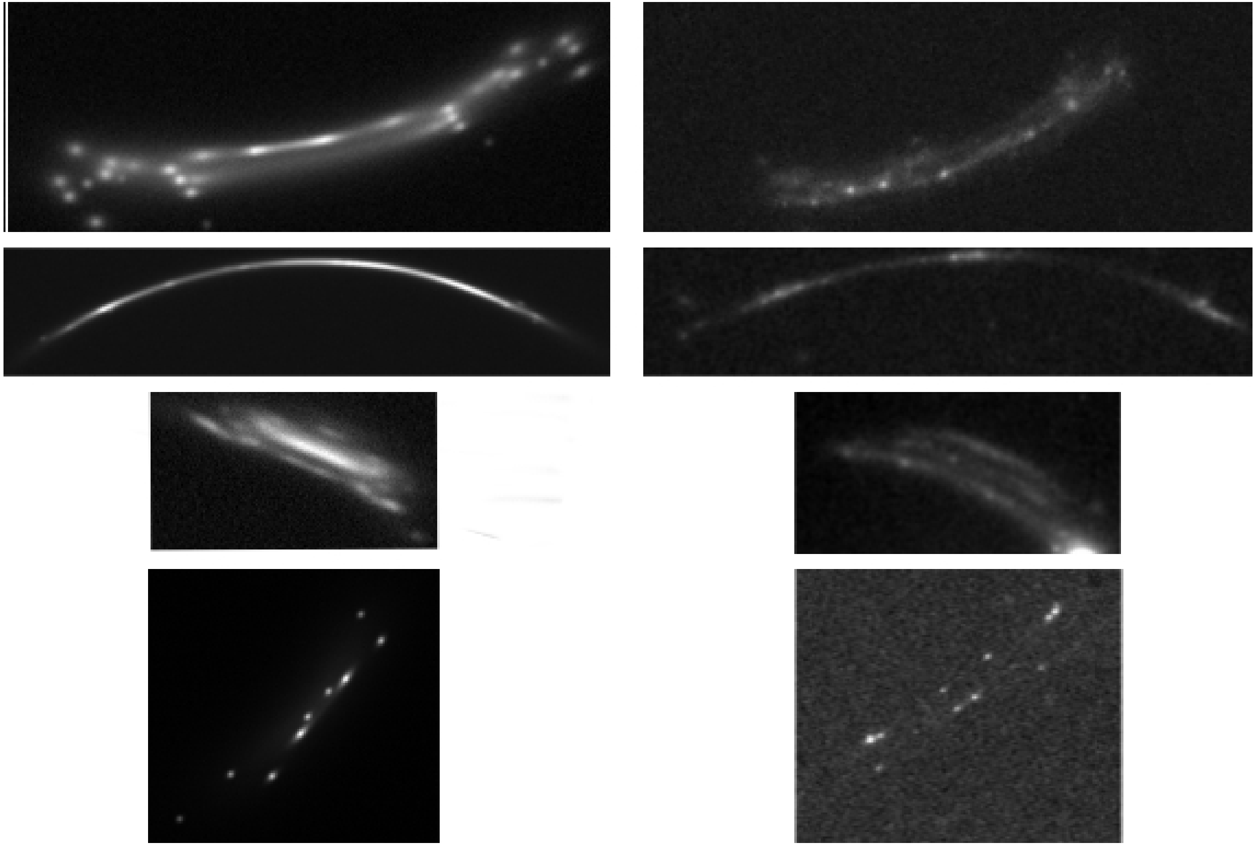
Lastly, the radius of each clump was allowed to be between 1 and 1000 parsecs, and scaled with the clump luminosity. These 5000 sources were then lensed by placing them near the caustics of numerically simulated halos drawn from the Outer Rim Simulation, a cosmological simulation led by the Argonne National Laboratory, and lensing them using the code described in Li et al. 2015, creating images with a 0.015 arcsecond/pixel scale in the image plane. For each realization of the lensing simulation, a source galaxy was drawn from the sample of 5000 input sources and a halo was drawn from a list of the 75 halos from the Outer Rim simulation that were between redshift 0.3 and 0.6 and that had Einstein radii greater than 5 arcseconds and masses greater than $2.25\times10^{14}\mathrm{M}_{\odot}$. Sources were placed at z = 2.0, since most presently observed arcs are from sources between about redshifts 1 and 3. The lensing halos were left at the redshift at which they were located in the Outer Rim Simulation (i.e., somewhere between 0.3 and 0.6). Each source was lensed by about 5 different halos, on average, to widely sample the combinations of possible pairs of source morphologies and lensing morphologies. This resulted, after accounting for multiple image systems, in more than 40,000 unique lensed images. These images were then PSF convolved with a Gaussian PSF with FWHM=0.067 arcseconds, similar to the F606W filter on HST's Wide Field Camera 3 and several of the expected JWST infrared PSFs. They were then rebinned down to 0.03 arcseconds per pixel, to approximate the pixel scales of HST and JWST. Individual images in multiple image systems were separated by taking everything that was connected within 2 effective radii (determined by lensing bitmasks that encoded the locations of each smooth and clumpy component). Individual aperture files were made for each of these arcs, masking out anything that fell outside of 2 effective radii (and that was not contained within the half-light radius of a clump). Anything that fell inside these apertures was considered signal, and noise was added to each image to bring the average signal to noise ratio per pixel to 0.01. Fig. 4.1 displays a selection of simulated arcs (on the left) and real observed arcs (on the right).

4.3.2 Clump-identification

Once the relevant areas corresponding to each "arc" or image were identified, the next step was to identify clump candidates. Keeping in mind that the purpose is to test measurements of clump-to-total luminosity fractions across a broad range of source galaxy morphologies and lens configurations, it was necessary to make this an automated process. Because there is a second step, after clump candidate identification, in which the components were decomposed into Gaussians, it is preferable to have a *complete but impure* list of clump candidates rather than a *pure but incomplete* list. Contamination can be dealt with in the decomposition process, using an iterative method.

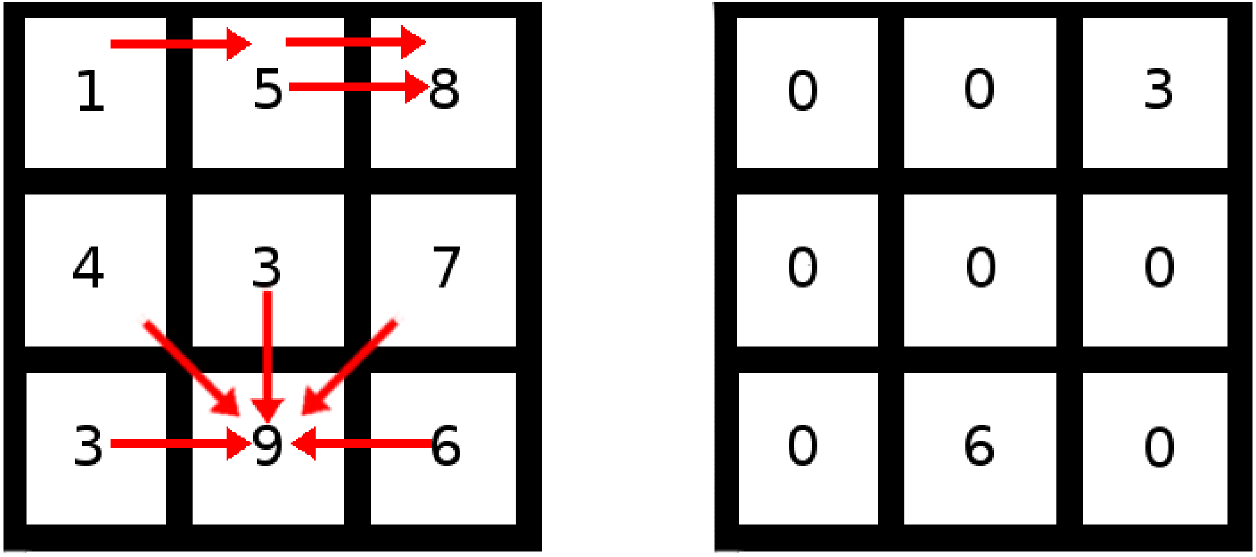
I used a fairly simple algorithm, similar to a watershed algorithm, to determine the initial list of clump candidates. For each pixel determined to be in the arc using the criteria in the previous section, the gradient of the flux was traced until a local maximum was found. In a separate map, of the same dimensions as the input image, a tally was kept for each local maximum, which was incremented each time a pixel, after following the local flux gradients,

Figure 4.1: Randomly Simulated Arcs vs. Real Arcs



Randomly simulated arcs (left) compared with real observed arcs (right). No attempt was made to reproduce the arcs on the right. These were randomly generated using the constraints described in this section and selected for this figure based on their similarity to real arcs, for illustrative purposes. Clearly this algorithm creates realistic images spanning a wide range of lensing configurations. Note: Noise levels in the simulated images are not necessarily matched to the noise levels of the observed images, and the stretch of each of these images and the residuals has been altered to best show contrast, and the pixel scales have been altered to better fit the page.

Figure 4.2: Schematic Illustration of Clump Finding Algorithm



An illustration of the clump finding algorithm for a very simple case. The grid on the left represents flux values for a 3x3 image. Red arrows show the direction of the maximum local gradient at each point, and therefore the "path" that a pixel would follow. The grid on the right represents the resulting weight image, where all points that are not local maxima are set to zero, and the local maxima are weighted by the number of pixels that "travel" to that pixel along the gradients (including that pixel itself).

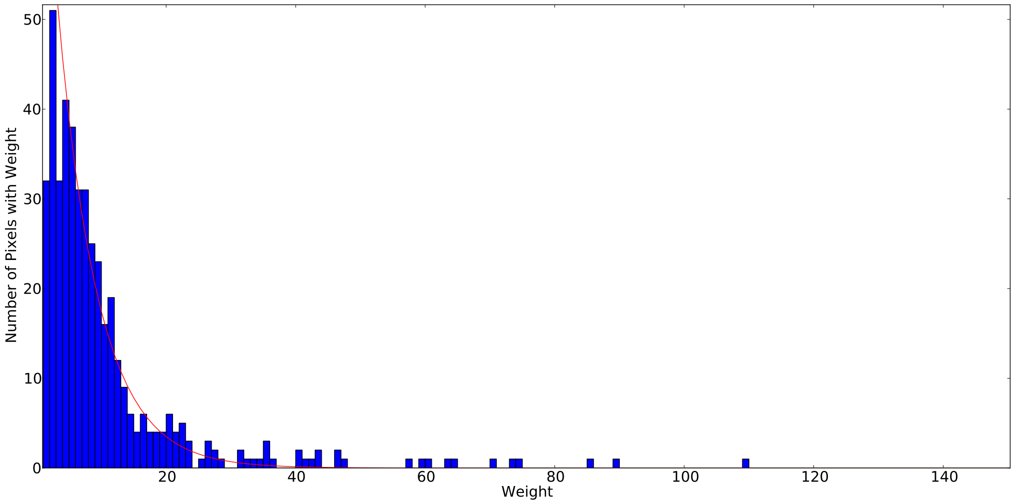
was traced to that location. Essentially, this creates a map of all of the local maxima, with a weight corresponding, roughly, to the size of the region for which it is a local maximum. See Fig. 4.2 for an illustration of this process in the very simple case of a 3×3 square.

This, of course, results in many very small local maxima which are not the result of detectable clumps, but are instead simply random noise spikes. These, however, especially in high noise images, tend to have very low weights because they are local maxima of very small regions. Additionally, because of the gradient in flux moving outward from the center of a galaxy or from the center of a clump, these noise spikes tend not to result in local maxima except near the outskirts of the visible portions of a galaxy (where the gradient in the underlying galaxy's light profile is flatter). So in addition to the low weights, there are spatial constraints that can theoretically be used to filter out spurious clump detections. In this work, I have not incorporated this spatial information because, as we shall see in section 4.4, the GALFIT modeling can be done very well using only a cut based on the weights.

To make this cut, it is sufficient to make a histogram of the weights and fit the histogram with a decreasing exponential function, using only weights of 2 or greater (it is only at higher weights that the distribution begins to resemble an exponential-like function, though it should be noted that in many cases, a partial Gaussian does a decent job of fitting the whole range of weights), and make the cut based on the discrepancy between the predicted number of clumps based on this curve-fitting and the actual number of detected clumps with that weight. Because of noise, there should be a rather large number of pixels that are local maxima for very small regions (i.e., for only a few pixels), whereas for bright and extended clumps, their brightest points tend to be local maxima for very large regions. And indeed, the largest clumps tend to be extreme outliers in these histograms. For this work, I used the following criterion: If the value of the fitted exponential is less than 1% of the actual number of clumps detected with that weight, the clump was determined to be "real" and included in the candidate clump list. An example of one of these histograms is shown in Fig. 4.3 where a few of the most significant clumps (weights that fall far to the right) have been omitted for the sake of showing better detail in the main distribution.

There are a few peculiarities of this method that provide advantages in certain cases. For example, this method works best for very large images and for very high resolution data. The more pixels in an image, the better one can determine the histogram used for making cuts and the higher the weights for even the smallest real clumps. Increasing noise levels do not necessarily make this algorithm perform worse, as one might expect (though at a certain point—when clumps aren't clearly detectable by eye, for instance—it obviously will). As mentioned before, higher levels of noise result in large numbers of spurious clump detections near the edges of a galaxy, but they will have very low weights. At very low noise levels, there are fewer noise spikes that result in local maxima, but that can mean that these pixels are local maxima for larger regions, giving them larger weights and making them harder to exclude with this type of cut. It may, therefore, occasionally be advantageous to add a small amount of noise to images before running this algorithm, though it appears this is only really

Figure 4.3: Example of Cut on Clump Candidates



A histogram of weights determined by the clump-finding algorithm, in blue, with an exponential fit of all weights greater than 2 overplotted in red. For any weight where the red curve was less than 1% of the histogram value, all clumps with that weight were kept. This histogram is zoomed in to better show detail in the histogram and fitted curve, but there are more weights that occur exactly once each far to the right of the window chosen here, one at 535 and one at 996. In this example, clumps with weights larger than about 40 would have been included in the clump candidate list.

a problem when signal to noise levels are so high that they are not representative of signal to noise levels typically attained in real data. A third peculiarity of this method is that when images are highly sheared, there can be very long clumps with several local maxima along them due to things like differential magnification across a clump (these features are present in the noise-free simulated images, so they do not come from noise). Because most of the pixels that trace gradients to maxima within these highly sheared clumps come from roughly perpendicular to the arc (if they go along the arc, they are likely to encounter one of the other maxima within the sheared clump before finding the actual maximum of the clump), they may have slightly lower weights than slightly fainter, but rounder clumps (where pixels from every direction can track to the same local maximum). However, these are typically still detectable even after cuts are made on the weight histograms. And in fact, because the longer clumps are best modeled with multi-component fits in the decomposition stage, finding multiple maxima within a single sheared clump actually results in better models. However, as we will see, this can result in increasing the difficulty of disentangling the clump components from the smooth components in the automated models.

4.3.3 Modeling clumpy and diffuse components

Once clumps were identified, GALFIT was used to model the clumpy and diffuse components of each of the lensed images. First, the curvature of the arc was determined based on the apertures created for each image as described earlier in this chapter. Images with large curvature were artificially straightened and resampled using an IDL script written by Prof. Gladders, and the lists of identified clump positions were transformed accordingly. Images with little or no curvature (i.e., images with low shear) were not transformed in this way. After this stage, the images and lists of identified clumps were used as input for an automated attempt to model the clumps and the diffuse components using GALFIT.

Automated application of GALFIT proceeded using the initial input clump list as the starting location for a sequence of Gaussian components. Using Gaussian components instead of more complicated profiles like Sersic profiles that might better match the components of an unlensed source (though it is not immediately clear that the basic structure of a Sersic profile would survive lensing anyway) decreases the parameter space that GALFIT must search (since Gaussians are Sersic profiles with the Sersic index locked at 0.5) and vastly improves the speed at which GALFIT can complete the modeling process. Based on the apparent success of Johnson et al. 2017a, where Gaussian components were also employed, the decrease in model complexity would seem to be a reasonable trade for the increase in speed. The initial clump lists served much of the function of the human analyst in the procedure reported in Johnson et al. 2017a. This initial model was refined to a best possible fit, with the clump locations confined to within a few pixels of the input location. This is never a sufficiently good fit. The residual map was used to guide the addition of further Gaussian components: a single further component was placed on the highest residual peak, and the model was then refit in a narrow spatial window +-20 pixels around that peak, with only components in that window allowed to vary. Up to 4 further single components were added in this manner, and then the entire model was re-optimized. At each complete re-optimization, any components that have converged to a magnitude much fainter than the other model components (signifying that they are unnecessary at that location with the total model) are removed. This process of adding further components one by one in small spatial windows, interleaved with complete model re-optimizations was iterated until no residual pixels were greater than 5-sigma, or until 10 complete optimization loops had been completed.

4.4 Results of this approach

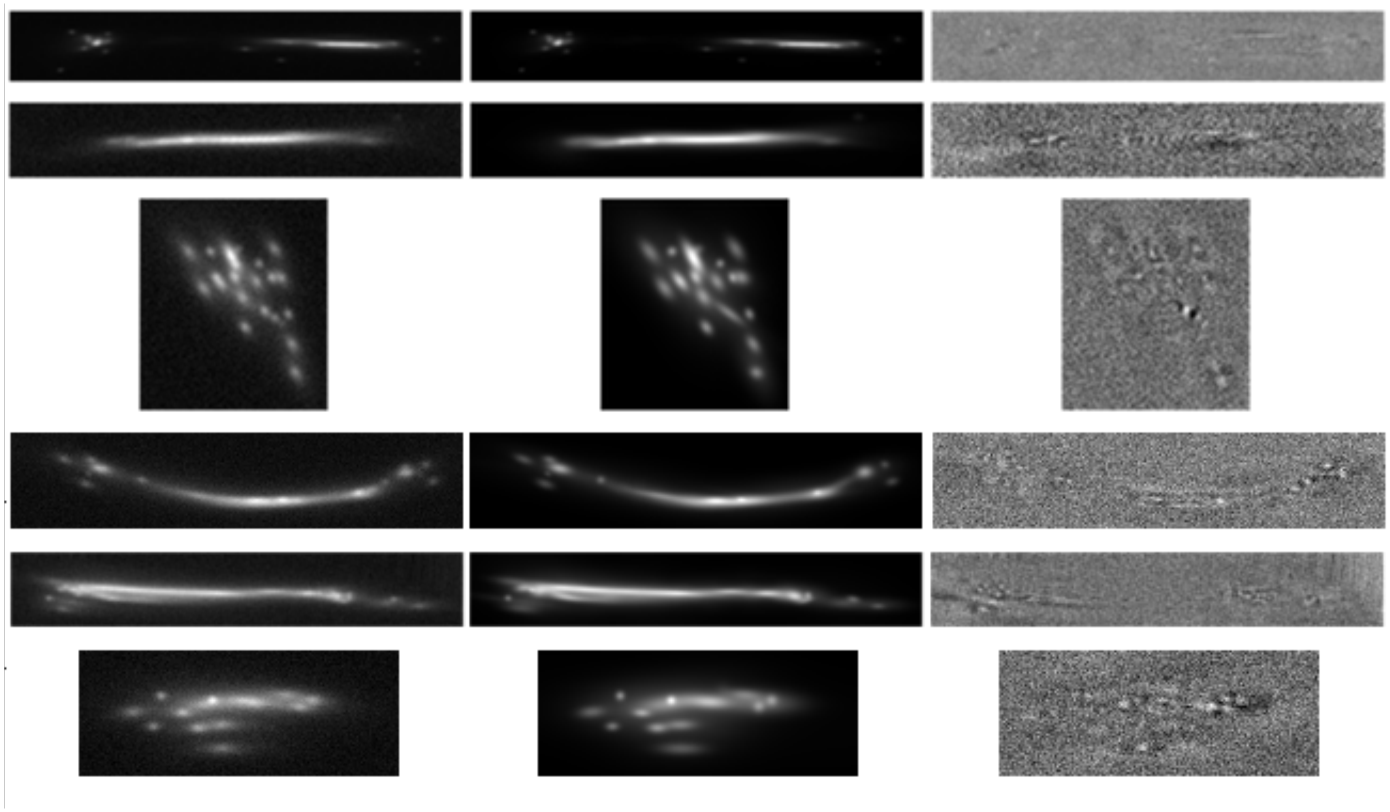
By eye, models do a reasonable job of fitting the lensed galaxy images in most cases. And indeed, examining the residual images confirms this assessment as shown in Fig. 4.4 where I have compiled a small sample of lensed model images in the left column (after straightening, if applicable), images produced by the automated modeling process in the center column, and residuals in the right column. The centers of clumps are not always fit perfectly, as is often the case with GALFIT and similar techniques because that is where the light profile is changing most rapidly. Occasionally, there will be some fringing in the wings of the rounder clumps, likely as a result of fitting Sersic profiles with n=2 using Gaussians (or because artdata's Sersic profiles deviate slightly from actual Sersic profiles), but the integrated magnitude determined by GALFIT is expected to still be fairly accurate in these instances.

From these models, I hope to calculate the fraction of each galaxy's light that is contained in clumpy components. To assess how well this method allows the clump luminosity fraction to be estimated (if at all), I will first consider the simplest cases: Those with high image fractions (as defined in chapter 2) where there is high magnification, but very little shear (and in particular, very little shearing of clumpy components, like the images in the bottom row of Fig. 4.1). If this method does not work on such a simple case, it cannot be expected to work on more complex images. Furthermore, simple cases like these allow for easier examination of the potential sources of uncertainty. For the rest of this chapter, we will be concerned only with these cases. In particular, I will discuss a subset of 100 such images, selected by eye to have this type of image morphology and a wide range of intrinsic clump luminosity fractions. I will examine the results of the fully automated method described in the preceding sections as well as attempts to further refine these Gaussian-based models to explore sources of uncertainty. Motivated by these results, I will explore further refinements of this method, including models based on Sersic profiles.

4.4.1 Results of the fully-automated Gaussian model

The modeling algorithm is given significant leeway to place new components in locations that were not marked by the initial clump-finding algorithm in order to fit diffuse components and to better fit clumps that have been magnified or sheared enough that they cannot be well-approximated with a single Gaussian. Because of this, it can be difficult to identify, after the modeling process, which GALFIT components correspond to clumps and which

Figure 4.4: Visual Results of Automated Modeling Process

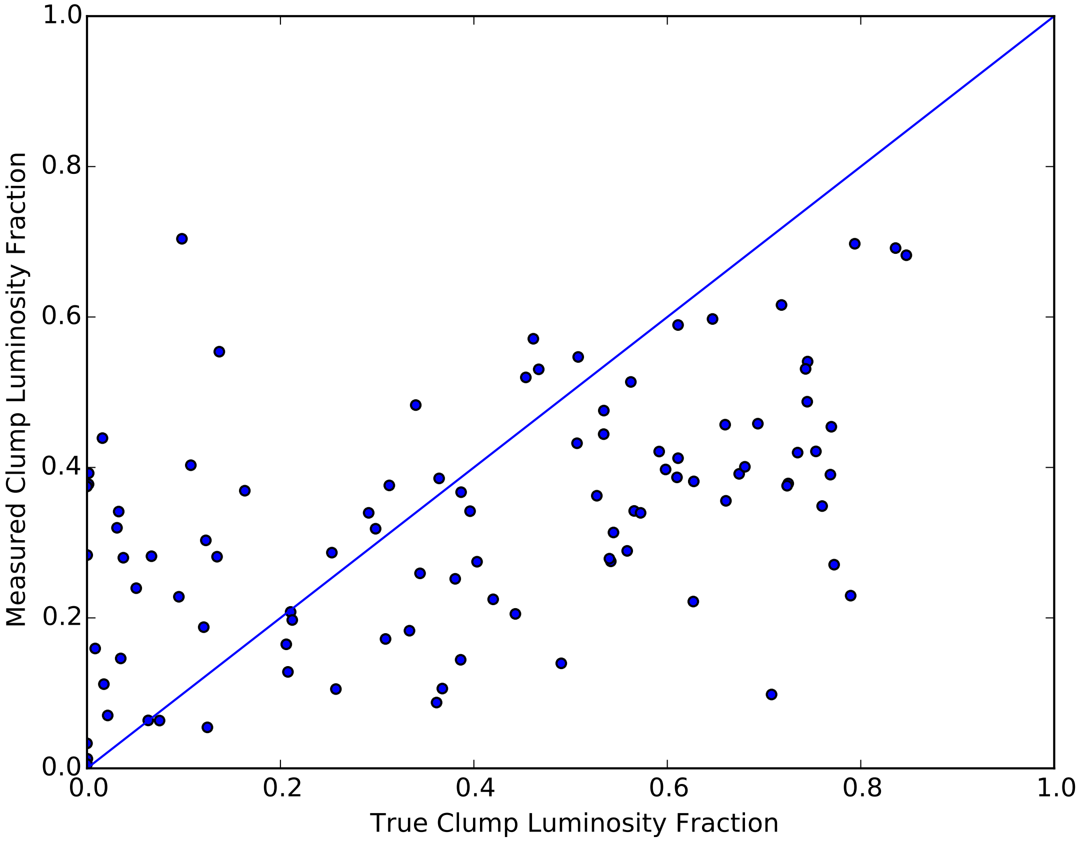


Results of the automated modeling process. Left: The simulated lensed images of a small sample of the sources. Middle: The model produced by the automated process. Right: The residual. Note: These images may have been rescaled to fit the page, so the pixel scale is not consistent across the sources. Also, patterned residuals near the corners of some of these images comes from correlated noise introduced by the straightening process.

correspond to smooth components. This can introduce a significant amount of uncertainty, but it is the last major step remaining after the modeling has been completed, but before the fluxes of the various components can be calculated and compared to the known values. There are slight correlations between the size of a component or the surface brightness of a component and its likelihood of being a smooth component, but neither of these alone provide a sufficient cut because of the variation in the intrinsic source properties. Similarly, there is helpful information contained in the positions of the components relative to the positions of the originally identified clumps, but because extra components of both types may have been added, this is does not provide a perfect criterion for classification either. Information is also available in the form of axis ratios (perhaps diffuse components have axis ratios more similar to the width to length ratio of the arc for highly sheared arcs) and position angles (should the smooth components align with the arc?). This is essentially all of the information available to use in the classification process. No single cut is sufficient to correctly classify these components (there are clear mistakes by eye no matter where they are cut, and histograms coded by type of component where components were classified by eye show significant overlap). However, it is not unreasonable that combinations of this information could be used in a cut, but not necessarily in a straightforward way (scatterplots of various pairs of these parameters and combinations and transformations of them—e.g., using FWHM, axis ratios, and magnitudes to get surface brightnesses—when color-coded by component type selected by eye, still show significant overlap).

One fairly simple cut that can allow for a first pass at calculating clump luminosity fractions that can be compared to the known input value for each source is to simply take the largest components, by area, and call them the diffuse components. Because the underlying galaxy light profiles—the diffuse components—are not actually Gaussians and are also extended enough that they still deviate significantly from Gaussians even after convolution with a Gaussian-like PSF, it may take multiple GALFIT components to fit the diffuse component of the galaxy. Similarly, when clumps are interspersed throughout the galaxy image

Figure 4.5: Clump Luminosity Fractions for Fully Automated Gaussian Model



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. This figure uses components determined and classified by a fully automated process. Each component is modeled by a Gaussian, and components are classified as relating to either clumps or smooth physical components by an automated process described in section 4.4.1.

or when there is notable shear in the image, different sections of the diffuse component will need to be fit with different Gaussians. As a result, the number of "large" Gaussians that are attributable to smooth components varies a little from object to object. If, for example, the three largest Gaussians by area are considered to be the smooth components, and the ratio of light contained in components designated as clumps (i.e., all other components) to the light contained in all components is calculated and compared to the known input values, the result is the scatterplot in Fig. 4.5.

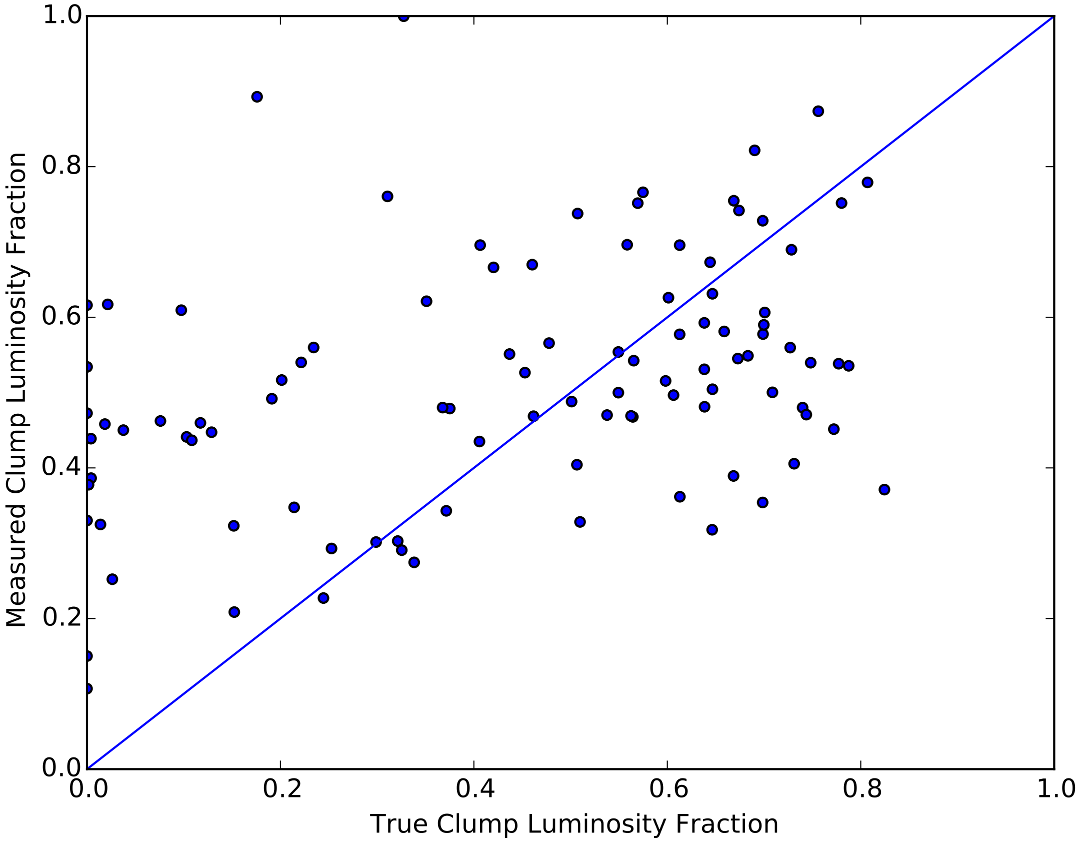
Perhaps somewhat surprisingly, even this simple cut shows that a correlation exists between the actual input clump luminosity fractions and the values measured in this way, albeit weak and with too shallow a slope, significant scatter, and notable clouds of outliers both above and below the 1-to-1 line. Still, though, it is suggestive that the information that we wish to extract remains mostly intact in the image plane, and that with more sophisticated techniques, it may be possible to extract it. Interestingly, when more highly sheared images are considered, a cut based on size, surface brightness, and axis ratio yields a similar result for a randomly selected sample of 100 such sources (Fig. 4.6). There is some correlation again, but significant scatter. Still, though, it suggests that even in image configurations with significant differential magnification, the conservation of surface brightness protects enough information to potentially measure the clump luminosity fractions of such objects in the image plane with a sufficiently sophisticated technique. This amounts to a strong argument that further refinements of this technique are worth considering. Therefore, it is necessary to explore the major sources of uncertainty in the existing technique. In doing so, three major potential sources of uncertainty immediately stand out.

Model uncertainty: Maybe the clump identification or the automated modeling are insufficient. The residuals are typically good, but are not perfect, and they can be quite bad in some cases. Furthermore, perhaps the clumps, the smooth components, or both, would be better modeled by another type of profile. After all, the profiles used for the clumps were Sersic profiles with n=2, and the profiles used for the smooth components were either exponential (Sersic with n=1) or deVaucouleurs(Sersic with n=4) profiles.

Classification uncertainty: Which model components correspond to which physical components? How can one best identify the GALFIT components that correspond to clumps in order to calculate the clump luminosity fraction?

Decomposition uncertainty: When the GALFIT model decomposes the light profile into various components, do these components correspond to real physical components (i.e., clump and smooth components), or can different physical components be blended into the

Figure 4.6: Clump Luminosity Fractions for Fully Automated Gaussian Model: High Shear Subsample



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 high-shear objects. This figure uses components determined and classified by a fully automated process. Each component is modeled by a Gaussian, and components are classified as relating to either clumps or smooth physical components by an automated process based only on size, surface brightness, and similarity of the axis ratio to the width to length ratio of the whole arc.

same model component? If this is the case, methods based on such a decomposition will be fundamentally limited in their ability to measure the clump luminosity fractions at all. And depending on the severity of this blending, such techniques could be rendered effectively useless, so it is critical to understand whether this is the case.

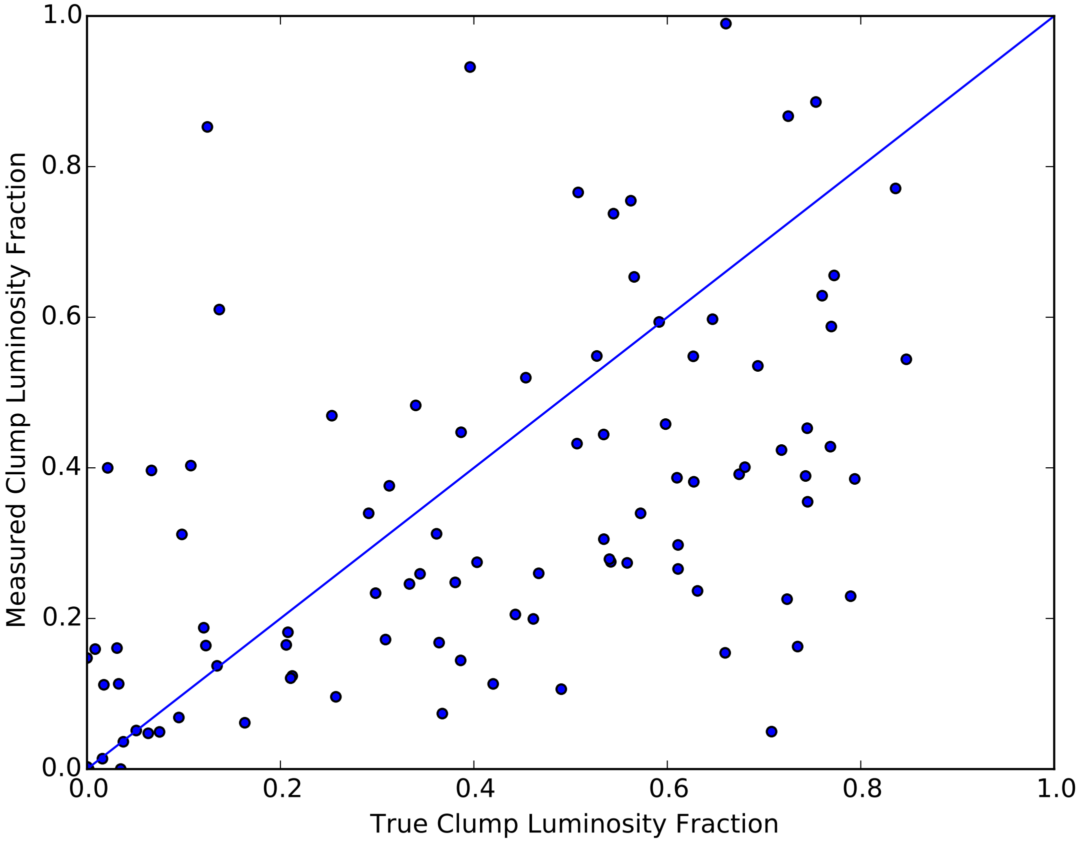
In the following subsections, I will explore the limits of Gaussian-based models, with an eye toward disentangling the effects introduced by these three types of uncertainties. Since Gaussians may not be sufficient, I will employ Sersic profiles in chapter 5 in order to further expand the capabilities of the model. This will allow an assessment of whether Gaussians are sufficient, but the information gained from more fully understanding model uncertainties will also make it possible to address whether classification uncertainties can be solved in an automated or semi-automated way and finally and conclusively address the issue of decomposition uncertainties.

4.4.2 Automated Gaussian models with by-eye classification

In order to address the issue of classification uncertainties, Prof. Gladders and I developed a tool designed to help visualize the various model components and select smooth or clumpy components by-eye. This visualization tool displayed each of the model components, and a three-panel window including the full model image, the image made by user-selected components, and the image consisting of the remaining components. After examining each of the images in the 100-image subsample and classifying the components in this manner, I calculated the resulting clump luminosity fractions and compared them to the true input values. Fig. 4.7 shows the resulting scatterplot.

By-eye classification in this manner, with these automated Gaussian decompositions, does not improve the quality of the measured clump luminosity fractions enough for this to be useful in practice. In particular, while there is some correlation, the bias toward underestimating the clump luminosity fraction in objects where the intrinsic clump luminosity fraction is above about 0.3 remains, and the scatter in such points is large. However, there

Figure 4.7: Clump Luminosity Fractions for Automated Gaussian Model with Manual Classification



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. This figure uses components determined by the automated Gaussian modeling process, but the components are classified as either clumps or diffuse components by eye.

are some notable improvements. For example, despite the cloud of poor measurements that lie below the 1-to-1 line, there are relatively fewer outliers that lie above the line compared to the fully automated case. Furthermore, at very low clump luminosity fractions (those below about 0.2), there is a marked improvement in the accuracy of the measured values. The reason for this improvement is two-fold. For one, at low clump fractions, any misidentification of a smooth component as a clump can drastically increase the inferred clump luminosity fraction. And due to the way the automated classification was conducted, if there are more than three smooth components (which is not unusual), this will necessarily happen. The second factor is that when multiple clumps are spatially close to each other, near the center of a galaxy, they can appear like one larger smooth component after they are convolved with the telescope PSF. In many cases, this can be discerned by eye, but as we shall see in chapter 5, it is by no means a simple task.

In any case, the improvement in classification is not sufficient to make this measurement method useful for real data, so we must turn our attention back to other sources of uncertainty, like model uncertainty. Will improving the models make this method more robust? And is it possible to do this by more carefully creating the Gaussian models (by, for example, doing them manually) or are more complicated profiles like Sersic profiles necessary? I consider these questions in the following subsection.

4.4.3 Hand-made Gaussian models with automated classification

In order to determine whether the clump identifications are insufficient or whether it is the automated modeling algorithm that is insufficient (or both), I repeated the analysis of the previous two sections using models that were created by hand. A summer REU student used GALFIT to model each of the 100 images, with the locations of the components chosen by hand. Residual maps show an improvement over the fully automated method. To compare with the fully automated version, where components were also selected automatically, I used the same simple classification algorithm on these GALFIT decompositions. Fig. 4.8 shows the relationship between the clump luminosity fractions measured in this way and the true values.

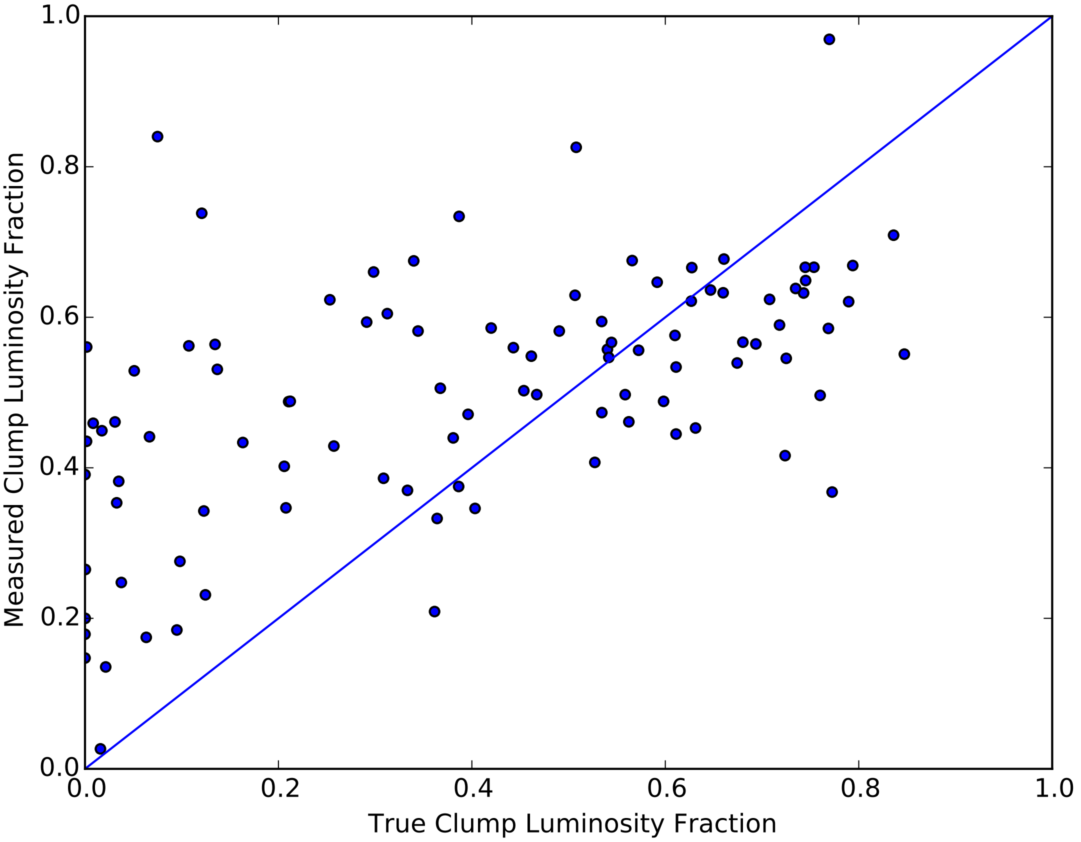
Compared to the fully automated approach, the correlation is still there and the scatter is perhaps a bit tighter, but the slope is still too shallow. Most notably, though, there are fewer outliers below the line, suggesting that more detailed models than the automated ones are necessary to properly model the light distributions in these images. Comparing this figure to Fig. 4.7, where the components for the automated models were classified by eye, it is apparent that improving the classification techniques will be important at least for objects with low intrinsic clump luminosity fractions. The automated cuts still tend to overestimate the values for those objects even when the models are created manually, despite the fact that in the automated version, classification by eye removed most such discrepancies. The obvious question now, though, is whether the cuts developed for the automated version are still appropriate for the models created manually. Furthermore, since the by-eye classifications already led to improvements in the measured values in the automated modeling case, we should expect that classification by eye would improve the values in the hand-modeled case as well, especially because it is already clear that the classification criteria should be much more complex than the simple cut being used. To truly test the limits of what a modeling method that uses Gaussians can do, we should examine a version of this method done entirely by hand—that is, when the models are created by hand and the classification is done by eye.

4.4.4 Hand made Gaussian models with by-eye classification

To truly test the effect of using a higher quality model, the hand made model should have component classification done by hand, as in section 4.4.2, and the resulting figure should be compared to Fig. 4.7. This should mostly isolate the effect of changing the modeling strategy, though some subjectivity will remain in the classification process. Fig. 4.9 shows the result of performing classification by eye after creating the models by hand.

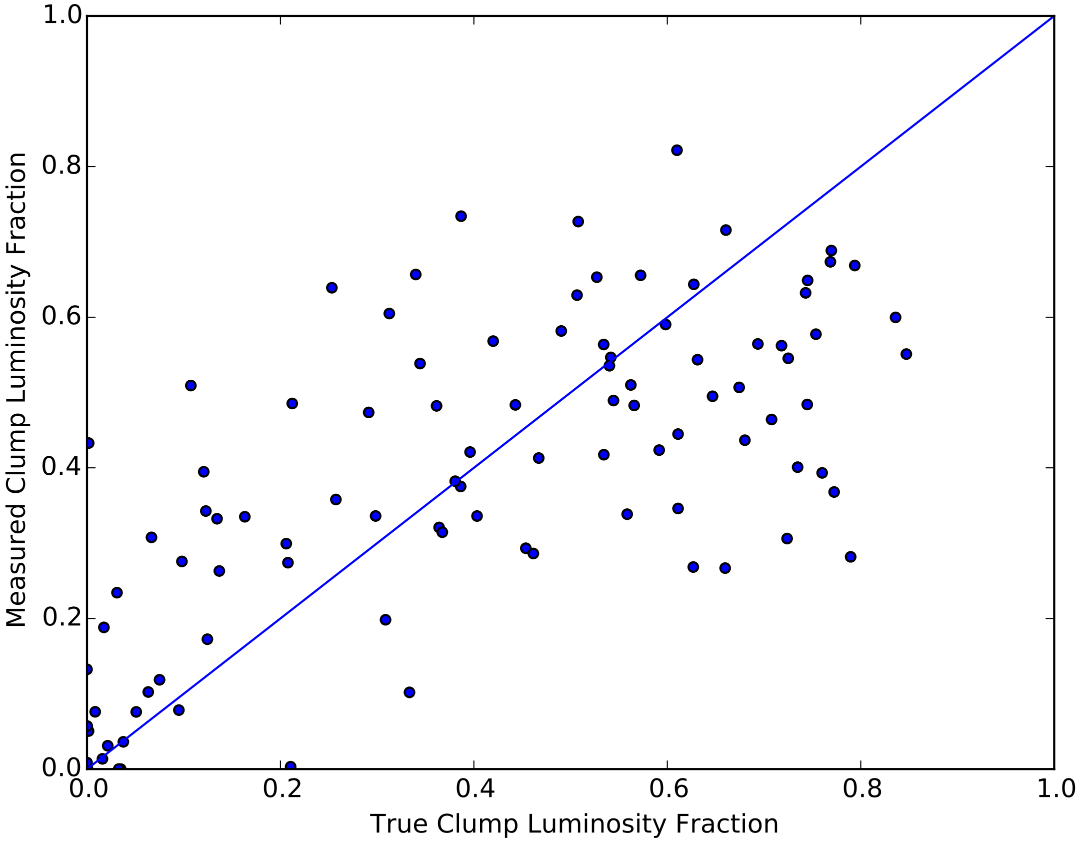
In this case, most of the outliers on both sides have disappeared and the slope more closely

Figure 4.8: Clump Luminosity Fractions for Manual Gaussian Model with Automatic Classification



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. The decompositions come from manual fitting using GALFIT with Gaussian components. Components were classified as either relating to clumps or smooth components by the same automated process as used for Fig. 4.5.

Figure 4.9: Clump Luminosity Fractions for Manual Gaussian Model with Manual Classification



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. This figure uses components determined by the hand-made Gaussian modeling process, and the components are classified as either clumps or diffuse components by eye.

matches the 1-to-1 line. However, the scatter still remains large. Clearly the more refined models have helped, but because of the large scatter that still remains, it is reasonable to ask whether a profile other than Gaussians could further improve the model. We have seen that the information necessary to make clump luminosity fraction measurements appears to still exist in the image plane, but we may be reaching the limits of the utility of a Gaussian-based model and several questions still remain. For example, can the classification be refined any further? But most importantly, does the scatter that remains represent a limitation due to using Gaussians, a limitation due to the classification process (and can a by-eye process be improved upon by using some other computational method), or is this a fundamental limitation due to decomposition uncertainties?

In chapter 5, I will further refine these models to use Sersic indices, explore different classification methods, and finally answer the question of whether the scatter in these figures is driven by things that can be improved like models or classifications, or whether it is driven by a fundamental limit on precision set by the decomposition uncertainty.

CHAPTER 5

EXTENSION OF THE CLUMP MODELING TECHNIQUE: SERSIC PROFILES, MORE CLASSIFICATION METHODS, AND THE DECOMPOSITION UNCERTAINTY

In the previous chapter, I developed an extremely large sample of realistic simulated lensed sources with known clump properties in order to test an automated method for measuring the clump luminosity fraction of individual strongly lensed galaxies in the image plane. As is evident from Fig. 4.5, there is—to put it mildly—a lot of room for improvement. However, Fig. 4.9 shows that overall, the basic method of using GALFIT to decompose lensed images into clump and smooth components has promise, but the implementation must be far more sophisticated than my attempted automated process that uses Gaussians.

As explained in section 4.4.1, there are three major types of uncertainty that may be driving the scatter and biases in the results so far. One of them was uncertainty coming from the model—are the residuals too large? Is a Gaussian profile sufficient to model these complicated structures? The second was the classification uncertainty. After GALFIT has decomposed the image into a set of Gaussians (or whatever profile is chosen), how does one tell—with confidence—which ones belong to clumps and which ones belong to smooth components? And lastly, there was the decomposition uncertainty—do the GALFIT components even correspond to physical components (i.e., clumps and diffuse components) at all? And is this related to the modeling uncertainty (i.e., would a more detailed model profile better separate these physical components)?

In the previous chapter, I tested the limits of what a model based on Gaussian profiles could do, even when the modeling was carefully conducted by hand, and the classifications were done manually and found that Gaussian profiles seemed to be insufficient for producing robust estimates of the clump luminosity fractions for these sources. This suggests that either the simplicity of Gaussians is driving the scatter, or that the decomposition uncertainty is.

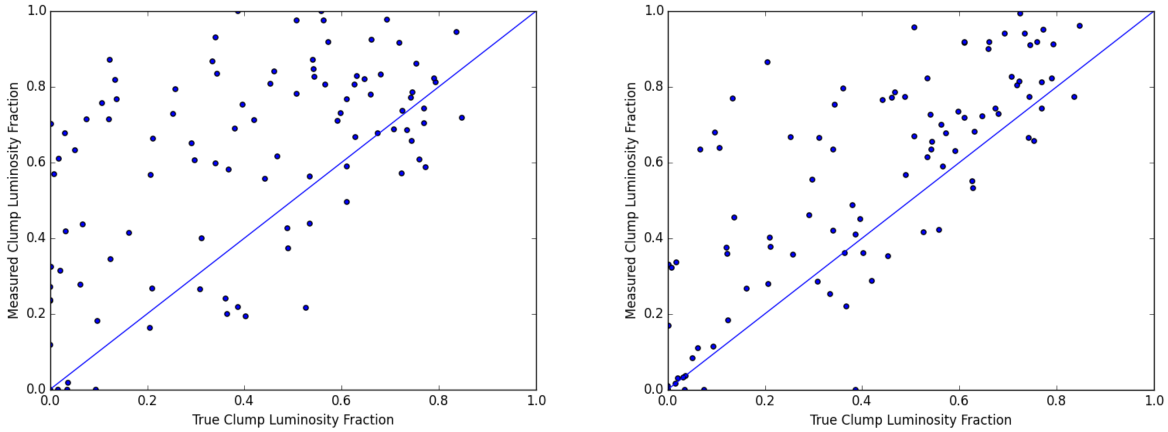
In this chapter, I will further explore the modeling uncertainty by using hand-crafted models based on Sersic profiles rather than Gaussians. As I will show, these more flexible profiles produce better results and allow a more serious investigation of the classification and decomposition uncertainties. I will also show that at least in the low-shear case, the classification uncertainty will be the main driver of scatter, and not the decomposition uncertainty, and I will suggest some avenues by which future work can further decrease the classification uncertainty.

5.1 Manual Sersic-based models, with automated classification: A first attempt

The first test of a Sersic-based modeling method was a simple one. The same 100 images used in the tests in the last section of chapter 4 were modeled again, this time using Sersic profiles instead of Gaussians. These were done manually, using GALFIT but with no pre-made clump list or algorithm for placing clumps. Positions were selected by hand, and components that by-eye seemed to be clumps were given a Sersic index of 2. For diffuse components, the Sersic index was allowed to vary. Fig. 5.1 shows the resulting clump luminosity fractions plotted against the true values for each source using two different automated classification methods. In the left panel, the classification was done using the same criteria as in the fully automated Gaussian case. On the right, only components with Sersic index between 1.9 and 2.1 were classified as clumps and everything else was determined to be a diffuse component. There are some striking differences between these and the analogous figure for the hand-made Gaussian models with automated component classification (Fig. 4.8).

For both of these, despite the scatter, the slope better approximates the 1-to-1 line than when manual Gaussian modeling was done and automated cuts were made. In fact, even compared to the Fig. 4.9, where the classifications for the manual Gaussian models were done by hand, the measured values at high intrinsic clump luminosity fractions are more

Figure 5.1: Clump Luminosity Fractions for Manual Sersic Model with Automated Classification



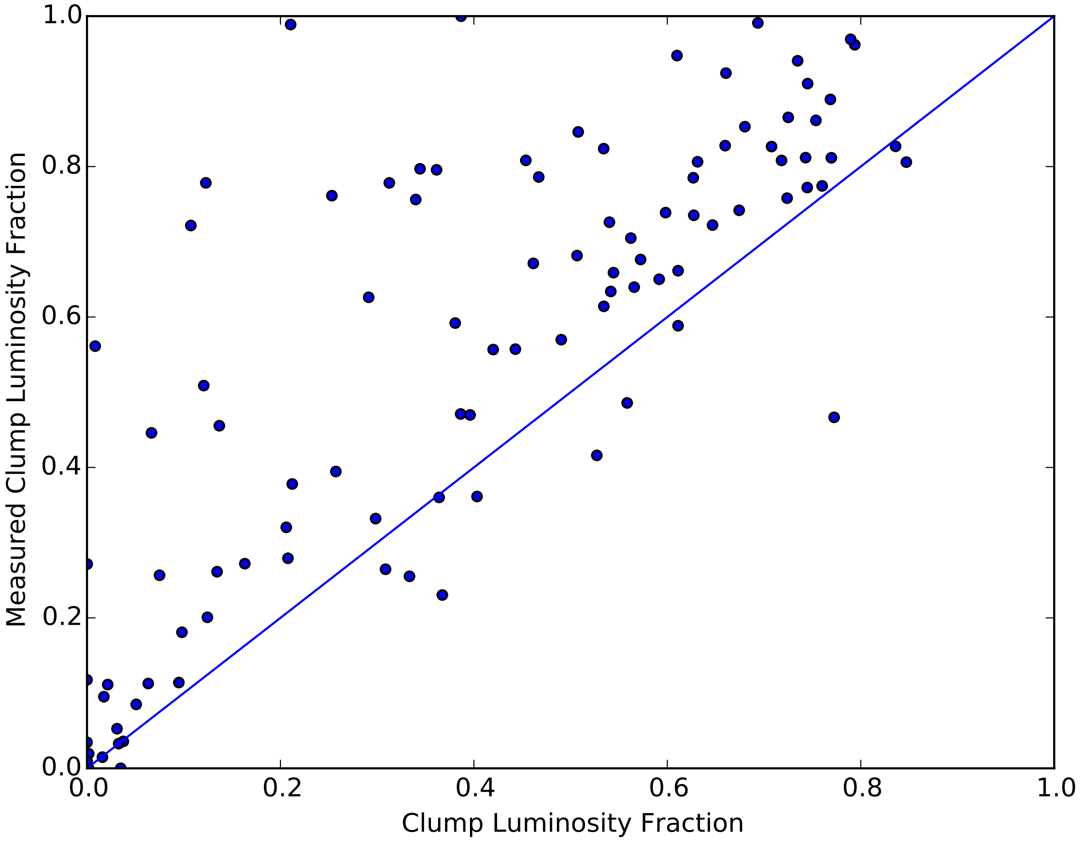
The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. For both plots, the decompositions come from the same manual fitting attempt using Sersic indices with n=2 for components identified as clumps by the modeler, and all other components are allowed to vary. On the left, the same criteria for automated component classification as was used for the Gaussian models in chapter 4 was employed. On the right, the only components with Sersic indices between 1.9 and 2.1 were classified as clumps, while the rest were assumed to be smooth components.

accurate. This suggests that perhaps the Sersic models are doing better job of mitigating the decomposition problem discussed in section 4.4.1 than the Gaussian does. Since Sersic profiles are less rigidly defined, they can be tweaked via the Sersic index to better match the wings (and perhaps the centers) of the various components. This could lead to reduced residuals, but does not necessarily have to. If it does, then one might expect more accurate estimates of the clump luminosity fraction than in the Gaussian case. But even if both Gaussians and Sersics perform similarly well in terms of modeling the overall light distribution, the more flexible profile—the Sersic profile—may do a better job of separating the physical components into model components.

Despite the strengths of this new version of the modeling, there are still some clear deficiencies. In both cases, there is still a large amount of scatter, particularly above the 1-to-1 line. In the left panel, the large scatter is not particularly surprising. The classification criteria used for that figure were developed based on models that used Gaussians and it is reasonable to think that the criteria might need to change when such an important piece of the modeling process changes. This is supported by the figure on the right, where the scatter is reduced due to a cut that comes mostly from human decisions during the modeling process. There is, however, still a tendency to overestimate the clump fractions in both panels. Is this because of the decisions made by the modeler, or is this because the model itself—despite being more general than in the Gaussian case—is still insufficient? The former possibility can be addressed by doing a manual re-classification, which I shall present in the next section. The latter can be addressed by further generalizing the model and allowing the Sersic indices of all of the components to vary. This will be done in the following section, 5.3.

While the analysis in this section has begun to hint at solutions to the question of decomposition uncertainty by raising the possibility that the more generalized models are able to better separate the components, the true impact of the decomposition uncertainties cannot be convincingly ascertained until model uncertainties have been fully ruled out. The

Figure 5.2: Clump Luminosity Fractions for Manual Sersic Model with Manual Classification



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. The decompositions come from the same manual Sersic-based fitting attempt as in the previous section. In this case, components were classified by eye after modeling.

work presented in the next two sections will help accomplish this.

5.2 Manual Sersic models with manual classification

In order to determine whether the scatter in the plots in section 5.1 is driven by model uncertainties or classification uncertainties, the components were classified by hand using the process described in section 4.4.2. Fig. 5.2 shows the impact of this classification strategy on the measured clump luminosity fractions.

Interestingly, there is very little change between the automated classification using Sersic index and the manual classification. This is not terribly surprising because there was some human intervention in the classification during the modeling process in that the modeler chose to enforce a Sersic index of 2 on components thought to be clumps. But because the modeling and the classification were done by two different researchers (Prof. Gladders did the modeling for this sample and I did the component classification), the bias in the classification should be different than the bias in the modeling. However, the classifications seem to agree quite well with the modeler's decisions anyway. This suggests that this is the limit that can be reached with human by-eye component classification with this type of model. Perhaps a more sophisticated classification process that includes all of the information available from the GALFIT parameters, maybe one based on machine learning if a suitably pure training set could be produced, might be better. But perhaps, instead, this is a limit imposed not by the human classification (since a simple automated classification is about as good), but rather by the model. If the model can be further generalized and still reduces the scatter or removes the bias toward overestimation, then there is still room to improve. I test this in the following section.

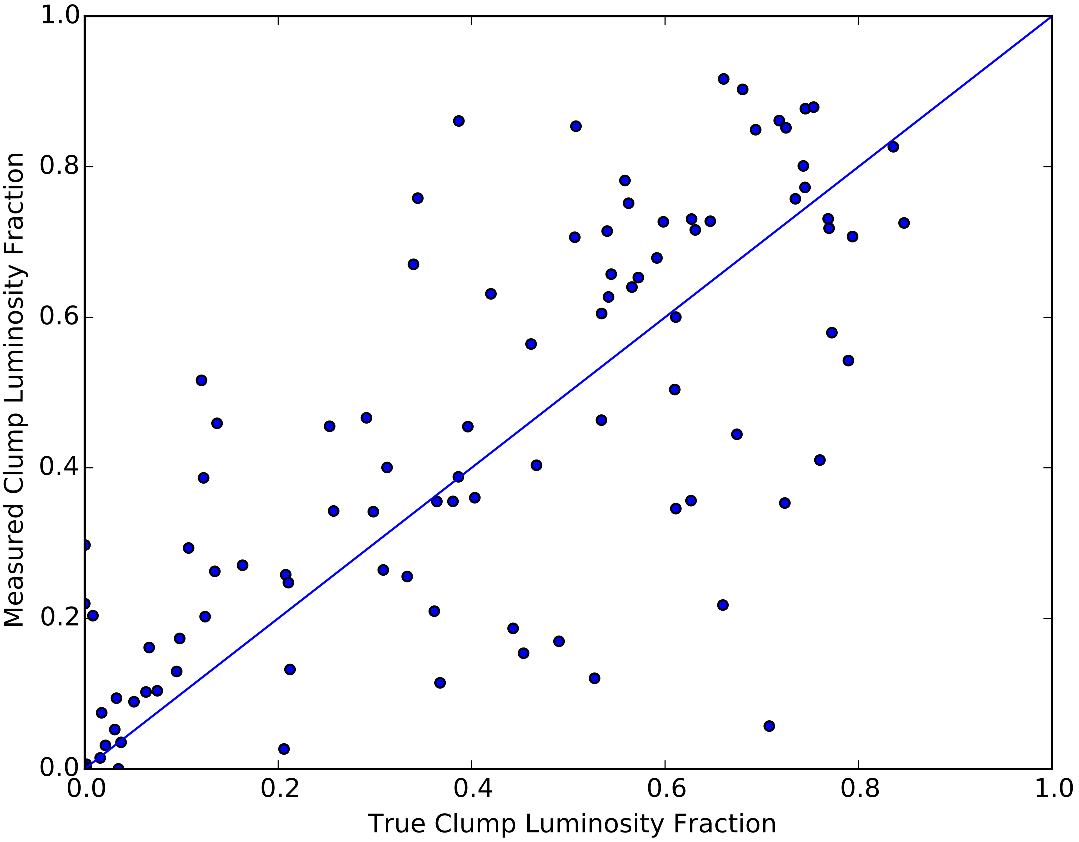
5.3 A more generalized Sersic model

One potential source of uncertainty in the previous analysis is the modeler's selection of what is and is not a clump, by eye. By not allowing the Sersic indices to vary for the components selected to be clumps, the modeler locks in a classification and potential decomposition problem during the modeling stage. This is particularly worrisome because, during the modeling process, it can be difficult to distinguish between a clump and the central peak of the light distribution in, for example, an elliptical galaxy. This becomes especially difficult when there are several clumps close together. When they are convolved with a telescope PSF, it can be hard to tell not only whether they are clumps or the center of a larger smooth component, but also whether it is one or many clumps and how many clumps there are. And in the worst case, it may be several clumps *and* the peak of a smooth component since these sources were designed so that the likelihood of finding a clump would be highest near the center. Furthermore, when there is shearing—and despite the selection for low shear in these 100 objects, there will still be some noticeable shear—even a clump whose Sersic index was 2 in the source plane may be better modeled by a component with a different Sersic index (or multiple components with different Sersic indices) in the image plane.

To avoid guaranteeing ourselves a classification uncertainty, and to allow better fitting of sheared components, we can simply let the Sersic indices of all of the modeler-selected components vary. This way, if a clump has been misidentified, or if it has been significantly sheared, or even if it was just small enough that PSF-convolution has distorted its profile significantly (i.e., made it more Gaussian), the model can better adjust to fit the resulting light distribution than it could when the Sersic indices of clumps were locked at n=2. In this section, I present results of a model where all of the components were allowed to vary. The locations of components were still set by the modeling effort in section 5.1, but the Sersic indices of modeler-identified clump components can vary away from n=2 and the other components can vary accordingly to compensate. This is the most general possible way of modeling using Sersic indices which are, themselves, one of the most general but realistic parametric light profiles. One could go back and re-model the sources using these more general profiles instead of using the previously determined locations and it might further improve the ability of the model to reproduce the lensed images, but I do not expect such an improvement to be substantial.

In Fig. 5.3, I present the results of modeling using Sersic profiles with varying indices and classifying the components as clumps or smooth components by eye after the modeling. The various biases apparent in all previous such figures nearly vanishes. There remains noticeable scatter, but there is no obvious systematic tendency to overestimate (as in Fig. 4.8, Fig. 5.1, and Fig. 5.2) or to underestimate (as in Fig. 4.7), though there might be a very slight bias toward overestimation (in section 5.5, once further refinements have been implemented, I will more thoroughly consider and quantify this bias). There is also no clear trend toward

Figure 5.3: Clump Luminosity Fractions for Most General Manual Sersic Model with Manual Classification



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. The decompositions come from the a manual Sersic-based model where the Sersic indices of all components were allowed to vary. In this case, components were classified by eye after modeling.

decreased accuracy at the extremes (e.g., major overestimates at low intrinsic clump luminosity fractions as in Fig. 4.5 or major underestimates at high intrinsic clump luminosity fractions as in Fig. 4.9).

Despite the relative successes of this modeling technique, there still exists some significant scatter as well as outliers on both sides of the 1-to-1 line suggestive of some type of misclassification or decomposition issue. With this most general Sersic-based model, I am beginning to reach the limits of modeling complexity for any method based around GALFIT. Therefore, it is worthwhile to ask whether further improvements to classification might be possible using more sophisticated algorithms that explicitly incorporate more information than simply looking at the components by eye or whether we have reached a fundamental limit in the efficacy of this type of modeling and measurement method due to the decomposition uncertainty imposed by models of even this level of complexity. Fortunately, as I will explain in the next section, it is possible to use maps of lensed clump positions to determine how well a method like this could do in the ideal case, where classifications are as accurate as possible (whether such accuracy is realistically attainable or not), thereby revealing the level of scatter introduced by the decomposition uncertainty. From this we can also gain insights into how to improve classification methods, how to incorporate other potential observations in order to do so, and to what level of accuracy and precision it is possible to measure a strongly lensed galaxy's intrinsic clump luminosity fraction in the image plane.

5.4 An attempt to quantify the decomposition uncertainty

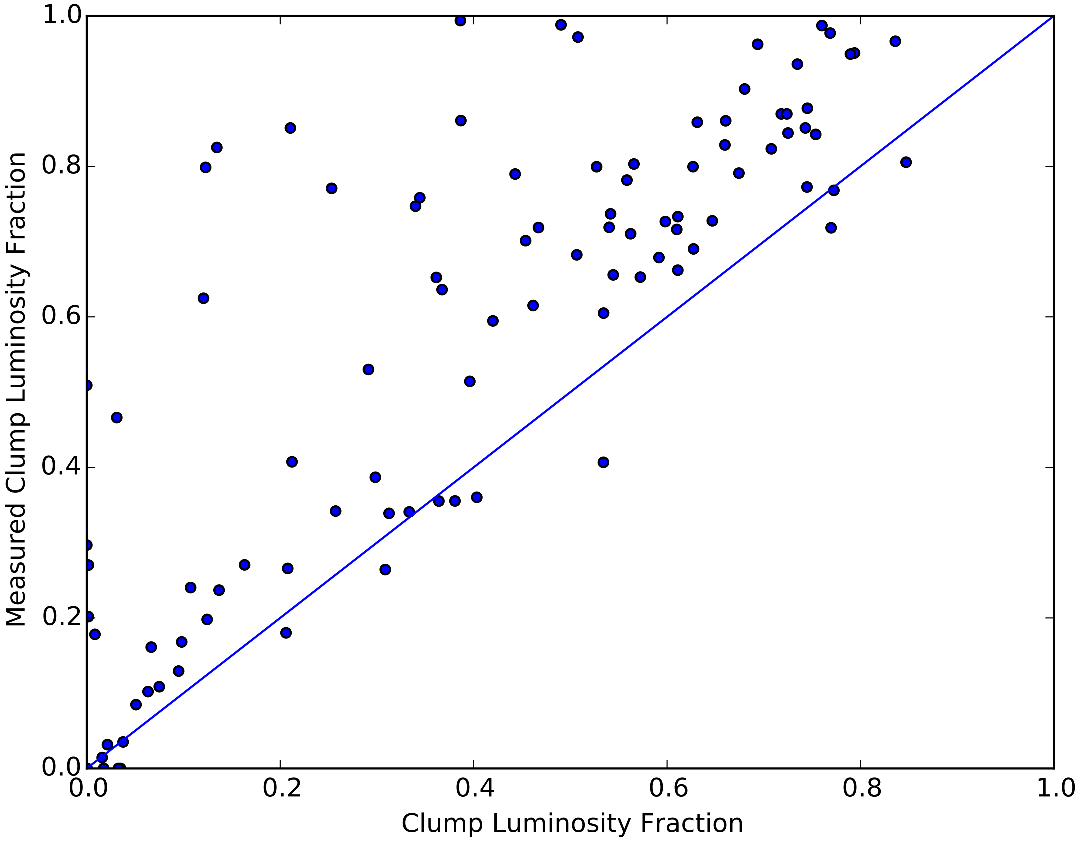
An extra piece of information that has not been brought into this analysis up to this point is the known input locations of the clumps. I have ignored this information so far because in real data, it would not be available and identification of clumps or classification of clump and smooth components would be limited—as it has been so far in this work—to the abilities of the researcher or of the modeling, classification, and measurement algorithms to discern between clumps and smooth components in data similar to the simulated images described in chapter 4. However, to further quantify the significance of the classification uncertainties and to be able to make definitive remarks about the decomposition uncertainty, it is useful to bring in this knowledge.

When the model galaxies were created by the process described in chapter 4, the locations and sizes of clumps in the source plane were also recorded. A map of the locations of each clump was created in the following manner. For each clump, all pixels within one halflight radius of the center of the clump were marked as part of the clump. A larger radius could reasonably have been used, but localization of each clump was more important than mapping out all light related to each clump, so it was unnecessary to use a larger radius. The one caveat to this is that a minimum radius of 4 pixels (20 parsecs) was imposed in order to make sure that after lensing and rebinning to a coarser grid, the clump would still be present in the clump map in at least one pixel. The clumps were labeled by using a bit mask, so that the first clump in the parameter list for each source (i.e., the list of clumps fed into the iraf artdata mkobjects function) was marked by taking all pixels within the halflight radius of that clump, or at least within 4 pixels of that clump, and assigning a value of 1 to its first bit. The pixels belonging to the second clump would get a 1 in the second bit, and so on. Mapping the clumps in this manner allows pixels where clumps overlap to be marked in such a way that all clumps overlapping at that pixel can be identified, but minimizes the storage space needed for these clump maps.

Using the fully general Sersic models, I classified the components by hand, comparing the locations of the model components to an overlaid clump map. This significantly reduces the classification uncertainties, but it still does not eliminate them entirely. Because multiple components can sometimes be required to model a single clump, it becomes difficult to identify clump components in regions where overlap between clumps occurs, especially if such a region also overlaps the center of the underlying smooth light profile of the source, which is often the case. Still, this by-eye identification, with the aid of clump maps is quite good at reducing the scatter caused by classification as is demonstrated in Fig. 5.4.

Almost all of the outliers below the 1-to-1 line disappear. These outliers are underestimates of the clump luminosity fraction and are typically caused by misidentifying clump components as being smooth. Based on my experience of classifying these by eye, the most common reason for these misidentifications is that several overlapping clumps near the center of a source become confused with the center of the source itself. The clump maps largely help resolve this issue. In doing so, however, it causes the opposite issue of accidentally misidentifying a smooth component that overlaps a blend of clump components. The knowledge that one or more clumps exists in a certain location biases the researcher toward classifying

Figure 5.4: Clump Luminosity Fractions for Most General Manual Sersic Model with Manual Classification Using Clump Maps



The measured clump luminosity fractions plotted against the true clump luminosity fractions for a sample of 100 low-shear objects. The decompositions come from the a manual Sersic-based model where the Sersic indices of all components were allowed to vary. In this case, components were classified by eye after modeling, using maps of known clump locations to minimize the classification uncertainty and begin to probe the decomposition uncertainty.

components in that location as clumps, resulting in a systematic overestimate of the clump luminosity fraction. This manifests as a tendency for points in Fig. 5.4 to lie just above the 1-to-1 line. Interestingly, when the components were classified by eye without these maps, the bias was noticeably less pronounced. This further suggests that the maps, while helpful in the underestimated cases, bias by-eye classification methods toward identifying smooth components as clumps too often.

Similarly there are several outliers on the high side—extreme overestimates of the clump luminosity fraction—that were not as severe or were not outliers at all when the clumps were identified by eye without the clump maps in Fig. 5.3, further suggesting that much of this remaining scatter above the 1-to-1 line is driven by the researcher's bias. A second review of several or these outliers shows that changing the identification of a single component from being a clump to being smooth can bring the measured value into much closer alignment with the 1-to-1 line. This, combined with the fact that most of the other points lie relatively close to a straight line (albeit one with a slightly steeper slope than the ideal 1-to-1 line) suggests that the decomposition uncertainty is actually relatively small, and that further refinements of the classification process may be able to bring the remaining outliers back toward the line. This seems to be the limit on what a by-eye classification can do with this type of model, so the next step is to bring more of the quantitative GALFIT parameters back into the picture and attempt to use either a more refined automated classification system or a hybrid system that involves both human and automated facets.

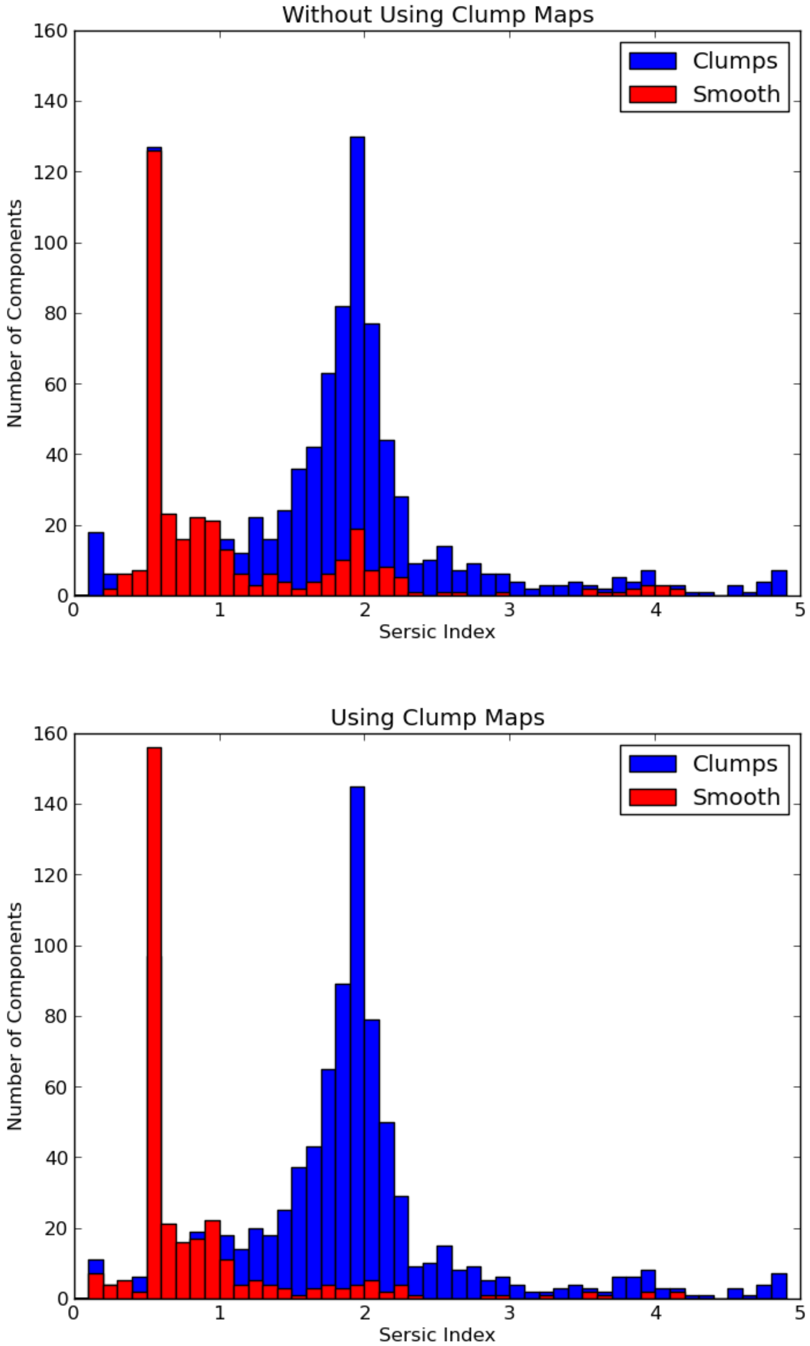
5.5 Toward an automated classification process and a final conclusion on the decomposition uncertainty

In addition to the parameters that were available in the Gaussian-based models, namely the positions, magnitudes, FWHMs, axis ratios, and position angles, the generalization to Sersic profiles also provides an additional parameter, the Sersic index, which can be used in an automated classification scheme. Because the Sersic index of clump and smooth components is expected to differ, might it be possible to solve the problem of misidentifying overlapping PSF-convolved clumps with smooth components by using the Sersic index? Fig. 5.4 suggests that these components are indeed well-separated (otherwise we would expect to see more underestimates and outliers below the 1-to-1 line), but Fig. 5.3 suggests that it is difficult to make this identification by eye. After all, it is not particularly easy to distinguish slight differences in the index of a Sersic profile by eye, especially when magnitudes, sizes, and axis ratios are also varying.

To see whether this is reasonable, I plotted the histogram of Sersic indices for clump components and smooth components in two cases: First, using the by-eye classification without clump maps, and second using the by-eye classification with clump maps. The assumption is that classification will be more accurate using the clump maps, even if it is not perfect. Therefore, any systematic change in the identifications based on Sersic index would correlate with actual differences in the light profiles of different physical components in the image plane. And the exact nature of these changes would point to what type of cut might need to be made to separate the clump components from the smooth components. The result is shown in Fig. 5.5, where clump components are shown in blue and smooth components are shown in red. The top panel shows the histograms for the case where the clump maps were not used, while the bottom panel contains the histograms for the case where they were.

The most obvious feature of Fig. 5.5 is that there is a clear reduction in the number of model components with Sersic indices near 2 that are identified as smooth components when the clump masks are taken into consideration. An easy question to ask, then, is: What happens if the classification is done with a simple automated cut, where anything near 2 is considered to be a clump? Fig. 5.6 shows what happens when objects are classified purely based on this single criterion. A component was considered to be a clump if its Sersic index was greater than 1.3 or less than 2.5 in order to capture most of the blue peak in Fig. 5.5.

Figure 5.5: Distribution of Sersic Indices for Clump and Smooth Components as Classified with and without Clump Maps



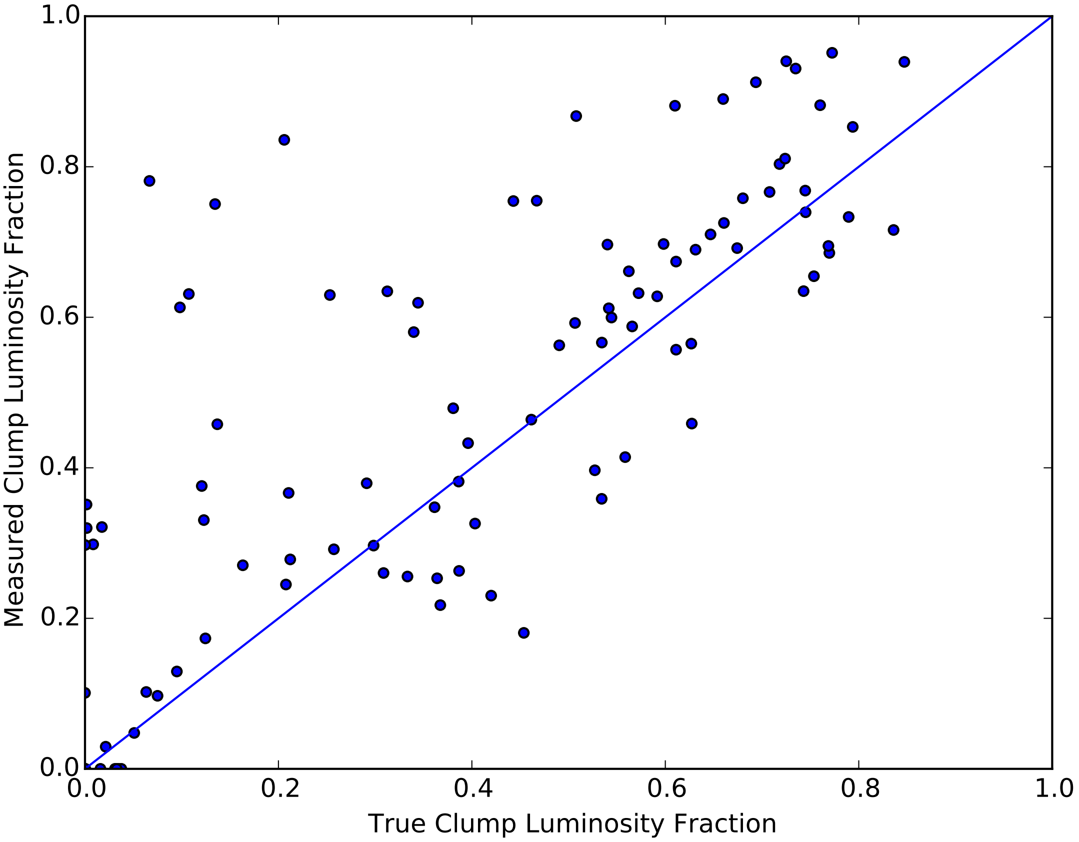
The distribution of Sersic indices for components identified as clumps (blue) and smooth components (red). *Top*: Distributions when components are classified by eye without use of clump maps to guide the classifications. *Bottom*: Distributions when components are classified by eye while using the known locations of clumps, via the clump maps, to guide the classification. Using clump maps drastically reduces the number of components with Sersic indices near 2 that are identified as being smooth components.

Otherwise, it was classified as a smooth component.

For the most part, such a cut does indeed help solve the misclassification problem in the cases where clumps near the center of a smooth component get confused with the smooth component, eliminating many of the lower outliers from the by-eye classification in Fig. 5.3 and suggesting that the incorporation of this extra information can improve upon what is possible with a simple by-eye classification method. However, scatter to the upper side of the line is increased. Specifically, at lower intrinsic clump luminosity fractions, there are several quite large overestimates that were not present when classification was done by eye for Fig. 5.3. This suggests that further improvement over this simple automated cut is indeed possible, either by combining the automated classification, with classification by eye in particular cases, or by using a more sophisticated automated classification method. But what would the results of such a method look like? How well could we expect to do? In the case of a hybrid method, the ideal case would be that when the human and automated classifications disagree, we manage to choose the more correct identification. The potential efficacy of such a hybrid method could therefore be approximated by taking the measured clump luminosity fractions from the automated classification process in Fig. 5.6 and from the by-eye classification process in Fig. 5.3 and plotting the result that more closely matches the true intrinsic clump luminosity fraction. This should provide an optimistic but realistic idea of how well a hybrid process could work for classification of these components. Fig. 5.7 shows the result of such an analysis.

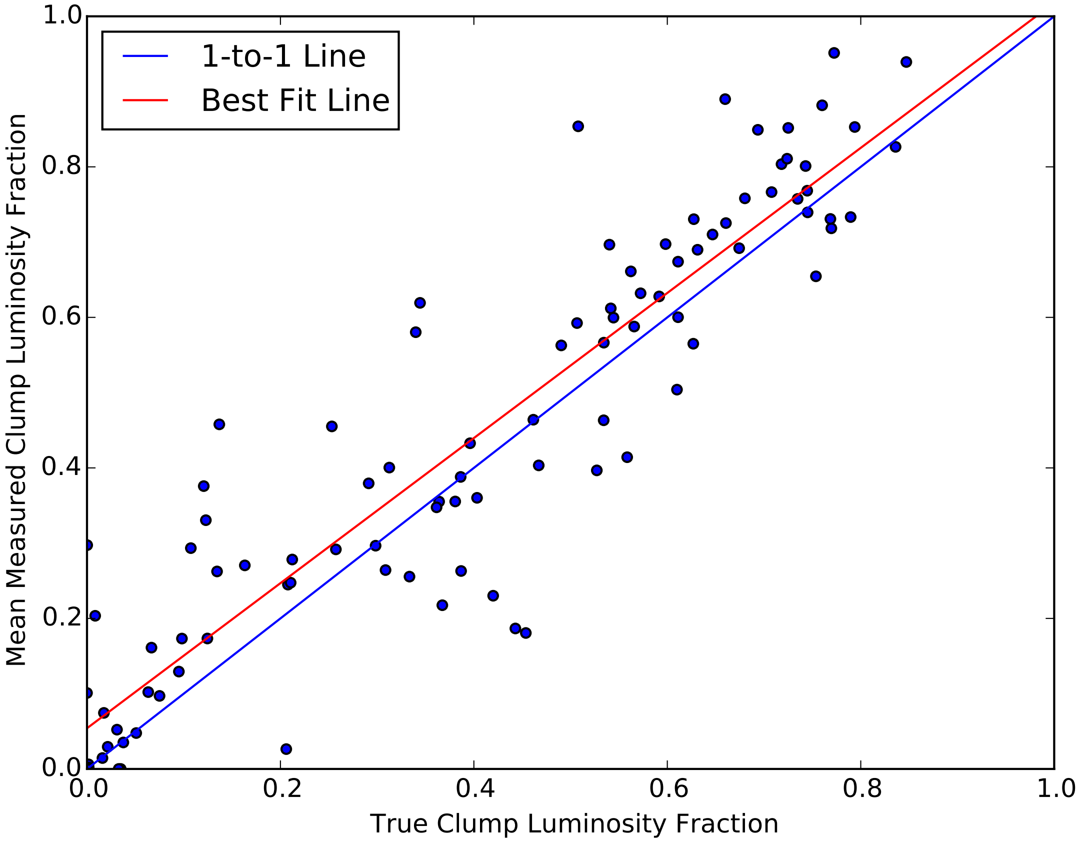
In this optimistic approximation of a hybrid method, the scatter is greatly reduced. The by-eye classification has eliminated the major outliers on the high end that remained after the automated classification, while the automated classification has eliminated most of the outliers underneath the 1-to-1 line that were present after by-eye classification. The best fit line for this data is slightly offset from the 1-to-1 line. The mean of the residuals is about 0.04, indicating a slight preference for overestimation of the clump luminosity fraction (perhaps a result of decomposition uncertainty in crowded parts of the image). The slope

Figure 5.6: Clump Luminosity Fractions for Most General Manual Sersic Model with Classification Based on Sersic Index



The measured clump luminosity fractions plotted against the true clump luminosity fractions for the same sample of 100 low shear objects, with the same models as in Fig. 5.3. In this case, though, classification was done based solely on whether the Sersic index of a component was greater than 1.3 or less than 2.5, in which case it was classified as a clump. If the Sersic index fell outside of that range, it was classified as smooth.

Figure 5.7: Clump Luminosity Fractions for Most General Manual Sersic Model with Idealized Hybrid Classification Method



The measured clump luminosity fractions plotted against the true clump luminosity fractions for the same sample of 100 low shear objects, with the manual Sersic models with the Sersic indices of all components allowed to vary. This figure represents the sort of result that could optimistically but realistically be obtained given hybrid classification method that combines automated cuts with by-eye classification.

of the best fit line, however, is 0.96, running almost perfectly parallel to the 1-to-1 line suggesting that there is very little bias as a function of intrinsic clump luminosity fraction. The standard deviation of the residuals is only 0.11, suggesting that the decomposition uncertainty is no larger than this, and that the major driver of the scatter in the previous figures was actually the classification uncertainty. This means that this type of method can indeed work and that it is probably not even necessary to use a model more sophisticated than Sersic profiles as long as a more sophisticated classification method can be developed. Furthermore, the fact that uncertainties introduced by model decomposition issues are so low validates the approach taken in Johnson et al. 2017a for SGAS1110+6459, though there may be an argument for re-analysis using Sersic profiles instead of Gaussians. On the other hand, the analysis for SGAS1110+6459 was not done in the image plane. A much more detailed analysis, involving source plane modeling and propagating the GALFIT model through the lens model, was applied in this case, which may well have resulted in more robust results. It would be interesting, though, to do a re-analysis of this object in the image plane using Sersic indices by hand to see whether the inferred clump luminosity fraction agrees with our collaboration's previous work, especially as the image-plane method becomes more refined.

I have used Fig. 5.7 to argue for a hybrid classification method, but how would such a hybrid method work in practice? Since it is already possible to use the classification tool described in chapter 4 to classify the components of approximately 100 images in an hour, one could imagine a process by which components were identified by eye, but difficult classifications of components near the center of a source could be flagged and determined by a cut on the Sersic index and potentially other properties without unreasonably slowing the pace of the classification. However, a hybrid classification scheme is not the only potential path forward.

Because of the large number of pieces of information available with which to attempt to make classifications and the complex relationships between those parameters and whether a component is a clump or a smooth component, a machine learning approach may be more appropriate. A training set could be determined in a way similar to the analysis that led to Fig. 5.7. If such a machine learning method could work, it would eliminate the need for human intervention in the classification process, speed up the analysis (which will be helpful when large samples of strongly lensed sources become available in the coming years from new observing resources like WFIRST), and potentially provide more reliable classifications than a human researcher could alone. Additionally, to truly quantify the uncertainties involved in using a method like this requires testing on a large dataset with known clump luminosity fractions and the widest possible range of realistic source images, like the dataset I developed and described in chapter 4. And being able to perform these calculations on such a large dataset requires an automated approach. In other words, regardless of whether a hybrid approach is efficient enough to be applied to real observational datasets, the validation of the approach requires automation.

5.6 Concluding Remarks and Outlook

This chapter and the preceding one have largely been a proof of concept of this technique. There is clearly a lot of work left to do and there are a few immediate questions that deserve to be addressed before work continues. For instance, will this method even work on other image classes?

In Fig. 4.6, I showed a scatter plot of measured vs. intrinsic clump luminosity fractions for highly sheared objects (in cases excluding merging pairs where significant portions of the source do not appear in the image plane—i.e., those with low image fractions as described in chapter 2). If there is a case where the lensing configuration provides large image fractions, but for which the clump luminosity fraction cannot theoretically be recovered without source reconstruction, one would expect it to be objects with high shear and high differential magnification. However, even in that case, a very rudimentary classification method following modeling with what we now know to be too simple a model (automated with Gaussians instead of hand-made with Sersic profiles) still results in a correlation between measured and true clump luminosity fractions, albeit a weak one. The correlation, of course, has too shallow a slope and has a large amount of scatter, but it is of similar quality to the corresponding figure for the low-shear subsample when it was also done using the automated Gaussian technique (correlation coefficient r=0.31 for low shear and r=0.24 for high shear) and many of the same correctable problems indicative of classification errors seem to exist—like the over-estimation at the lowest intrinsic clump luminosity fractions. These similarities, and the fact that more sophisticated methods resulted in tighter less biased correlations in the low-shear case are reasons to be optimistic that a similar effort could result in a tighter and less biased relationship in the high-shear case. The existence of a correlation at all suggests that at least some of the information we wish to extract from the source is preserved in the image plane and the fact that the scatter plot does not look much worse for the high-shear case than it did for the low-shear case in Fig. 4.5 suggests that the main sources driving the scatter are similar to those driving it in the low-shear case (modeling and classification uncertainties) and not something more fundamental and insurmountable like differential magnification or decomposition uncertainties.

Another question one might ask is: is the cut based on Sersic index applicable to real objects or even to the highly sheared simulated objects? The clumps that were inserted into the model sources had Sersic indices of 2, so it is reasonable to wonder whether this is physically realistic and whether that choice artificially determined the parameters by which the cut for Fig. 5.6 was carried out. It likely was the reason that the cut of n between 1.3 and 2.5 worked so well. However, the Sersic index of the source clumps was chosen because sources with clumps with n=2 did such a good job of reproducing real images of lensed sources after being artificially lensed in our pipeline. So while the cut on Sersic index in real data might need to vary, it may not need to vary much. It is reasonable to ask whether Sersic profiles can still fit a highly sheared clump since a highly stretched Sersic profile in one dimension would likely change the best-fit Sersic index of the components used to model it. But in two-dimensional images that have been lensed, only one dimension will be significantly stretched, and the Sersic index may still be well-preserved in the other dimension. A series of Sersic profiles, closely packed together, trying to fit a light profile like this may still preserve the Sersic index since it better fits the profile in the less distorted dimension and the profile in the more distorted dimension could be fit by carefully packing many components with that same Sersic index next to each other. But regardless, the cut on Sersic index is less important for its literal interpretation as a specific parameter of a physical or model component, and more important as a proxy for any process by which clumps can be reliably separated from smooth components. So the real question of interest here is whether there is some other piece of information that can be brought to bear on the question of component classification that can reliably distinguish between clumps and smooth components. Sersic index may be that piece of information, but another candidate would be colors. Remember, this entire analysis has been done using monochromatic images. But since central bulges and elliptical galaxies are expected to be composed of older stellar populations than would be found in actively star-forming clumps (Noguchi 1999, Dekel et al. 2009, etc.), color could potentially be used to separate clump and smooth components. In a modeling process where imaging data from multiple bands was used and the components in each band were linked so that each component could be assigned a color, it is possible that color could provide a dividing line for clump and smooth components in even the most ambiguous cases. Indeed, in many of the lensed images in SGAS, the clumps appear a different color than other components. However, in the case of SGAS1110+6459, it has been shown that the clump components have the same color as the smooth components (Rigby et al. 2017) meaning that such a discriminator, while potentially quite useful in many cases, will still not be universally applicable. Of course, SGAS1110+6459 is not typical of the arcs that we have observed. A further reason to be optimistic about the utility of using other bands for classification is that since JWST will cover a much broader wavelength range than HST, especially in the infrared, and with a similarly sized PSF to that of HST, it is possible that a different set of measured colors could better reveal the contrast between components even for SGAS1110+6459 in the near future. Still, regardless of whether color will ever separate the clump and smooth components of this one object, the strong possibility of being able to find such a discriminator that works in most cases, whether it be Sersic index, color, a combination, or something else, is more reason for optimism.

The work presented in chapters 4 and 5 is by no means a complete and polished solution to the problem of measuring the clump luminosity fraction in the image plane of strongly lensed objects, but it is a sort of proof of concept and a necessary first step. This method is not ready to be used on real data, but as I have argued, there is good reason to believe that a method based on similar principles can work and is possible to create. Using Sersic profiles instead of Gaussians or making tweaks to the clump identification or modeling algorithms may be able to improve the automated modeling process enough to allow researchers to bypass the manual modeling process. This is helpful for conducting this type of analysis on a large sample of objects, but absolutely necessary for application to a large simulated dataset like the one I have created and described in chapter 4 in order to quantify uncertainties and fully understand the advantages and limitations of the method. Component classification could be conducted fairly quickly even with human intervention, but it is possible that an automated method, perhaps based around machine learning, could remove the need for human intervention at this stage as well. And the ability to reproduce the intrinsic clump luminosity fractions of the simulated sources when the models and classification methods are refined only to the point that I have been able to reach in this work shows that effects like differential magnification and decomposition uncertainties will not impede the ability of a suitably sophisticated method to measure these values in the image plane, which would increase the sample of objects for which such values could be measured (since not all lenses can be modeled) and facilitate quick measurements (since there would be no need for the extremely time-consuming processes of lens modeling or source reconstruction). While there is still a lot of work to do before a method like this could be considered robust, I remain optimistic that one eventually will be.

CHAPTER 6

A BRIEF SUMMARY AND FUTURE APPLICATIONS OF THESE TECHNIQUES

The methods that I have begun to develop in chapters 4 and 5 are clearly not yet sufficient to robustly measure the clump luminosity fraction of lensed objects in real data. However, the results of chapter 5 show that methods for measuring the clump luminosity fraction that depend on decomposition into model components using GALFIT *can work* on lensed objects as long as the profiles used are sufficiently similar to the real profiles of the components. Furthermore, it seems that Sersic profiles are sufficient in at least the simple case examined in chapter 5. Of course, more work needs to be done in order to make this method robust for even these low-shear objects, possibly by incorporating some of the other observations or algorithmic approaches to classification that I discussed in chapter 5. Even then, additional effort will need to be put into broadening the set of image morphologies on which this type of measurement can be done, but there is reason to be optimistic.

Being able to perform these sorts of measurements at high redshifts will allow the study of the evolution of clump luminosity fractions (or more broadly, "clumpiness") over cosmic time, providing insight into some of the questions about galaxy mass assembly raised in chapter 1. If the clump luminosity peaks early, for example, it implies a rapid period of galaxy formation in the early universe, which would require a fast mass assembly processes like hot mode accretion. Conversely, if it peaks late, closer to redshift zero, it would mean that slower formation processes could have dominated at early times. Indeed, the fraction of galaxies that are considered "clumpy" has already been used to try to make similar inferences by, for example, Shibuya et al. 2016. But as I discussed in chapter 4, these analyses were either limited to giant clumps, whose relationship to the smaller clumps more common in nearby galaxies is unclear, or were identified as being "clumpy" subjectively (e.g., by visual inspection). A more inclusive concept of clumps could conceivably change these results. And ultimately, what we really want to know about are the star-forming processes at all scales, not just the largest ones. A more rigidly defined metric for clumpiness that keeps the idea of star-forming regions at its core—a metric like the fraction of a galaxy's light contained in clumps larger than a certain size, for instance—is more objective and quantitative and therefore easier to compare across redshifts. In the coming years, JWST will provide access to HST-like spatial scales at similar rest-frame wavelengths to what is available for low redshift objects with HST, but it will do it for higher redshift objects. It will be important to have the right metrics at our disposal when these data become available.

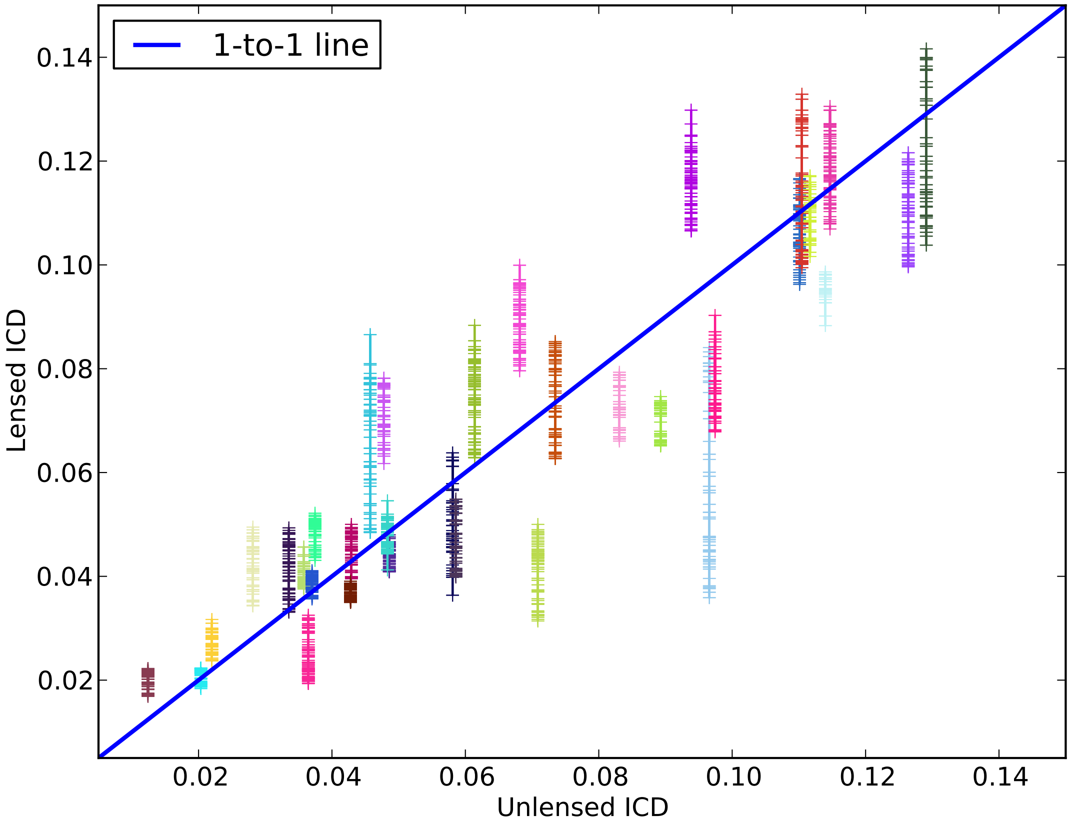
In addition to addressing the contributions of cold and hot mode accretion to galaxy mass assembly using characteristics of the star-forming clumps in galaxies, the contributions of minor and major mergers can now be probed using the Gini coefficient even in lensed objects. Higher Gini coefficients (i.e., light distributions that are most concentrated), tend to correlate with more elliptical or bulgy morphologies, at least at low redshift. As I mentioned in chapter 1, these types of morphologies are more indicative of the effects of mergers. Gini coefficients have already been measured in unlensed galaxies up to at least redshift 4 (Lotz et al. 2006), but since the lensed sample is selected differently and may, for example, include fainter objects, it is worthwhile to measure the Gini coefficients of these objects as well and compare them to the larger unlensed sample.

Furthermore, since the Gini coefficient has the fairly unique property that it can be measured robustly both in high redshift unlensed sources as well as in the image plane of lensed sources, it could be an effective tool for comparing these two samples and beginning to quantify any morphological biases introduced by the selection functions of the two samples. Understanding these biases will be necessary for eliminating bias in conclusions about mass assembly scenarios and formation histories inferred from observations of galaxies in these samples. A moderately sized sample of lensed objects already exists from our collaboration's HST program GO13003 (PI: Gladders) and samples extending to higher redshift will surely be generated by JWST. Not all of the lensed sources in the existing sample have been observed with high enough signal to noise ratios to effectively measure the Gini coefficient in accordance with chapter 2 of this thesis, but there are enough to form the beginning of a sample of lensed objects for comparison with unlensed objects. As for unlensed galaxies, Gini coefficients have been measured for many such objects observed in both widefield and deep surveys. A true comparison between them and their lensed counterparts would require carrying out these measurements again using the aperture definitions given in chapter 2, but it would be fairly straightforward to do this for sources in, say, the Hubble Ultra Deep Field, which would be a particularly convenient sample to use because of its depth and because of the number of spectroscopic and photometric redshifts that have already been published for this field (e.g., Coe et al. 2006).

In addition to what can be gleaned from the Gini coefficient, it may be possible to extract further morphological information from lensed objects in the image plane using other metrics. For example, the internal color dispersion (ICD), introduced by Papovich et al. 2003 may be preserved by lensing due to gravitational lensing being achromatic and because the ICD only depends the variation of fluxes between filters at each pixel, but not on the relative spatial locations of an image's constituent pixels. In fact, a preliminary attempt to test this hypothesis using the lensed sources from chapter 2 (lensed with a halo with a spherical NFW profile) suggests that this is indeed the case, but with the caveat that the uncertainties increase as the unlensed ICD increases Fig. 6.1. To do this analysis properly would require using more general profiles, at least an elliptical NFW profile like the one used for section 2.4 because it is capable of producing more types of images and is therefore more representative of how real observations would look. Still, this initial look is promising and shows that this idea deserves further examination.

If the ICD is preserved by lensing, then it too could be used to separate image families in the same way that I argued the Gini coefficient could in chapter 3. Still, as with the Gini coefficient, there has not been a true in-situ test of these metrics' abilities to separate image families in real data. However, the Frontier Fields present a rich testing ground for these

Figure 6.1: Lensed vs. Unlensed ICD



Lensed internal color dispersions plotted against unlensed ones for the same simulations as in chapter 2. Each color corresponds to a different source. For each source, the ICD for all realizations of the model were plotted, with error bars as calculated using the prescription in Papovich et al. 2003 at a signal-to-noise level of 0.01 per pixel. These sources were lensed using a spherical NFW profile, so a more complex profile is required to fully sample the range of possible image configurations. However, the agreement with the 1-to-1 line is encouraging in this simple case, despite the slight tendency for the spread in lensed ICD values to increase as the unlensed ICD increases.

methods thanks to their complex image configurations and multitude of lensed background sources, and to the extensive work that has already gone into image family identification by eye by lens modelers (e.g., Jauzac et al. 2014, Johnson et al. 2014, Richard et al. 2014, Grillo et al. 2015, Ishigaki et al. 2015). The main hurdle in such a test would be eliminating the contamination from intracluster light, but there are methods that are likely capable of doing this (e.g., parametric fitting techniques like GALFIT).

With metrics like the Gini coefficient and perhaps the clump luminosity fraction and the ICD, we can begin to extract detailed morphological information from high redshift objects magnified by gravitational lensing and we should soon be able to start making inferences about the formation and evolution of galaxies and their mass assembly histories. HST has provided a wealth of data that can be mined for these types of analyses already, and while it continues to operate, we will be able to further explore the details of the low redshift universe and, using lensing, probe the high redshift universe in the wavelengths accessible by HST with new observations. But such observations will be limited to the observed-frame optical, ultraviolet, and near-infrared, restricting our ability to make true apples-to-apples comparisons across redshift. What we really want, comparisons of morphologies at the same *rest frame* wavelengths across redshift, will likely have to wait until JWST and eventually WFIRST are launched, since they are better optimized for infrared observations—a necessity for studying rest-frame optical, ultraviolet and near-infrared features at high redshift. JWST will enable deep observations of individual high redshift lensed objects which will allow us to begin to perform this sort of analysis on small high redshift samples. And ultimately, WFIRST will observe large samples of these objects and enable a true statistical study of the change in star-formation morphologies in entire populations of galaxies over cosmic time with the lessons learned from the JWST samples.

Futhermore, while this thesis concerns itself with the type of information that can be gleaned from broadband imaging data, the spectroscopic abilities of JWST—especially its spatially-resolved spectroscopic capabilities—represent an entirely separate, but complementary universe of observations that can be brought to bear on these same questions of galaxy structure. In this chapter, I have speculated on several directions for future research, but that will only be the beginning of what is possible with JWST and WFIRST, and I hope that my contributions in this work will be able to help leverage the capabilities of these new instruments as we continue our quest to explain the diversity of structure in the universe.

REFERENCES

Abraham, R. G., van den Bergh, S., & Nair, P. "A New Approach to Galaxy Morphology. I. Analysis of the Sloan Digital Sky Survey Early Data Release." 2003, *Astrophysical Journal*, 588, 1, 218

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., & Bolton, A. S. "Two-dimensional kinematics of SLACS lenses–III. Mass structure and dynamics of early-type lens galaxies beyond $z \cong 0.1$." 2011, *Monthly Notice of the Royal Astronomical Society*, 415, 3, 2215

Baugh, C. M. "A primer on hierarchical galaxy formation: the semi-analytical approach." 2006, *Reports on Progress in Physics*, 69, 12, 3101

Bershady, M. A., Jangren, A., & Conselice, C. J. "Structural and Photometric Classification of Galaxies. I. Calibration based on a Nearby Galaxy Sample." 2000, *Astronomical Journal*, 119, 6, 2645

Bertin, E., & Arnout, S. "SExtractor: Software for source extraction." 1996, *Astronomy and Astrophysics Supplement Series*, 117, 393

Blanton, E. L., Gregg, M. D., Helfand, D.J., et al. "The Environments of a Complete Moderate-Redshift Sample of First Bent-Double Radio Sources." 2001, *Astronomical Journal*, 121, 6, 2915

Bouwens, R. J., Bradley, L., Zitrin, A., et al. "A Census of Star-Forming Galaxies in the $z\sim9$-10 Universe Based on HST+Spitzer Observations Over 19 CLASH Clusters: Three Candidate $z\sim9$-10 Galaxies and Improved Constraints on the Star Formation Rate Density at $z\sim9.2$." 2014, *Astrophysical Journal*, 795, 2, 126

Coe, D., Benítez, N., Sánchez, S. F., et al. "Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology." 2006, *Astronomical Journal*, 132, 2, 926

Coe, D., Bradley, L., & Zitrin, A. "Frontier Fields: High-Redshift Predictions and Early Results." 2015, *Astrophysical Journal*, 800, 2, 84

Conselice, C. J. "The Relationship Between Stellar Light Distributions of Galaxies and their Formation Histories." 2003, *Astrophysical Journal Supplement Series*, 147, 1, 1

Conselice, C. J. & Arnold, J. "The structures of distant galaxies–II. Diverse galaxy structures and local environments at $z=4$–6; implications for early galaxy assembly." 2009, *Monthly Notices of the Royal Astronomical Society*, 397, 1, 208

Conselice, C. J., Bershady, M. A., & Jangren, A. "The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies." 2000, *Astrophysical Journal*, 529, 2, 886

Conselice, C. J., Rajgor, S., & Myers, R. "The structures of distant galaxies–I. Galaxy structures and the merger rate to z$\sim$3 in the Hubble Ultra-Deep Field." 2008, *Monthly Notices of the Royal Astronomical Society*, 386, 2, 909

Dekel, A. & Birnboim, Y. "Galaxy bimodality due to cold flows and shock heating." 2006, *Monthly Notices of the Royal Astronomical Society*, 368, 1, 2

Dekel, A., Sari, R., & Ceverino, D. "Formation of Massive Galaxies at High Redshift: Cold Streams, Clumpy Disks, and Compact Spheroids." 2009, *Astrophysical Journal*, 703, 1, 785

Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C., "Discovery of Face-On Counterparts of Chain Galaxies in the Tadpole Advanced Camera for Surveys Field." 2004, *Astrophysical Journal*, 604, 1, L21

Elmegreen, D. M., Elmegreen, B. G., Ravindranath, S., & Coe, D. A. "Resolved Galaxies in the Hubble Ultra Deep Field: Star Formation in Disks at High Redshift." 2007, *Astrophysical Journal*, 658, 2, 763

Elmegreen, B. G., Elmegreen, D. M., Vollbach, D. R., Foster, E. R., & Ferguson, T. E. "On the Origin of Exponential Disks at High Redshift." 2005, *Astrophysical Journal*, 634, 1, 101

Elmegreen, B. G. & Struck, C. "Exponential Galaxy Disks from Stellar Scattering." 2013, *Astrophysical Journal Letters*, 775, 2, L35

Florian, M. K., Gladders, M. D., Li, N., & Sharon, K. "The Gini Coefficient as a Tool for Image Family Identification in Strong Lensing Systems with Multiple Images." 2016a, *Astrophysical Journal Letters*, 816, 2, L23

Florian, M. K., Li, N., & Gladders, M. D. "The Gini Coefficient as a Morphological Measurement of Strongly Lensed Galaxies in the Image Plane." 2016b, *Astrophysical Journal*, 832, 2, 168

Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. "The SINS Survey: SINFONI Integrated Field Spectroscopy of z$\sim$2 Star-Forming Galaxies." 2009, *Astrophysical Journal*, 706, 2, 1364

Freeman, P. E., Izbicki, R., Lee, A. B., et al. "New image statistics for detecting disturbed galaxy morphologies at high redshift." 2013, *Monthly Notices of the Royal Astronomical Society*, 434, 1, 282

Galametz, A., Grazian, A., Fontana, A., et al. "CANDELS Multiwavelength Catalogs: Source Identification and Photometry in the CANDELS UKIDSS Ultra-Deep Survey Field."

2013, *Astrophysical Journal Supplement Series*, 206, 2, 10

Genzel, R., Newman, S., Jones, T., et al. “The SINS Survey of z∼2 Galaxy Kinematics: Properties of Giant Star-Forming Clumps.” 2011, *Astrophysical Journal*, 733, 2, 101

Grillo, C., Suyu, S.H., Rosati, P., et al. “CLASH-VLT: Insights on the Mass Substructures in the Frontier Fields Cluster MACS J0416.1-2403 through Accurate Strong Lens Modeling.” 2015, *Astrophysical Journal*, 800, 1, 38

Groeneboom, N. E., & Dahle, H. “Introducing GAMER: A Fast and Accurate Method for Ray-Tracing Galaxies Using Procedural Noise.” 2014, *Astrophysical Journal*, 783, 2, 138

Grogin, N. A., Kocevski, D., Faber, S. et al. “CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey.” 2011, *Astrophysical Journal Supplement Series*, 197, 2, 35

Guo, Y., Ferguson, H. C., Bell, E. F., et al. “Clumpy Galaxies in CANDELS. I. The Definition of UV Clumps and the Fraction of Clumpy Galaxies at $0.5<z<3$.” 2015, *Astrophysical Journal*, 800, 1, 39

Hatton, S., Devriendt, J. E. G., Ninin, S., et al. “GALICS–I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation.” 2003 *Monthly Notices of the Royal Astronomical Society*, 343, 1, 75

Hoag, A., Huang, K.-H., Treu, T., et al. “The Grism Lens-Amplified Survey from Space (GLASS). VI. Comparing the Mass and Light in MACS J0416.1-2403 Using Frontier Field Imaging and GLASS Spectroscopy.” 2016, *Astrophysical Journal*, 831, 2, 182

Hubble, E. “Extra-Galactic Nebulae.” 1926, *Astrophysical Journal*, 64, 321

Ishigaki, M., Kawamata, R., Ouchi, M., et al. “Hubble Frontier Fields First Complete Cluster Data: Faint Galaxies at $z \sim 5$–10 for UV Luminosity Functions and Cosmic Reionization.” 2015, *Astrophysical Journal*, 799, 1, 12

Jauzac, M., Clément, B., Limousin, M., et al. “Hubble Frontier Fields: a high-precision strong-lensing analysis of galaxy cluster MACS0416.1-2403 using ∼200 multiple images.” 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 2, 1549

Jauzac, M., Richard, J., Jullo, E., et al. “Hubble Frontier Fields: a high-precision strong-lensing analysis of the massive galaxy cluster Abell 2744 using ∼180 multiple images.” 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 2, 1437

Johnson, T. L., Sharon, K., Bayliss, M. B., et al. “Lens Models and Magnification Maps of the Six Hubble Frontier Fields Clusters.” 2014, *Astrophysical Journal*, 797, 1, 48

Johnson, T. L., Sharon, K., Gladders, M. D., et al. “Star Formation at z = 2.481 in the Lensed Galaxy SDSS J1110 = 6459. I. Lens Modeling and Source Reconstruction.” 2017a, *Astrophysical Journal*, 843, 2, 78

Johnson, T. L., Rigby, J. R., Sharon, K., et al. “Star Formation at z = 2.481 in the Lensed Galaxy SDSS J1110+6459: Star Formation Down to 30 Parsec Scales.” 2017b, *Astrophysical Journal*, 843, 2, L21

Jones, T. A., Swinbank, A. M., Ellis, R. S., Richard, J., & Stark, D. P. “Resolved spectroscopy of gravitationally lensed galaxies: recovering coherent velocity fields in subluminous z $\sim$ 2–3 galaxies.” 2010, *Monthly Notices of the Royal Astronomical Society*, 404, 3, 1247

Jullo, E., Natarajan, P., Kneib, J.-P., D’Aloisio, A., & Limousin, M. “Cosmological Constraints from Strong Gravitational Lensing in Clusters of Galaxies.” 2010, *Science*, 329, 5994, 924

Kawamata, R. Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. “Precise Strong Lensing Mass Modeling of Four Hubble Frontier Field Clusters and a Sample of Magnified High-Redshift Galaxies.” 2016, *Astrophysical Journal*, 819, 2, 114

Kennicutt Jr., R. C. “Structural Properties of Giant HII Regions in Nearby Galaxies.” 1984, *Astrophysical Journal*, 1, 287, 116

Kennicutt Jr., R. C. “The Global Schmidt Law in Star-Forming Galaxies.” 1998, *Astrophysical Journal*, 498, 2, 541

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. “How do galaxies get their gas?” 2005, *Monthly Notices of the Royal Astronomical Society*, 363, 1, 2

Koekemoer, A. M., Avila, R. J., Hammer, D., et al. “The HST Frontier Fields: Science Data Pipeline, Products and First Data Release.” 2014, *American Astronomical Society*, Meeting 223, 254.02

Koekemoer, A. M., Faber, S., Ferguson, H. et al. “CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics.” 2011, *Astrophysical Journal Supplements*, 197, 2, 36

Laporte, N., Streblyanska, A., Clement, B., et al. “The first Frontier Fields cluster: 4.5μm excess in a z$\sim$8 galaxy candidate in Abell 2744.” 2014, *Astronomy & Astrophysics*, 562, L8

Lefor, A. T., Futamase, T., & Akhlaghi, M. “A systematic review of strong gravitational lens modeling software.” 2013, *New Astronomy Reviews*, 57, 1-2, 1

Li, N., Gladders, M. D., Rangel, E. M. et al. “PICS: Simulations of Strong Gravitational

Lensing in Galaxy Clusters."2016, *Astrophysical Journal*, 828, 1, 54

Lisker, T. "Is the Gini Coefficient a Stable Measure of Galaxy Structure?" 2008, *Astrophysical Journal Supplement Series*, 179, 2, 319

Lotz, J. M., Primack, J., & Madau, P. "A New Nonparametric Approach to Galaxy Morphological Classification." 2004, *Astrophysical Journal*, 128, 1, 163

Lotz, J. M., Madau, P., Giavalisco, M., Primack, J., Ferguson, H.C. "The Rest-Frame Far-Ultraviolet Morphologies of Star-Forming Galaxies at z$\sim$1.5 and 4." 2006, *Astrophysical Journal*, 636, 2, 592

Lotz, J. M., Jonsson, P., Cox, T. J., et al. "The Major and Minor Galaxy Merger Rates at z $<$ 1.5" 2011, *Astrophysical Journal*, 742, 2, 103

Lotz, J. M., Bountain, M., Grogin, N. A., et al. "The HST Frontier Fields." 2014, *American Astronomical Society*, Meeting 223, 254.01

Madau, P., Pozzetti, L., & Dickinson, M. "The Star Formation History of Field Galaxies." 1998, *Astrophysical Journal*, 498, 1, 106

Mahdi, H. S., van Beek, M., Elahi, P. J., et al. "Gravitational lensing in WDM cosmologies: the cross-section for giant arcs." 2014, *Monthly Notices of the Royal Astronomical Society*, 441, 3, 1954

Moster, B. P., Macciò, A. V., Somerville, R. S., Johansson, P. H., & Naab, T. "Can gas prevent the destruction of thin stellar discs by minor mergers?" 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 2 1009

Naab, T., Johansson, P. H., Ostriker, J. P., & Efstathiou, G. "Formation of Early-Type Galaxies from Cosmological Initial Conditions." 2007, *Astrophysical Journal*, 658, 2, 710

Noguchi, M. "Early Evolution of Disk Galaxies: Formation of Bulges in Clumpy Young Galactic Disks." 1999, *Astrophysical Journal*, 514, 1, 77

Papovich, C., Giavalisco, M., Dickinson, M., Conselice, C. J., & Ferguson, H. C. "The Internal Ultraviolet-Optical Color Dispersion: Quantifying the Morphological K-Correction." 2003, *Astrophysical Journal*, 598, 2, 827

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. "Detailed Structural Decomposition of Galaxy Images." 2002, *Astronomical Journal*, 124, 1, 266

Porter, L. A., Somerville, R. S., Primack, J. R., & Johansson, P. H. "Understanding the structural scaling relations of early-type galaxies." 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 1, 942

Richard, J., Jauzac, M., Limousin, M., et al. "Mass and magnification maps for the Hubble Space Telescope Frontier Fields clusters: implications for high-redshift studies." 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 1, 268

Rigby, J. R., Johnson, T. L., Sharon, K., et al. "Star Formation at z=2.481 in the Lensed Galaxy SDSS J1110+6459, II: What is missed at the normal resolution of the Hubble Space Telescope?" 2017, *Astrophysical Journal*, 843, 2, 79

Shibuya, T., Ouchi, M., & Harikane, Y. "Morphologies of ∼190,000 Galaxies at z = 0-10 Revealed with *HST* Legacy Data. I. Size Evolution." 2015, *Astrophysical Journal Supplement Series*, 219, 2, 15

Shibuya, T., Ouchi, M., Kubo, M., & Harikane, Y. "Morphologies of ∼190,000 Galaxies at z=0–10 Revealed with HST Legacy Data. II. Evolution of Clumpy Galaxies." 2016, *Astrophysical Journal*, 821, 2, 72

Sharon, K., Gladders, M. D., Rigby, J. R., et al. "Source-plane Reconstruction of the Bright Lensed Galaxy RCSGA 032727-132609." 2012, *Astrophysical Journal*, 746, 2, 161

Simard, L., Willmer, C. N. A., Vogt, N. P., et al. "The Deep Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip." 2002, *Astrophysical Journal Supplement Series*, 142, 1, 1

Somerville, R. S. & Davé, R. "Physical Models of Galaxy Formation in a Cosmological Framework." 2015, *Annual Reviews in Astronomy and Astrophysics*, 53, 51

Stark, D. P., Swinbank, M. A., Ellis, R. S., et al. "The formation and assembly of a typical star-forming galaxy at redshift z∼3." 2008, *Nature*, 455, 7214, 775

Swinbank, A. M., Webb, T. M., Richard, J., et al. "A spatially resolved map of the kinematics, star formation and stellar mass assembly in a star-forming galaxy at z=4.9." 2009, *Monthly Notices of the Royal Astronomical Society*, 400, 3, 1121

Tacconi, L. J., Neri, R., Genzel, R., et al. "Phibbs: Molecular Gas Content and Scaling Relations in z ∼ 1-3 Massive, Main-sequence Star-forming Galaxies." 2013, *Astrophysical Journal*, 768, 1, 74

Toomre, A. "Mergers and Some Consequences." 1977, *Evolution of Galaxies and Stellar Populations, Proceedings of a Conference at Yale University*, 401

Treu, T., Brammer, G., Diego, J. M., et al. "'Refsdal' Meets Popper: Comparing Predictions of the Re-Appearance of the Multiply Imaged Supernova Behind MACSJ1149.5+2223." 2016, *Astrophysical Journal*, 817, 1, 60

Umetsu, K., Medezinski, E., Nonino, M., et al. “CLASH: Mass Distribution in and around MACS J1206.2-0847 from a Full Cluster Lensing Analysis.” 2012, *Astrophysical Journal*, 755, 1, 56

van Dokkum, P. G. “The Recent and Continuing Assembly of Field Elliptical Galaxies By Red Mergers.” 2005, *Astrophysical Journal*, 130, 6, 2647

Walker, I. R., Mihos, J. C., & Hernquist, L. “Quantifying the Fragility of Galactic Disks in Minor Mergers.” 1996, *Astrophysical Journal*, 420, 121

Yasuda, N., Fukugita, M. Narayanan, V. K., et al. “Galaxy Number Counts from the Sloan Digital Sky Survey Commissioning Data.” 2001, *Astronomical Journal*, 122, 3, 1104