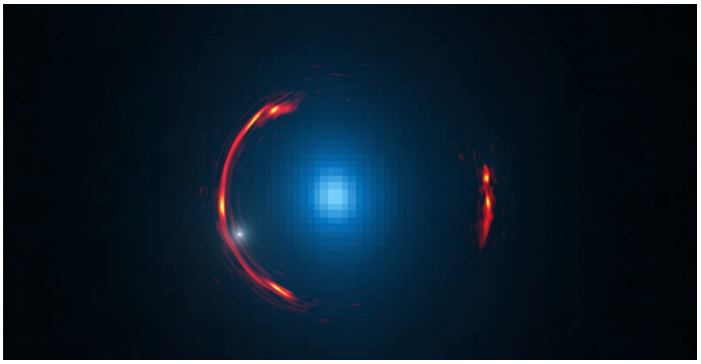
University of Nottingham
UK | CHINA | MALAYSIA

Modern Methods and Their Applications to Strong Gravitational Lensing

Jacob Maresca



Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2022

Have you finished your thesis yet?

– Simon Dye

Supervisors:	Dr Simon Dye	
Examiners:	Dr Graham Smith	(University of Birmingham)
	Prof. Omar Almaini	(University of Nottingham)
Submitted:	22/07/2022	
Examined:	15/09/2022	
Final version:	07/01/2023	

Contents

Abstract **v**

Acknowledgements **vi**

Published Work **vii**

Statement of Originality **viii**

1 Introduction **1**

1.1 The ΛCDM Cosmological Model 1

1.2 Galaxy Evolution 6

1.2.1 Dusty Star Forming Galaxies 7

1.2.2 The Evolution of Mass Density Profiles in Early Type Galaxies 14

1.3 Gravitational Lensing Theory 19

1.3.1 The Deflection Angle 19

1.3.2 The Lens Equation 23

1.3.3 The Deflection Potential 25

1.3.4 Magnification and Distortion 26

1.3.5 Critical Curves and Caustics 29

1.3.6 The Mass Sheet Degeneracy 29

1.3.7 The Fermat Potential 31

1.4 Basic Lens Models 33

1.4.1 Axially Symmetric Lenses 34

1.4.1.1 The Point Mass Lens 36

1.4.1.2 The Singular Isothermal Sphere 37

1.4.2 Elliptical lens models . . . 38
1.5 Gravitational Lensing Science . . . 39
1.5.1 The Einstein Mass . . . 39
1.5.2 Measuring The Hubble-Lemaître Constant . . . 41
1.6 Interferometry and Synthesis Imaging . . . 43
1.6.1 Interferometery . . . 43
1.6.2 Synthesis Imaging . . . 44
1.7 Semi-Linear Gravitational Lens Inversion . . . 47
1.7.1 Regularised Linear Inversion . . . 50
1.7.2 Adaptive Semi Linear Inversion . . . 51
1.7.3 The semi-linear inversion method in the uv-plane . . . 53
1.8 Machine Learning . . . 56
1.8.1 Neural Networks . . . 57
1.8.2 Convolutional neural networks . . . 59
1.9 Outline . . . 60

2 Comparing and combining BNNs and parametric profile fitting 63
2.1 Gravitational Lens Simulation . . . 68
2.1.1 Training set . . . 68
2.1.2 Test sets . . . 71
2.2 Lens modelling methodology . . . 73
2.2.1 `PyAutoLens` . . . 73
2.2.2 Convolutional Neural Networks . . . 75
2.2.2.1 Architecture . . . 75
2.2.2.2 Uncertainties . . . 77
2.2.2.3 Training . . . 77
2.2.3 `PyAutoLens` + CNN . . . 78
2.3 Results . . . 79
2.3.1 SIE Lenses + Parametric Sources . . . 81
2.3.2 SIE Lenses + HUDF Sources . . . 83
2.3.3 Power Law Lenses + HUDF Sources . . . 84

2.3.4 EAGLE Lenses + HUDF Sources . . . 89
2.3.5 Modelling Speed . . . 94
2.4 Discussion . . . 98
2.4.1 Modelling Accuracy . . . 100
2.4.2 Modelling Speed . . . 101
2.4.3 Modelling Difficulties and Limitations . . . 102
2.5 Summary . . . 103

3 Auto-identification of unphysical source reconstructions 106
3.1 Erroneous Solutions and Their Inversions . . . 109
3.2 Methodology . . . 114
3.2.1 Lensing Simulations . . . 114
3.2.2 Training Data . . . 117
3.2.3 Testing Data . . . 123
3.2.4 CNN Architecture . . . 126
3.2.5 Combining CNN and Lens Modelling . . . 129
3.3 Results . . . 132
3.3.1 CNN performance . . . 133
3.3.2 Performance of `PyAutoLens` combined with CNN . . . 136
3.4 Conclusions . . . 138

4 High-resolution ALMA observations of strongly lensed DSFGs 142
4.1 Data . . . 145
4.2 Methodology . . . 151
4.2.1 The semi-linear inversion method in the uv-plane . . . 151
4.2.2 Lens Model . . . 151
4.3 Results . . . 152
4.3.1 Intrinsic source properties . . . 156
4.3.2 Object notes . . . 161
4.4 Discussion . . . 165
4.5 Conclusions . . . 170

5 Lens modelling biases in CLEANed images **173**

5.1 Simulated Data 175

5.2 Methodology 178

5.2.1 Image-plane modelling 178

5.2.2 uv-plane modelling 181

5.2.3 Lens Model 182

5.3 Results 182

5.3.1 Time-binning visibilities 182

5.3.2 Modelling results 187

5.4 Discussion 192

5.5 Conclusions 209

6 Conclusions and Future Work **213**

6.1 Summary of main results 213

6.2 Future improvements 216

6.3 Outlook 218

Bibliography **221**

Abstract

This thesis presents a collection of works, some methodological and others observational, with the common theme of strong gravitational lensing. We have leveraged a wide variety of techniques, from machine learning algorithms to uv-plane modelling, and used a multitude of simulated and real data products to achieve the results contained within this thesis.

In Chapter 2, we explored the feasibility of combining the predictions of an approximate Bayesian neural network with the parametric profile fitting of `PyAutoLens`. By using the predicted 1-σ uncertainties on the lens model parameters from the CNN to inform our choice of prior within `PyAutoLens`, we were able to show that a technique that combines both of these tools is able to outperform either method alone. We applied our methodology to a series of increasingly complex and realistic simulated strong gravitational lenses, beginning with simple parametric lenses and sources, and graduating to HUDF sources and lenses extracted from the EAGLE simulation.

In Chapter 3, we addressed the issue of the existence of unphysical source reconstructions by applying a convolutional neural network to detect these solutions, and developed a simple prescription to re-initialise the lens modelling process in a new region of parameter space to help ensure convergence upon the global solution. Such tools are necessary components to developing a truly automated lens modelling pipeline, as will become increasingly necessary in the era of LSST and Euclid.

In Chapter 4, we presented and modelled $\sim$0.1 arcsec resolution ALMA imaging of seven strong gravitationally lensed galaxies detected by the Herschel Space Observatory. We inferred the mass profiles of the lensing galaxies and by determining the magnification factors, we investigated the intrinsic properties and morphologies of the lensed submillimetre sources. We found that these submillimetre sources all have ratios of star formation rate to dust mass that are consistent with, or in excess of, the mean ratio for high-redshift submillimetre galaxies and low redshift ultra-luminous infrared galaxies. Reconstructions of the background sources reveal that the majority of our sample display disturbed morphologies. The majority of our lens models have mass density slopes close to isothermal, but some systems show significant differences.

In Chapter 5, we present a comparison between performing lens modelling in the uv-plane versus the image-plane. When dealing with interferometer observations, one must make a choice of whether to perform the analysis on the raw visibility data, or to employ an algorithm such as CLEAN to produce an image to work with. When producing an image from visibility data, there are several choices one must make that can impact the final image, such as the adopted weighting scheme. We investigated whether these choices can lead to biases in the inferred lens model parameters, and whether some choices outperform others. We found that in general, direct modelling of the visibilities provided the most robust means for recovering the lens model parameters, but also that using the Briggs weighting scheme performed better than the natural weighting scheme.

Acknowledgements

Firstly, I would like to thank my supervisor, Simon Dye. You have given me the freedom to work as independently as I desired, and yet have always been ready to give up your time when I needed help. It has been invaluable to me knowing that I can knock on your door for a chat about a strange result or interesting plot. I have learned a great deal from conversations with you, from your incredible knowledge of all things lensing, to simply how to approach challenging problems.

I have also been incredibly lucky to work with some excellent colleagues during my time at Nottingham. Nan, thank you for your patience in helping me with a myriad of problems I was a little too embarrassed to bring to our group meetings. You saved me countless hours of scratching my head. Thank you to James Pearson for being a pleasure to work with and all-round machine learning whizz. Listening to you talk about your work during group meetings inspired my first proper independent project. I would be remiss if I did not also thank James Nightingale, for tirelessly fielding my questions regarding PyAutoLens, the software that underpins so much of the work in this thesis. Thank you for your endless troubleshooting, bug fixing and patience with my ability to generate new and exciting errors!

This acknowledgment section would be woefully incomplete without thanking my long-suffering housemates and fellow publicans of the Dunkirk Arms (DA); Mick, Roan and Liza. I can't overstate how much I have enjoyed our time in the DA and the general nonsense that has accompanied living with you all. From bean-eating competitions to blowing up the garden. From ceaseless circulation to sustained screeching. I wouldn't change any of it, and you are all going to be deeply missed.

There are so many more people to thank in Nottingham that have made my life here as wonderful as it is; Sukhi, Leo, Kellie, CB, Karel, Cha, Jéssica, Tomáš, Simon W, Finlay, Tom C, Steph, the list goes on... I have many wonderful memories with all of you, and photographic evidence of far too much of it! I would also like to thank Dan, Brad, Paris, Tim, Vinay, Ellie, Katie, Vinnie, Jasper, Anna P, Sam C, Mr S., Colin, Wenqi and countless other members/instructors of the UoN Taekwondo club. You have all been incredible friends and encouraged me to do things that I never would have previously dreamed of. Being able to kick each other and de-stress from my PhD certainly helped too!

Lastly, but certainly not least, I would like to thank my family. Without your constant support, this truly would not have been possible, and I would be a miserable architect. In particular, thank you Mum for always taking an interest in what I was doing and being there for me to vent to when things were not going quite so well. And thank you, Marcus, for dedicating so much time and energy to helping me with maths; I really wouldn't be doing this otherwise.

Published works

Most of the content in this thesis has already been published in the following three works:

i Pearson, J., **Maresca, J.**, Li, N., Dye, S., *Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting, Monthly Notices of the Royal Astronomical Society, Volume 505, Issue 3, August 2021, Pages 4362–4382*

ii **Maresca, J.**, Dye, S., Li, N., *Auto-identification of unphysical source reconstructions in strong gravitational lens modelling, Monthly Notices of the Royal Astronomical Society, Volume 503, Issue 2, May 2021, Pages 2229–2241*

iii **Maresca, J.**, Dye, S., Amvrosiadis, A., Bendo, G., Cooray, A., De Zotti, G., Dunne, L., Eales, S., Furlanetto, C., González-Nuevo, J., Greener, M., Ivison, R., Lapi, A., Negrello, M., Riechers, D., Serjeant, S., Tergolina, M., Wardlow, J., *Modelling high-resolution ALMA observations of strongly lensed dusty star-forming galaxies detected by Herschel, Monthly Notices of the Royal Astronomical Society, Volume 512, Issue 2, May 2022, Pages 2426–2438*

Chapter 2 is based on material published in Paper i, Chapter 3 is based on Paper ii, and Chapter 4 is based on Paper iii.

The vast majority of the work presented in this thesis was carried out by the author, with supplementary advice from various co-authors listed above. In any instances where the work includes the product of larger collaborations, this is explicitly mentioned in the relevant chapter.

Statement of Originality

Here, I describe contributions I made to each of the chapters presenting new results in my thesis.

i Chapter 2 presents the modelling results of combining a Bayesian neural network with a parametric profile fitting approach to lens modelling. Whilst this section is entirely written in my own words, the content of this chapter is based upon the Pearson et al. (2021) publication. My contributions to this paper were chiefly that I performed all the parametric modelling of the simulated strong gravitational lenses. This includes cases where PyAutoLens was employed in isolation, and the different ways in which we used the CNN to inform the choice of priors used by PyAutoLens.

ii Chapter 3 deals with the automated removal of unphysical source reconstructions in the lens modelling process. This chapter is almost entirely my own work, except for the production of the simulated lensed images of HUDF sources, which were supplied to me by my co-author Nan Li.

iii Chapter 4 presents an analysis of ALMA observations of lensed SMGs. This chapter presents all my own work, except where it has been made clear that I am including the results of Dye et al. (2018). I carried out all the analysis myself, with useful input and suggestions from my co-authors.

iv Chapter 5 showcases some unpublished work I have carried out, investigating the relative performance of carrying out lens modelling in the uv-plane vs the image plane. This work is my own, having created the simulated data products and performed all the modelling and analysis myself. I benefitted from many useful conversations with my supervisor, Simon Dye, in the process.

Chapter 1

Introduction

1.1 The ΛCDM Cosmological Model

Currently, our most successful cosmological model is the flat ΛCDM model of the Universe (Planck Collaboration et al., 2020a), which describes the formation of structures in a hierarchical fashion. Observations of the universe on large scales show it to be expanding (Hubble, 1929), with the average density of matter, in large volumes, decreasing. If we imagine rewinding the clock on this cosmic evolution, we infer that at some point in the past, there must have been a point where the density of the universe was infinite. It is this moment that we refer to as the "Big Bang". The Universe evolves from $t = 0$ (see Figure 1.1) through a period of cosmic inflation (Guth, 1981; Linde, 1982), during which an extremely rapid expansion of spacetime occurs between 10^{-36} and 10^{-32} seconds after the big bang. At such early times, the universe is nearly homogeneous, except for small quantum mechanical perturbations in density (Mukhanov & Chibisov, 1981), and it is these regions of over and under density that sow the seeds of structure formation.

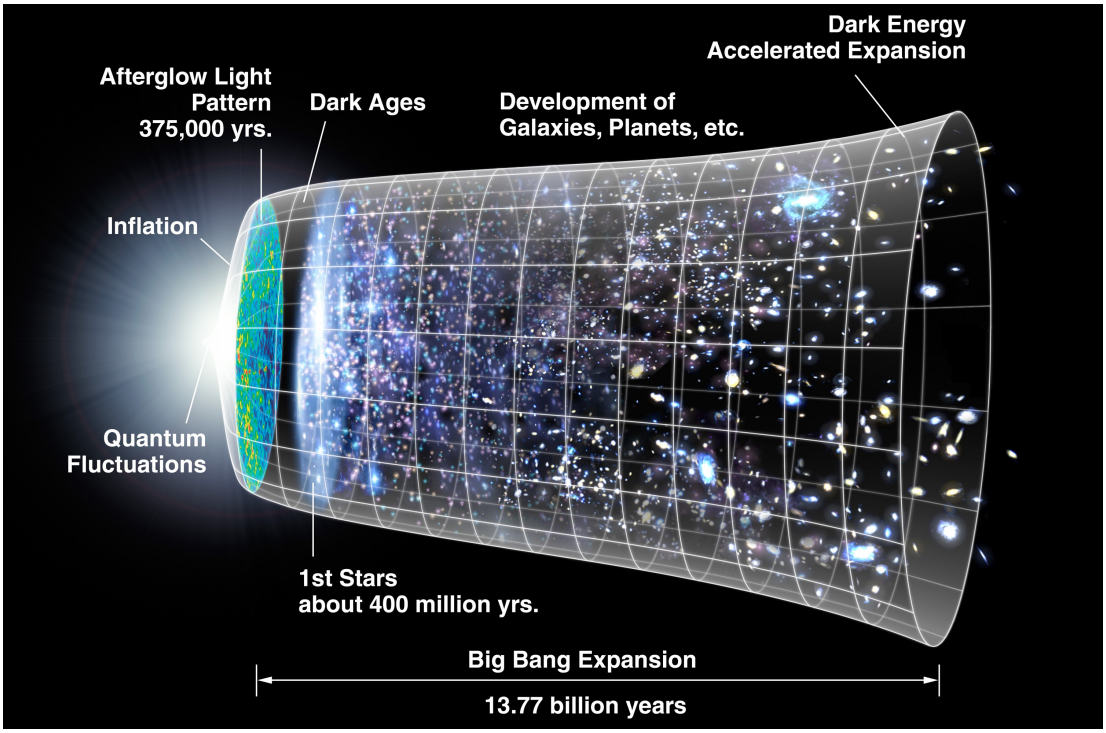


Figure 1.1: Representation of the expansion history of the Universe. Image reproduced from NASA & WMAP Science Team (2006)

As the Universe continues to grow and cool over the next second, the first baryons are able to form, though it is still far too hot for neutral matter to exist, and so the universe is filled by an opaque plasma. One second after the big bang, the first baryons have formed, forming the coupled photon-baryon soup that permeates the early universe. The matter density fluctuations propagate at the speed of sound in this opaque plasma, forming what are known as Baryon Acoustic Oscillations (Peebles & Yu, 1970; Sunyaev & Zeldovich, 1970). After 380,000 years, the Universe had cooled to an extent that allowed the first atoms to form and photons are finally able to stream freely throughout the Universe, creating the Cosmic Microwave background (CMB) that we observe today (Penzias & Wilson, 1965). At this point, the photons and baryons have become decoupled from one another, leaving the matter density perturbations frozen into the dark matter distribution. On large scales, the CMB is observed to be homogeneous and isotropic, with a near perfect black body spectrum characterised by a temperature of $T \sim 2.7$K today (Fixsen, 2009). The CMB can be characterised by its power spectrum, which describes the relative contributions to the anisotropies of the CMB as a function of angular scale (see Figure 1.2). The CMB power spectrum is a powerful tool for cosmologists to study the early Universe. For example, the location of the peaks in the power spectrum can be used to determine the curvature of the Universe (Hu & White, 1997).

The period of time after this initial release of photons is known as the "Dark Ages", whilst the clouds of hydrogen gas slowly collapse to form the first galaxies and stars. This period of time is referred to as "Dark" because until the first stars have formed, the only sources of light are the rapidly red-shifting photons of the CMB and the occasional 21-cm radio emissions of Hydrogen atoms. As time continues, the dark matter and baryons continue to slowly collapse into virialised haloes (Kravtsov & Borgani, 2012). The hierarchical growth of the ΛCDM model tells us that smaller

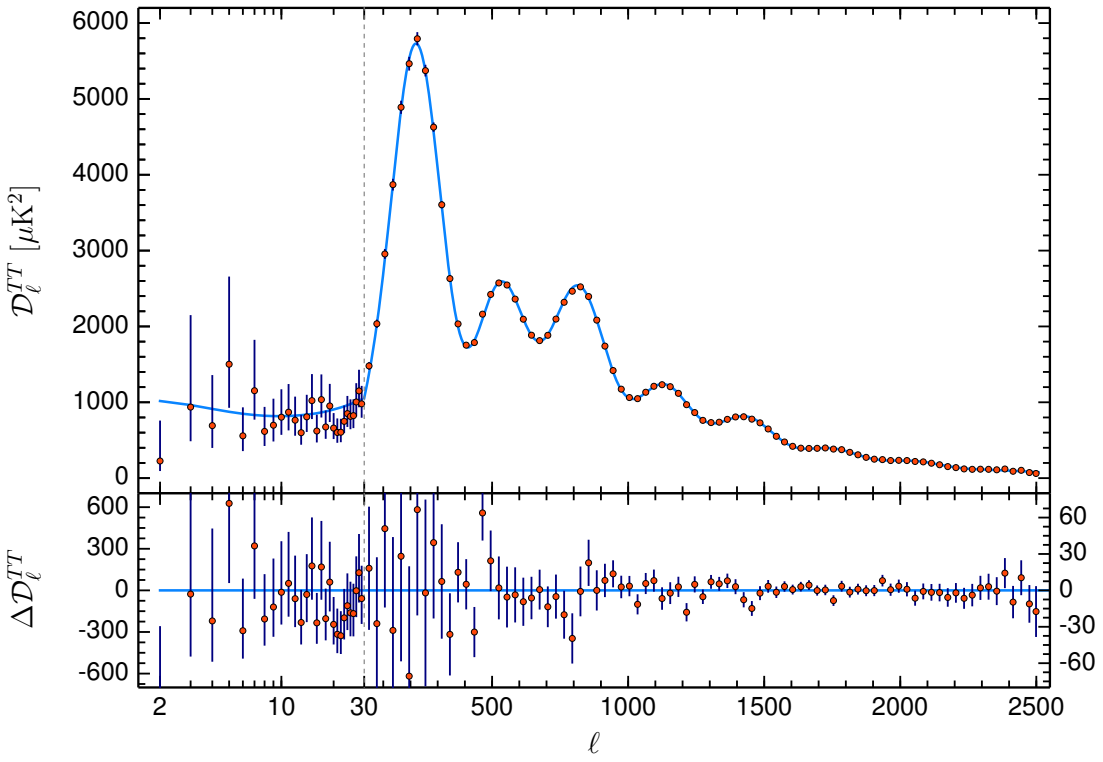


Figure 1.2: Planck 2018 CMB temperature power spectrum from Planck Collaboration et al. (2020b). The solid blue line shows the best fit to the ΛCDM theoretical predictions.

haloes form first, with larger haloes growing more slowly. This can be understood in terms of larger over densities, typically be less dense than smaller ones. As such, the haloes that host galaxies form earlier than the much larger haloes that form galaxy clusters. These dark matter haloes provide the potential wells for baryons to collapse onto, and as the primordial Hydrogen and Helium cools to temperatures of around $T \sim 20 - 30\text{K}$, molecular gas clouds are able to collapse to the point that the first stars can form (Couchman & Rees, 1986).

The intense ultraviolet (UV) radiation emitted by the first stars, quasars and galaxies began the gradual process of reionisation that began approximately 300 Myrs after the Big Bang, and finishes by approximately 1 Gyr after the Big Bang (Loeb & Barkana, 2001). This process of reionisation turns the inter-galactic medium (IGM) from neutral Hydrogen into a highly ionised state, allowing for the passage of Lyman continuum photons (Loeb & Barkana, 2001; Barkana & Loeb, 2001).

As cosmic time continues to progress towards the present day, the dark matter haloes distributed throughout the Universe, are drawn together by the attractive forces of gravity, forming ever more massive haloes (Somerville et al., 2012). The galaxies that call these haloes home also undergo mergers and this plays an important role in their evolution, as will be discussed in the following section. The observed distribution of galaxy masses in the local universe, or number density of galaxies per logarithmic mass bin, is one that follows a power law for low masses, and has an exponential cutoff at the high-mass end, often fitted with the Schechter function (Baldry et al., 2012; Press & Schechter, 1974). The location of the exponential cut-off, or "characteristic mass" in the local universe is approximately $M^* \sim 10^{10.6} - 10^{11}\text{M}_\odot$ (Panter et al., 2007; Li & White, 2009). As such, there are many more low mass galaxies than those with masses $m > M^*$. In order to explain the shape of this galaxy mass function, we require that the fraction of baryonic mass converted into stars increases with mass to

a particular value before decreasing again (Shankar et al., 2006; Marinoni et al., 2002). Theories of galaxy formation must therefore explain the suppression of star formation efficiency at the low mass end of the slope, as well as the exponential cut-off at high masses. In low mass galaxies, galactic winds from supernovae feedback is thought to play a critical role in regulating star formation (Dekel & Silk, 1986; Larson, 1974), whilst at the high mass end of galaxies feedback from active galactic nuclei (AGN) is thought to prevent the necessary cooling of gas for star formation activity (Kereš et al., 2005; Bower et al., 2006).

1.2 Galaxy Evolution

There exists a number of bimodalities in galaxy populations, both at high redshift and in the local universe, in terms of their morphology, colour and star formation. In 1926, Hubble famously discovered that galaxies can be split in terms of their morphology into two broad groups of early and late-type galaxies (see Figure 1.3). The early-type classification includes elliptical and lenticular galaxies, whilst the late-type classification includes spiral and barred spiral galaxies. The nomenclature of this classification scheme has often led to the presumption that Hubble implied an evolutionary pathway from early to late-type galaxies, which is in fact the opposite direction of evolution we currently expect. However, Hubble never intended this to be interpreted in such a way, saying that "The nomenclature, it is emphasized, refers to position in the sequence, and temporal connotations are made at one's peril. The entire classification is purely empirical and without prejudice to theories of evolution..." (Hubble, 1927).

A related bimodality in their colour is observed, such that galaxies typically fall into either the passively evolving red-sequence or the star-forming blue cloud (Strateva et al., 2001; Brammer et al., 2009). It is widely believed that the quenching of

star formation is responsible for causing the migration of galaxies in the blue cloud to the red sequence. In addition to being quenched, there exists a morphological difference between the two populations, with passive galaxies usually having early-type morphologies and star-forming galaxies preferentially forming discs (Kennicutt, 1998).

A wide variety of galaxy evolution models have been proposed to explain both the quenching of star formation and morphological changes required to move a galaxy from the blue cloud into the red sequence (Dekel & Birnboim, 2006; Martig et al., 2009; Matteo et al., 2005; Eales et al., 2018). The underpinning of many of these models is the major merger scenario, where a starburst is triggered by the merging event, and AGN feedback drives away the remaining gas (particularly relevant in high mass galaxies) (Hopkins et al., 2006b; Trayford et al., 2016). Alternatively, models have shown that the inflow of cold gas can cause disc instabilities and form compact spheroids (Dekel et al., 2009b)

1.2.1 Dusty Star Forming Galaxies

Dusty Star Forming Galaxies (DSFGs) (see Blain et al., 2002; Casey et al., 2014; Shapley, 2011, for a comprehensive review) at high redshift play host to some of the most intense rates of star formation in the Universe. The rest frame UV emission associated with young stellar populations is often completely obscured by the large dust reservoirs in DSFGs (e.g Dudzevičiūtė et al., 2020), produced by low/intermediate mass stars and supernovae. This absorbed radiation is re-emitted by the high opacity dust in the sub-millimetre wavelength regime. Resultantly, DSFGs are most commonly detected via their continuum emission in the far-IR to millimetre wavelengths, with typical dust temperatures in the range of 30-50K (Casey et al., 2014; Bussmann et al., 2013).

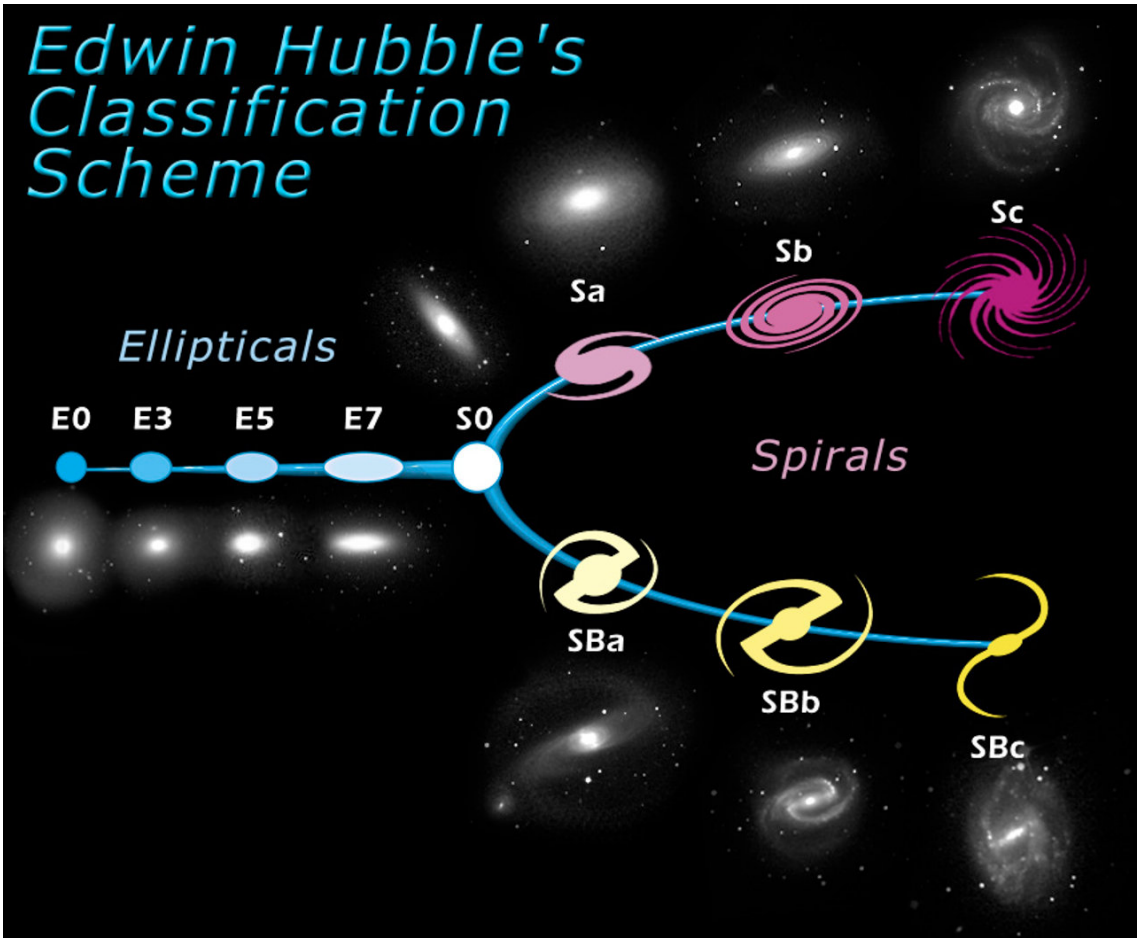


Figure 1.3: Hubble's tuning fork diagram showing that galaxies can be broadly divided into two categories; ellipticals and spirals. Image reproduced from NASA & ESA (2013)

The first DSFGs were discovered in the 850μm sub-mm band by the Submillimeter Common-User Bolometer Array (SCUBA; Holland et al., 1999; Hughes et al., 1998), an instrument on the James Clerk Maxwell Telescope (JCMT). For this reason, they are often referred to as sub-millimetre galaxies (SMGs) and this term will be used for the remainder of this thesis.

Studies of the star formation rate (SFR) volume density as a function of redshift show that the peak of cosmic star formation activity occurred at a redshift of $1 < z < 3$ (Madau & Dickinson, 2014). This peak, occurring between 1.5-3.5 Gyrs after the big bang, corresponds to $\sim 0.2 \mathrm{M}_{\odot} \mathrm{yr}^{-1} \mathrm{Mpc}^{-3}$, which is an order of magnitude higher than the present day value of $0.01 \mathrm{M}_{\odot} \mathrm{yr}^{-1} \mathrm{Mpc}^{-3}$ (Hopkins & Beacom, 2006). SMGs contribute an estimated 20 per cent of the cosmic star formation rate (SFR) density up to a redshift of $z \sim 4$ (Swinbank et al., 2013; Lapi et al., 2011) and hence provide a valuable aid to our understanding of the formation of galaxies during this period. Additionally, ~ 50 per cent of all the extra-galactic background radiation is from the sub-mm regime, and so understanding the sources of this radiation is vital to obtaining a complete picture of the contributors to the star formation density of the universe (Fixsen et al., 1998; Hauser & Dwek, 2001; Hill et al., 2018). The redshift distribution of SMGs peaks at $z \sim 2.5$ (Chapman et al., 2005; Simpson et al., 2014), which is coincident with the epoch associated with the greatest AGN activity (Richards et al., 2006; Assef et al., 2011). Typical SFRs of SMGs, measured from their far-IR/sub-mm continuum Spectral Energy Distribution (SED), are on the order of hundreds or even thousands of solar masses per year, which are extreme when compared with the SFR of our galaxy of $2 \mathrm{M}_{\odot} \mathrm{yr}^{-1}$ (Chomiuk & Povich, 2011).

The origin of these extreme SFRs is still a contested subject, but there are two key proposed mechanisms; short-lived starbursts induced by major mergers (Narayanan et al., 2010a) and cold mode accretion (Davé et al., 2010; Dekel et al., 2009a).

Short-lived intense starbursts, triggered by gas-rich major mergers, are a plausible explanation for the observed SFRs of SMGs. Although considerably less abundant, Ultra Luminous Infrared Galaxies (ULIRGs; Sanders & Mirabel, 1996) found in the local Universe are often seen as analogues to high-redshift SMGs due to their strongly dust-obscured UV luminosities, high infrared luminosities, and comparable bolometric luminosities (Alaghband-Zadeh et al., 2012; Rowlands et al., 2014), and as such ULIRGs can provide useful insight into the processes at play in SMGs. The SFRs of ULIRGs are thought to be driven by the mechanisms of gas-rich major mergers, where gas is compressed during the coalescence, causing strong inflows towards the centre of the galaxy. These inflows support the star formation activity until the gas is depleted or expelled from the galaxy by AGN or supernova feedback (Casey et al., 2014). It has been demonstrated by hydrodynamic simulations that the observed SEDs of SMGs can be reproduced by this major merger induced starburst activity (Narayanan et al., 2010a,b). Furthermore, the clumpy morphology due to gas compression in ULIRGs appears to match well the morphologies observed in SMGs.

It is not clear, however, that the major merger rate at high-redshift is capable of accounting for the observed numbers of SMGs. Cosmological simulations suggest that only a small fraction of the host haloes of SMGs will undergo major mergers between $2 < z < 4$ (Whitney et al., 2021; Narayanan et al., 2015), and whilst the frequency of mergers increases with redshift, the fraction of extremely star-forming galaxies does not (Bournaud et al., 2014).

The second scenario, of cold mode accretion, posits that the majority of SMGs are orderly rotating disks, with star formation supported by cold gas accretion and minor mergers (Dekel et al., 2009a; Kereš et al., 2009). Hydrodynamical simulations have been used to show that only those starburst galaxies with the most intense SFRs require major mergers, with the rest being stream-fed (Dekel et al., 2009a). Major

mergers play a more dominant role at higher redshifts, but simulations have indicated that the peak brightness of SMGs typically occurs 1 Gyr after the last major merger. This paradigm has successfully reproduced the observed characteristics of SMGs, such as the stellar masses and clustering properties.

SMGs are not only interesting for the potential mechanisms of their formation, but also for the role they play in our overall picture of galaxy evolution. Consider a typical SMG of mass $10^{11}\mathrm{M}_{\odot}$ at $z \sim 2$, with a star formation rate of $500\mathrm{M}_{\odot}\mathrm{yr}^{-1}$ and gas reserves of $5 \times 10^{11}\mathrm{M}_{\odot}$. After approximately 1 Gyr, this SMG will have converted all of its gas into stars and have a stellar mass on the order of $10^{12}\mathrm{M}_{\odot}$. Such stellar masses are typically found in massive elliptical galaxies at $z = 0$. Similarities in their size, number density and clustering properties with quiescent galaxies at low redshifts (e.g., Simpson et al., 2014; Toft et al., 2014) is suggestive of an evolutionary connection (An et al., 2019; Dudzevičiūtė et al., 2020). An explanation for how the population of SMGs became the red and dead population of massive elliptical galaxies observed in the local Universe is underpinned by the process of gas quenching and the mechanism of gas-poor mergers (Oogi & Habe, 2012; Guo & White, 2008; Lapi et al., 2018).

In addition to the reprocessing of UV photons from young stars by the dust in SMGs, Active Galactic Nuclei (AGN) may also contribute to this emission, but studies have shown that approximately 10 per cent of SMGs have associated AGN (Wang et al., 2013). Characterising the properties of the dust in SMGs requires fitting an SED model to observed flux densities in the far-IR to mm wavelengths. Typical SED models are formed from modified black bodies, and can have multiple components to deal with hot and cold dust respectively (Pearson et al., 2013). One such dual temperature SED has the form

$$S_{\nu} = \nu^{\beta}[N_{\mathrm{c}}B(\nu, T_{\mathrm{c}}) + N_{\mathrm{w}}B(\nu, T_{\mathrm{w}})], \tag{1.1}$$

where N is the weighting of the cold and warm components (denoted with subscripts), T is the dust temperature of the two components (denoted by the subscripts) and B_ν is the Planck function. The spectral index, β, controls the dust emissivity, modifying the standard black body equation, and is a property of the crystallinity of the dust grains. Theoretically, we expect $\beta = 2$ for crystalline dielectrics and metals, whilst $\beta = 1$ for amorphous, layer-lattice materials (Tielens & Allamandola, 1987). With an SED model fitted to the observed photometry for an SMG, the total infrared luminosity can be computed from its integral over the rest frame range of 3-1100μm

$$L_{IR} = 4\pi D_L^2(z) \int S_\nu d\nu, \tag{1.2}$$

where D_L is the luminosity distance of the SMG. Typical populations of SMGs have total IR luminosities between $10^{12} - 10^{13} L_\odot$, which are comparable to the values found in ULIRGs (Rowlands et al., 2014; Swinbank et al., 2013). At the extreme, some SMGs have been observed to have $L_{IR} \geq 10^{13} L_\odot$, characterising them as Hyper-Luminous Infrared Galaxies (HyLIRGs; Rowan-Robinson & Wang, 2010; Ivison et al., 2013; Bussmann et al., 2015).

If we make the simplifying assumption that all the IR emission has its roots in the dust reprocessed stellar radiation, the star formation rates of SMGs can be derived from the relation,

$$\log(\mathrm{SFR}) = \log(\mathrm{L_{IR}}) - 43.41, \tag{1.3}$$

(Kennicutt & Evans, 2012) which assumes a Kroupa initial mass function (Kroupa, 2001).

There have been many notable surveys conducted in the sub-millimetre and millimetre wavelength regimes to date. Some examples of these are:

- The Herschel Astrophysical Terahertz Large Area Survey (H-ATLAS; Eales et al., 2010) observed 660 deg^2, at 100 and 160 μm using the Photoconductor Array Camera and Spectrometer (PACS; Poglitsch et al., 2010), and also at 250, 350 and 500 μm using the Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al., 2010). Both of these instruments are on-board the Herschel Space Observatory. Results from this survey are used in Chapter 4 to constrain the SEDs of strongly lensed SMGs.

- The Herschel Multi-tiered Extragalactic Survey (HerMES; Oliver et al., 2012) was carried out using the PACS and SPIRE instruments to map a series of nested fields totalling 380 deg^2, with approximately 270deg^2 having only SPIRE photometry measurements. Again, results from this survey are used in the work of Chapter 4.

- The SCUBA-2 Cosmology Legacy Survey (S2CLS; Geach et al., 2013, 2017) observed an area of 5 deg^2 using the SCUBA-2 camera on the JCMT at 450 and 850 μm. We have made use of some 850μm flux densities in Chapter 4.

- The South Pole Telescope Survey (SPT; Carlstrom et al., 2011), making use of the 10-metre single dish South Pole Telescope, observed an area of 2500 deg^2 at 1.4 and 2.0 mm.

The H-ATLAS and SPT surveys have been particularly successful at discovering hundreds of thousands of SMGs, in part due to the large areas of observation. However, another factor at play in their success is the negative K-correction (Blain, 1996). This correction is applied to convert an object's flux density from the observed frame to that of the rest frame. Measurement's of a galaxy's SED S_ν, at an observed frequency ν, are given by

$$S_\nu = \frac{L_\nu}{4\pi D_L^2(z)}, \tag{1.4}$$

where L_ν is the luminosity of the galaxy at that frequency. Making use of the Rayleigh-Jeans approximation, we can write $L_\nu \sim \nu_0^{2+\beta}$, where $\nu = \nu_0/(1+z)$. The redshift dependence on the luminosity distance scales as $D_L \sim (1+z)^2$ (this approximation is valid at $0.5 < z < 3$) and thus we can rewrite Equation 1.4 as

$$S_\nu \sim \frac{(1+z)^{2+\beta}}{(1+z)^4} \sim (1+z)^{\beta-2}. \tag{1.5}$$

The spectral index β for typical dust SEDs varies between 1.5 and 2.0 (Casey et al., 2014), and so the resulting observed flux density remains constant over a wide range of redshifts. This effect is demonstrated in Figure 1.4, and is more prominent at $1.0 < \lambda < 2.0$mm. The effect of this is that SMGs are readily detectable up to very high redshifts ($z \sim 8$).

There exists a strong lensing bias in the sub-mm regime due to the steep number counts of sub-mm galaxies and the conservation of surface brightness by gravitational lensing. Magnifying this population leads to an increase in the surface density of galaxies above a given flux threshold, making it possible to find lensed sources in wide surveys with a simple cut in flux density above 100 mJy at 500 μm (Blain, 1996; Negrello et al., 2007, 2010; Perrotta et al., 2003). Figure 1.5 shows how this simple flux cut selects lensed SMGs extremely efficiently. One can easily remove any contaminating objects, which are typically local galaxies or radio-loud AGN, via cross correlation with relevant catalogues.

1.2.2 The Evolution of Mass Density Profiles in Early Type Galaxies

There are many open questions in the field of the formation and evolution of early type galaxies (ETGs). Their sizes appear to increase with decreasing redshift at a fixed stellar mass (Carollo et al., 2013; Huertas-Company et al., 2013; Newman

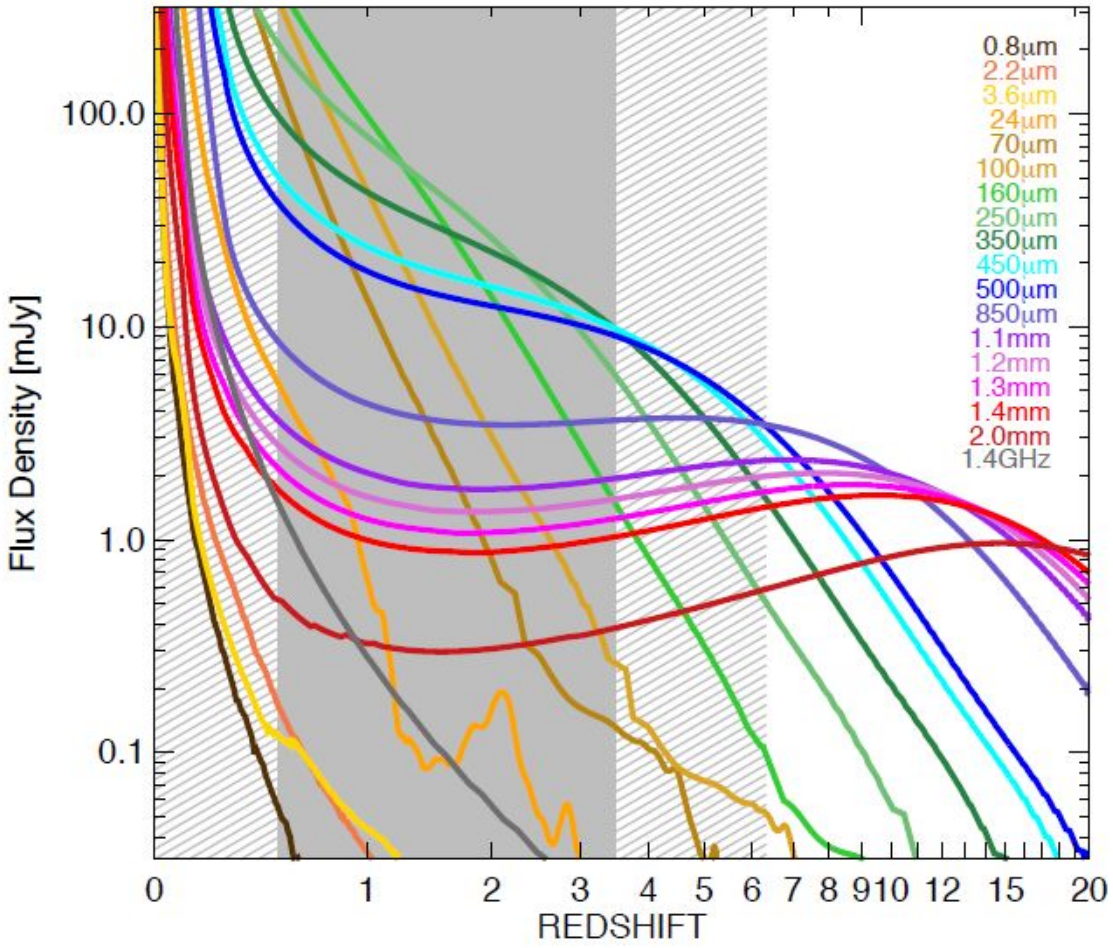


Figure 1.4: Observed flux densities for a typical IR-luminous galaxy as a function of redshift (Casey et al., 2014). In particular, this highlights the almost constant flux densities that SMGs have in the millimetre/sub-millimetre regime over a redshift range of $1 \leq z \leq 10$. Figure reproduced from Casey et al. (2014).

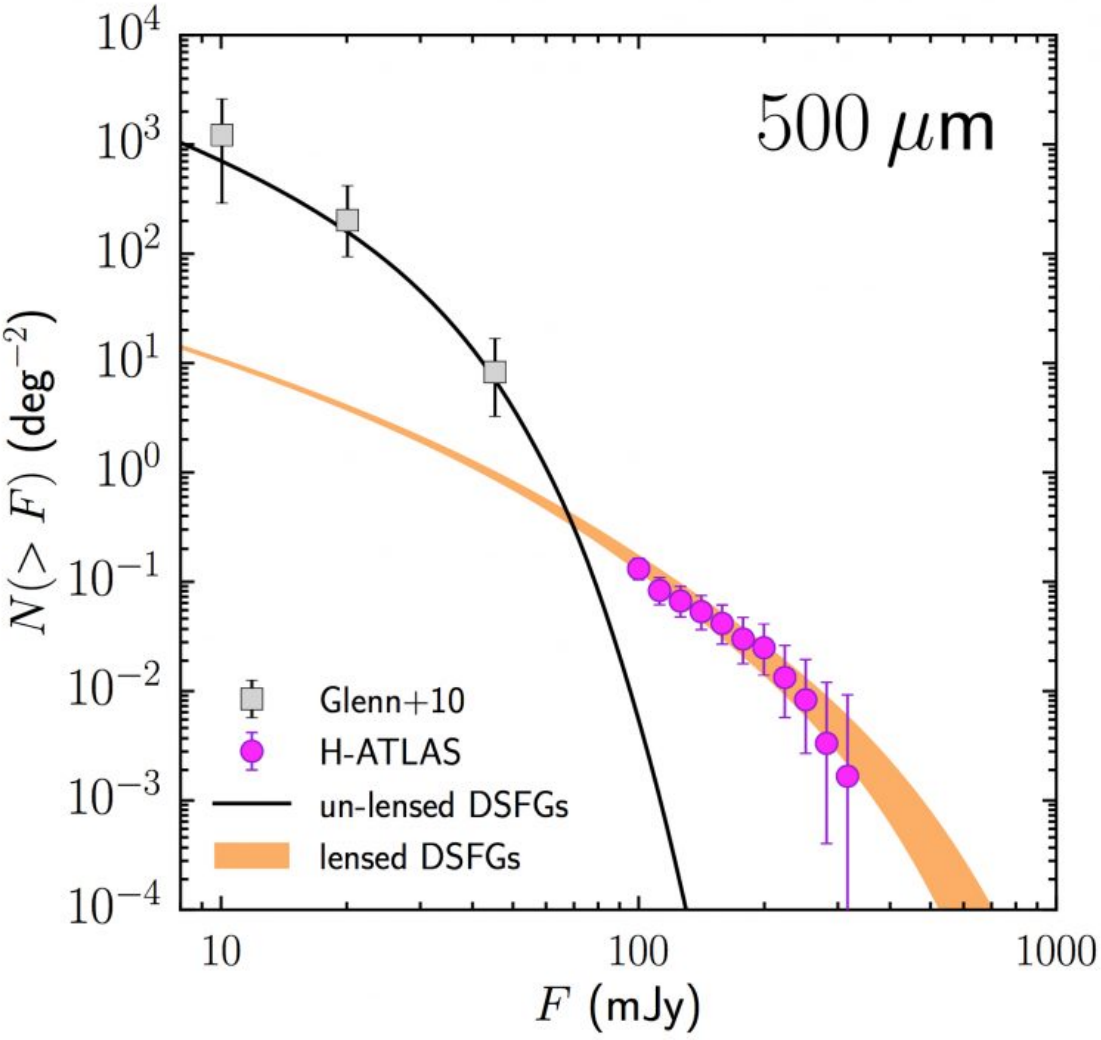


Figure 1.5: Integral number counts of $z > 1$ DSFGs at 500 microns. The lensed sub-mm galaxies contribution to the number density at fluxes > 100mJy is dominant and thus allows for efficient lens selection. Figure reproduced from Negrello et al. (2017).

et al., 2012), the star formation mode seems to differ from that seen in late-type galaxies (Cappellari et al., 2013; Brewer et al., 2012), and their number density has evolved dramatically since $z \sim 2$ (Ilbert et al., 2013). It is thought that major and minor mergers are responsible for the evolution of the structural and morphological characteristics of ETGs, but reproducing these observations remains a challenge for theoretical models (Oser et al., 2012; Remus et al., 2013).

Gravitational lensing, both by itself and in conjunction with other probes, can be effectively used to study the mass profiles of ETGs in the local universe and over cosmological distances (Koopmans et al., 2006a; Auger et al., 2010; Lagattuta et al., 2010). Extending the range of redshift over which we can study the mass profiles of ETGs, as well as simply increasing the number of known strongly lensed galaxies, has been a focus of several surveys in recent years (Lagattuta et al., 2010; Vieira et al., 2013; Gavazzi et al., 2012).

Tentative evidence exists that the average density slope inside the effective radius of the ETG (since both the Einstein radius and velocity dispersion probe regions $< \mathrm{R_{eff}}$), steepens with cosmic time (Bolton et al., 2012; Ruff et al., 2011; Koopmans et al., 2009). These trends are in agreement with some theoretical work (Dubois et al., 2013), but contrasting trends have been found by others (Remus et al., 2013). This dependence of density slope on redshift might be due to a dependence of the slope on other properties of ETGs, such as their mass and size, which are known to grow during their evolution. The density slope of ETGs is closely linked to the concentration of the stellar distribution, with more centrally concentrated galaxies having steeper overall density profiles. It has been shown by Sonnenfeld et al. (2013) that the average total mass density slope of galaxies, for a fixed stellar mass, decreases with redshift. However, (Sonnenfeld et al., 2013) also find that individual ETGs grow with approximately constant density slopes, and so the apparent dependence of

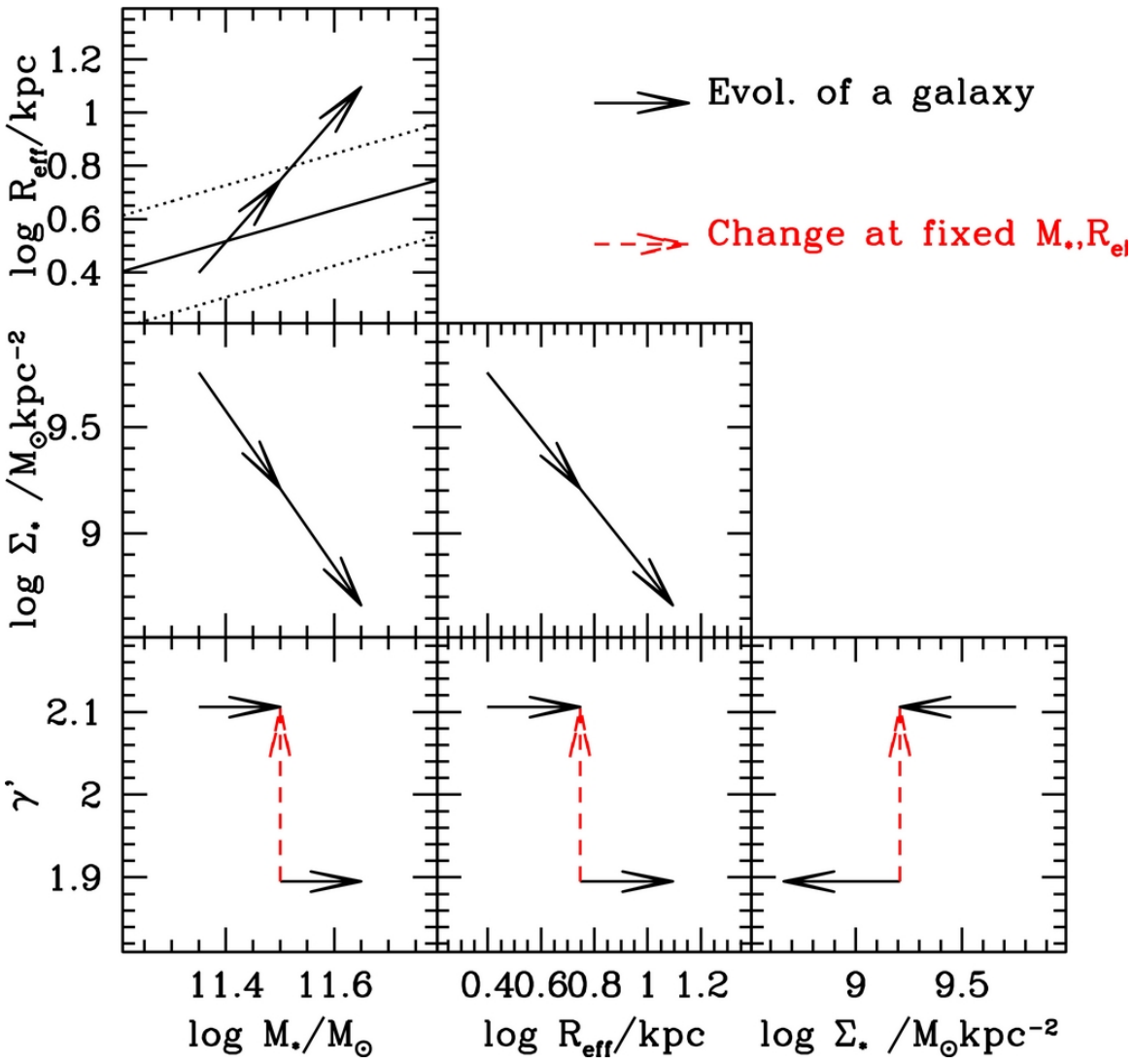


Figure 1.6: Illustration of a galaxy evolution scenario consistent with observations. The solid black arrows show the evolutionary tracks of two galaxies, between $z = 1$ and $z = 0$, in the parameter space of stellar mass, effective radius, effective density, and mass density profile slope. If the density profile is assumed not to change for an individual galaxy, at fixed stellar mass and size, it apparently increases due to the difference in the initial ($z = 1$) stellar densities of the two galaxies. Figure reproduced from Sonnenfeld et al. (2013).

density slope on redshift can be understood in terms of the evolution of the mass-size relation and a dependence on the stellar mass density. Figure 1.6 illustrates a possible evolutionary scenario consistent with these findings; the evolutionary paths of two galaxies, between $z = 1$ and $z = 0$ are shown in the parameter space of stellar mass, effective radius, effective density, and mass density profile. These two galaxies are chosen such that at $z = 1$, one galaxy has the same stellar mass and effective radius as the other at $z = 0$. Mass and size are evolved with the empirical relations found by Sonnenfeld et al. (2013), and the value of the $z = 0$ slope of the density profile based on the correlation with mass and size is set, whilst assuming it does not evolve for the individual galaxies. An apparent evolution of the density slope is observed, due to the difference in the initial stellar density being larger for the more massive galaxy.

1.3 Gravitational Lensing Theory

In this section, we introduce the relevant gravitational lensing theory required to understand the proceeding chapters. We discuss the lens equation, the deflection angle, deflection potential, the formation of multiple images, and explore some basic lens models. The derivations in this section follow closely those presented in Springer (2006).

1.3.1 The Deflection Angle

Given a mass M with spherical symmetry, let us consider the deflection of a ray of light travelling close-by. Assuming the impact parameter ξ is significantly larger than the Schwarzschild radius, the General Relativity prediction for the deflection angle $\hat{\alpha}$

is

$$\hat{\alpha} = \frac{4GM}{c^2\xi}, \tag{1.6}$$

which is twice the prediction from Newtonian gravity. Our condition on the impact parameter implies that $\hat{\alpha} \ll 1$. This leads to the conclusion that the strength of the Newtonian gravitational field is very weak, $\frac{\phi_N}{c^2} \ll 1$.

Einstein's field equations can be linearised in the weak field limit. This allows us to express the deflection angle for a collection of point masses, as a vectorial sum of the individual deflection angles. Consider the mass distribution in 3D with density $\rho(\mathbf{r})$, that we partition into elements of volume dV and mass $dm = \rho(\mathbf{r})dV$. A ray of light, whose trajectory can be described by $(\xi_1(\lambda), \xi_2(\lambda), r_3(\lambda))$, with the coordinates chosen such that in the far field, the incoming light ray propagates along r_3, passes nearby the mass distribution.

For a small deflection angle, the Born approximation tells us that the integral along the null geodesic through the deflection field can be approximated by the integral along a straight line path. The result of this is that we can approximate the actual path of the null geodesic as a straight line with an instantaneous deflection in the lens plane. A system which satisfies the conditions of this approximation is referred to as a geometrically thin lens. In this case, the trajectory does not depend on the affine parameter λ, and so $\boldsymbol{\xi}(\lambda) \approx \boldsymbol{\xi}$, where $\boldsymbol{\xi} = (\xi_1, \xi_2)$ is a two-dimensional vector. The impact vector, with respect to the mass element at $\mathbf{r}' = (\xi_1', \xi_2', r_3')$, is given by $\boldsymbol{\xi} - \boldsymbol{\xi}'$ and does not depend on r_3'.

Consider the mass element $dM_i = \rho(\mathbf{r}_i)dV_i$, then the deflection due to this element is

$$d\hat{\boldsymbol{\alpha}}_{\boldsymbol{i}}(\boldsymbol{\xi}) = \frac{4G}{c^2}\frac{dM_i(\boldsymbol{\xi}-\boldsymbol{\xi}'_i)}{|\boldsymbol{\xi}-\boldsymbol{\xi}'_i|^2}. \tag{1.7}$$

To find the total deflection of the light ray we sum over all the individual elements,

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2}\sum_i \frac{\mathrm{d}M_i(\boldsymbol{\xi}-\boldsymbol{\xi}'_i)}{|\boldsymbol{\xi}-\boldsymbol{\xi}'_i|^2} \tag{1.8}$$

and to make the summation exact, we take the limit

$$\lim_{dM\to 0}\hat{\boldsymbol{\alpha}}(\xi) = \frac{4G}{c^2}\int d^2\boldsymbol{\xi}'\int \mathrm{d}r'_3\rho(\xi'_1,\xi'_2,r'_3)\frac{(\boldsymbol{\xi}-\boldsymbol{\xi}'_i)}{|\boldsymbol{\xi}-\boldsymbol{\xi}'_i|^2}. \tag{1.9}$$

We define a surface mass density containing all the dependence on r'_3, which is the projected mass density on a plane perpendicular to the incident light ray

$$\Sigma(\boldsymbol{\xi}) = \int \mathrm{d}r_3\rho(\xi_1,\xi_2,r_3). \tag{1.10}$$

Substituting this back into Equation 1.9 yields an equation for the deflection produced by a general density distribution

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2}\int \mathrm{d}^2\xi'\Sigma(\boldsymbol{\xi}')\frac{(\boldsymbol{\xi}-\boldsymbol{\xi}'_i)}{|\boldsymbol{\xi}-\boldsymbol{\xi}'_i|^2}, \tag{1.11}$$

assuming that any deviation in the path that an actual photon takes compared to a straight line, is small with respect to the characteristic scale over which the density of the distribution varies significantly. This condition is fulfilled by almost all relevant astrophysical situations and will be assumed to hold throughout this work.

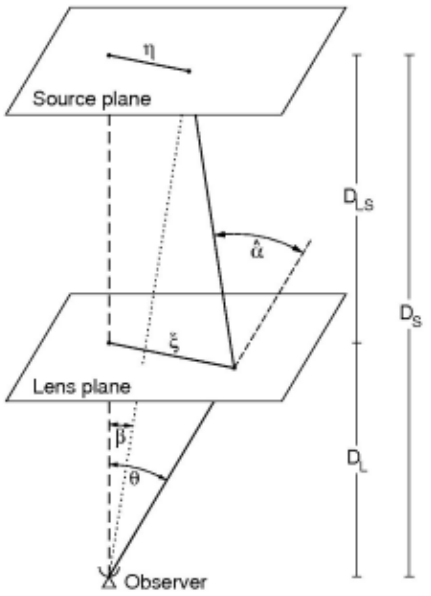


Figure 1.7: Diagram of a gravitational lens system. The lensing mass is at a distance D_L from the observer, with the distance to the background source being given by D_S. The distance between the lens plane and source plane is D_{LS}. The source lies at the angular position $\boldsymbol{\beta}$ in the source plane, at a distance $\boldsymbol{\eta}$ from the optic axis. $\hat{\boldsymbol{\alpha}}$ denotes the deflection angle of the light ray, and $\boldsymbol{\xi}$ the impact parameter of the light ray in the lens plane, with $\boldsymbol{\theta}$ being the apparent position of the lensed source on the sky. Figure reproduced from Schneider et al. (2005).

1.3.2 The Lens Equation

Figure 1.7 depicts a typical situation in gravitational lensing. A mass concentration at angular diameter distance D_L deflects light originating from a source at angular diameter distance D_S. We define the lens and source plane as perpendicular to the optical axis, whose exact definition is unimportant due to the small angles involved in such a system. Using our straight line approximation of the smoothly curved light paths, along with a flat sky approximation, allows us to develop the lens equation.

Consider a source at an angular position $\boldsymbol{\beta}$, lying in the source plane at a distance $\boldsymbol{\eta} = \boldsymbol{\beta} D_S$ from the optical axis. The deflection angle $\hat{\boldsymbol{\alpha}}$ of the light ray originating from the source, having impact parameter $\boldsymbol{\xi} = \boldsymbol{\theta} D_L$ in the lens plane, is given by equation 1.11. The deflection of the light ray causes an observer to receive the light as if it were coming from the angular position $\boldsymbol{\theta}$. Making use of the small angle approximation, the true and apparent position of the source in the sky are related by the lens equation

$$\boldsymbol{\theta} D_S = \boldsymbol{\beta} D_S + \hat{\boldsymbol{\alpha}} D_{LS}, \tag{1.12}$$

where D_{LS} is the angular diameter distance between the source and lens planes. We now divide through by D_S and define the scaled deflection angle,

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_{LS}}{D_S} \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}), \tag{1.13}$$

from which we transform the lens equation to

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}), \tag{1.14}$$

where we have changed the dependent variable of $\boldsymbol{\alpha}$ from $\boldsymbol{\xi}$ to $\boldsymbol{\theta}$.

The physical interpretation of Equation 1.14 is that an image of a source at true position $\boldsymbol{\beta}$ will be observed at angular positions $\boldsymbol{\theta}$ satisfying Equation 1.14. If the lens equation has multiple solutions, then the lens system produces multiple images.

We want to express the scaled deflection angle as a function of the surface mass density, and so we start from Equation 1.11 and invert the definition in Equation 1.13 to find

$$\boldsymbol{\alpha} = \frac{D_{LS}}{D_S}\frac{4G}{c^2}\int \mathrm{d}^2\boldsymbol{\xi}'\Sigma(\boldsymbol{\xi}')\frac{\boldsymbol{\xi}-\boldsymbol{\xi}'}{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2}. \tag{1.15}$$

Using $\boldsymbol{\xi} = D_L\boldsymbol{\theta}$ we can make a change of variables from $\boldsymbol{\xi}$ to $\boldsymbol{\theta}$,

$$\boldsymbol{\alpha} = \frac{1}{\pi}\int \mathrm{d}^2\boldsymbol{\theta}'\frac{D_L D_{LS}}{D_S}\frac{4\pi G}{c^2}\Sigma(D_L\boldsymbol{\theta}')\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}. \tag{1.16}$$

We now define a dimensionless surface mass density, otherwise called the convergence,

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_L\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}}, \tag{1.17}$$

which is the surface mass density as defined before, divided by a critical surface mass density. The critical surface mass density will be discussed in a later section, but it is defined as

$$\Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G}\frac{D_S}{D_{LS}D_L}. \tag{1.18}$$

Using these definitions we rewrite Equation 1.16 as

$$\boldsymbol{\alpha} = \frac{1}{\pi}\int \mathrm{d}^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}. \tag{1.19}$$

1.3.3 The Deflection Potential

We introduce the deflection potential $\psi(\boldsymbol{\theta})$ by projecting the Newtonian potential onto the two-dimensional plane,

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi}\int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\ln(\boldsymbol{\theta}-\boldsymbol{\theta}'). \tag{1.20}$$

For any two-dimensional vector, we have that $\boldsymbol{\nabla}\ln(|\boldsymbol{\theta}|) = \dfrac{\boldsymbol{\theta}}{|\boldsymbol{\theta}|^2}$ and we can use that fact to write the scaled deflection angle as the gradient of the deflection potential

$$\boldsymbol{\alpha} = \boldsymbol{\nabla}\psi, \tag{1.21}$$

$$\begin{aligned}\boldsymbol{\alpha} &= \frac{1}{\pi}\int \mathrm{d}^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\boldsymbol{\nabla}\ln(\boldsymbol{\theta}-\boldsymbol{\theta}') \\ &= \frac{1}{\pi}\int \mathrm{d}^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')\frac{\boldsymbol{\theta}-\boldsymbol{\theta}'}{|\boldsymbol{\theta}-\boldsymbol{\theta}'|^2}.\end{aligned} \tag{1.22}$$

In addition to this, using the identity $\nabla^2\ln(|\boldsymbol{\theta}|) = 2\pi\delta_D(\boldsymbol{\theta})$ we can write the convergence in terms of the Laplacian of the deflection potential,

$$\nabla^2\psi = \frac{1}{\pi}\int \mathrm{d}^2\boldsymbol{\theta}'\kappa(\boldsymbol{\theta}')2\pi\delta(\boldsymbol{\theta}-\boldsymbol{\theta}'), \tag{1.23}$$

$$\nabla^2\psi = 2\kappa(\boldsymbol{\theta}), \tag{1.24}$$

which is the two-dimensional Poisson equation. There is a parallel between these equations and Newtonian gravity. ψ is analogous to the gravitational potential, $\vec{\alpha}$ is an acceleration vector and κ represents the volumetric mass density.

1.3.4 Magnification and Distortion

Gravitational lensing conserves surface brightness, however, the deflection angle is dependent upon the impact parameter. This leads to a modified apparent surface of a lensed object and an apparent change in total brightness. This effect is known as magnification and is driven by the change in solid angle of the image with respect to the source. If we have that $I^{(s)}(\beta)$ describes the distribution of surface brightness for a source, then the distribution observed in the lens plane will be

$$I(\boldsymbol{\theta}) = I^{(s)}[\boldsymbol{\beta}(\boldsymbol{\theta})]. \tag{1.25}$$

As such, the mapping of the coordinates in the source plane to the lens plane is determined by the Jacobian,

$$\begin{aligned} \mathbf{A}(\boldsymbol{\theta}) &= \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}}, \\ &= \begin{pmatrix} \frac{\partial \beta_1}{\partial \theta_1} & \frac{\partial \beta_1}{\partial \theta_2} \\ \frac{\partial \beta_2}{\partial \theta_1} & \frac{\partial \beta_2}{\partial \theta_2} \end{pmatrix}, \end{aligned} \tag{1.26}$$

illustrating how a small displacement in the source plane propagates to the lens plane. We can compute the components of this matrix by taking the derivative of Equation 1.14 with respect to θ,

$$\begin{aligned} A_{ij} &= \frac{\partial \theta_i}{\partial \theta_j} - \frac{\partial \alpha_i(\boldsymbol{\theta})}{\partial \theta_j}, \\ &= \begin{pmatrix} 1 - \frac{\partial \alpha_1}{\partial \theta_1} & -\frac{\partial \alpha_2}{\partial \theta_1} \\ -\frac{\partial \alpha_1}{\partial \theta_2} & 1 - \frac{\partial \alpha_2}{\partial \theta_2} \end{pmatrix}. \end{aligned} \tag{1.27}$$

As was previously shown, the deflection angle is the gradient of the deflection potential and so,

$$\alpha_1 = \frac{\partial\psi(\theta_1,\theta_2)}{\partial\theta_1}, \tag{1.28}$$

$$\alpha_2 = \frac{\partial\psi(\theta_1,\theta_2)}{\partial\theta_2}, \tag{1.29}$$

which we use to rewrite the matrix A_{ij} as

$$\begin{aligned} A_{ij} &= \begin{pmatrix} 1-\frac{\partial^2\psi}{\partial\theta_1^2} & -\frac{\partial^2\psi}{\partial\theta_1\partial\theta_2} \\ -\frac{\partial^2\psi}{\partial\theta_1\partial\theta_2} & 1-\frac{\partial^2\psi}{\partial\theta_2^2} \end{pmatrix}, \\ &= \begin{pmatrix} 1-\psi_{11} & -\psi_{12} \\ -\psi_{12} & 1-\psi_{22} \end{pmatrix}, \end{aligned} \tag{1.30}$$

where the subscripts on ψ correspond to partial derivatives with respect to the vector components, (θ_1,θ_2) respectively. We further manipulate the matrix A_{ij} by introducing the shear $\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\varphi}$, with components

$$\gamma_1 = \frac{1}{2}(\psi_{11}+\psi_{22}), \quad \gamma_2 = \psi_{12}. \tag{1.31}$$

The amplitude of γ provides information about the amount of distortion, whilst its phase φ, tells us the direction of distortion. The factor of two in the phase angle is there because an ellipse transforms to itself under a rotation by π.

Rewriting the A_{11} component as

$$1-\psi_{11} = 1-\frac{1}{2}(\psi_{11}+\psi_{22})-\frac{1}{2}(\psi_{11}-\psi_{22}), \tag{1.32}$$

we see that by employing the definition of the shear components and Equation 1.24 we can write

$$1-\psi_{11} = 1-\kappa-\gamma_1. \tag{1.33}$$

Similar treatment for the other components results in the matrix being written as

$$A_{ij} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}. \tag{1.34}$$

Finally, we rewrite the Jacobian by factoring out $(1-\kappa)$ as,

$$A_{ij} = (1-\kappa) \begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1+g_1 \end{pmatrix}, \tag{1.35}$$

where we have introduced the reduced shear

$$g \equiv \frac{\gamma}{1-\kappa} = \frac{|\gamma|}{1-\kappa} e^{2i\varphi}. \tag{1.36}$$

From Equation 1.25, we see that for a point $\boldsymbol{\theta}_0$ within an image, that corresponds to a point $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\theta}_0)$ within the source, then

$$I(\boldsymbol{\theta}) = I^{(s)}[\boldsymbol{\beta}_0 + \boldsymbol{A}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)], \tag{1.37}$$

where we have used the linearised lens equation. The magnification tensor, $M(\boldsymbol{\theta})$, is defined by taking the inverse of the Jacobean,

$$M(\boldsymbol{\theta}) = \mathbf{A}^{-1}, \tag{1.38}$$

and tells us how to map points from the source plane to the image plane. The ratio of observed fluxes from an image, and it's unlensed source constitutes a definition of the magnification μ, of a source. From Equation 1.37 we have that the magnification

is

$$\begin{aligned} \mu &= \det(\mathbf{M}), \\ &= \frac{1}{\det(\mathbf{A})}, \\ &= \frac{1}{(1-\kappa)^2 - |\gamma|^2}. \end{aligned} \tag{1.39}$$

1.3.5 Critical Curves and Caustics

In a lens system, smooth closed curves can exist, called critical curves, where the determinant of the Jacobian vanishes, $\det A(\theta) = 0$. Mapping these curves to the source plane, with the lens equation, gives us the so-called caustic curves; which are not necessarily smooth. The magnification of an image, $\mu = \frac{1}{\det(A)}$ diverges on these critical curves. Of course, infinite magnifications are unphysical and do not occur in reality. This issue is resolved by moving away from the geometric optics approximation to a wave optics approach. Even so, sources located near caustics can have extremely high magnifications.

Another important point to note is that the location of a source relative to the caustics determine the number of images produced by the lens system. The number of images produced by a source will not change unless it's position moves across a caustic (Schneider et al., 2005).

1.3.6 The Mass Sheet Degeneracy

Upon observing a lensed object, we wish to find a mass distribution for the lens system that is able to account for the observations. It is natural to wonder whether such a model is unique, or if there exist multiple models capable of explaining the data. Letting $\kappa(\theta)$ represent a model that produces a good fit to the observed data, one can

show that there exists a set of lens models with mass distribution

$$\kappa_\lambda(\boldsymbol{\theta}) = (1-\lambda) + \lambda\kappa(\boldsymbol{\theta}) \tag{1.40}$$

that are equally good. The first term in this new mass distribution is equivalent to adding a homogeneous sheet to the distribution, which is where the term *mass sheet degeneracy* gets its name (Falco et al., 1985). The second term is simply a rescaling of the original mass distribution.

The deflection angle produced by the mass distribution, described by Equation 1.40, is given by

$$\boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}) = (1-\lambda)\boldsymbol{\theta} + \lambda\boldsymbol{\alpha}(\boldsymbol{\theta}). \tag{1.41}$$

Inserting this deflection angle into the lens equation yields

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\theta} - \boldsymbol{\alpha}_\lambda(\boldsymbol{\theta}), \\ &= \boldsymbol{\theta} - (1-\lambda)\boldsymbol{\theta} - \lambda\boldsymbol{\alpha}(\boldsymbol{\theta}), \\ &= \lambda(\boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})), \end{aligned} \tag{1.42}$$

which is equivalent to $\dfrac{\boldsymbol{\beta}}{\lambda} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})$, or a rescaling of the unlensed source position, which is not an observable.

Through application of the definition of the deflection potential (Equation 1.20), we see that the scaled deflection angle $\alpha_\lambda(\theta)$ corresponds to a deflection potential

$$\psi_\lambda(\theta) = \frac{1-\lambda}{2}|\theta|^2 + \lambda\psi(\theta), \tag{1.43}$$

satisfying the Poisson equation $\nabla^2\psi_\lambda = 2\kappa_\lambda$.

1.3.7 The Fermat Potential

An alternative approach to explaining lensing phenomena can be found through application of Fermat's principle. This principle states that images will form at locations corresponding to the minimum travel time for the light rays. We can express the additional light travel time, t, due to the deflection of a gravitational lens in terms of the Fermat potential, τ,

$$t = \frac{1+z_L}{c}\frac{D_L D_S}{D_{LS}}\left(\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta})\right), \tag{1.44}$$

where we define τ to be

$$\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\beta})^2 - \psi(\boldsymbol{\theta}). \tag{1.45}$$

There are two distinct components of the time delay induced by gravitational lensing. Firstly, there is the geometric delay due to the difference in path length of deflected light rays, and secondly there is the so-called Shapiro time delay due to the slowing of a light ray in a gravitational field, encoded within the gravitational potential $\psi(\theta)$.

Application of Fermat's principle to the Fermat potential,

$$\nabla\tau(\boldsymbol{\theta};\boldsymbol{\beta}) = 0 \tag{1.46}$$

recovers the lens equation and shows that images will form at the stationary points of the time delay surface. This is generalised to include images formed at the minima, maxima and saddle points of the time delay surface (see Figure 1.8). For smooth, finite mass distributions, there will always be at least one image formed at a minimum of τ.

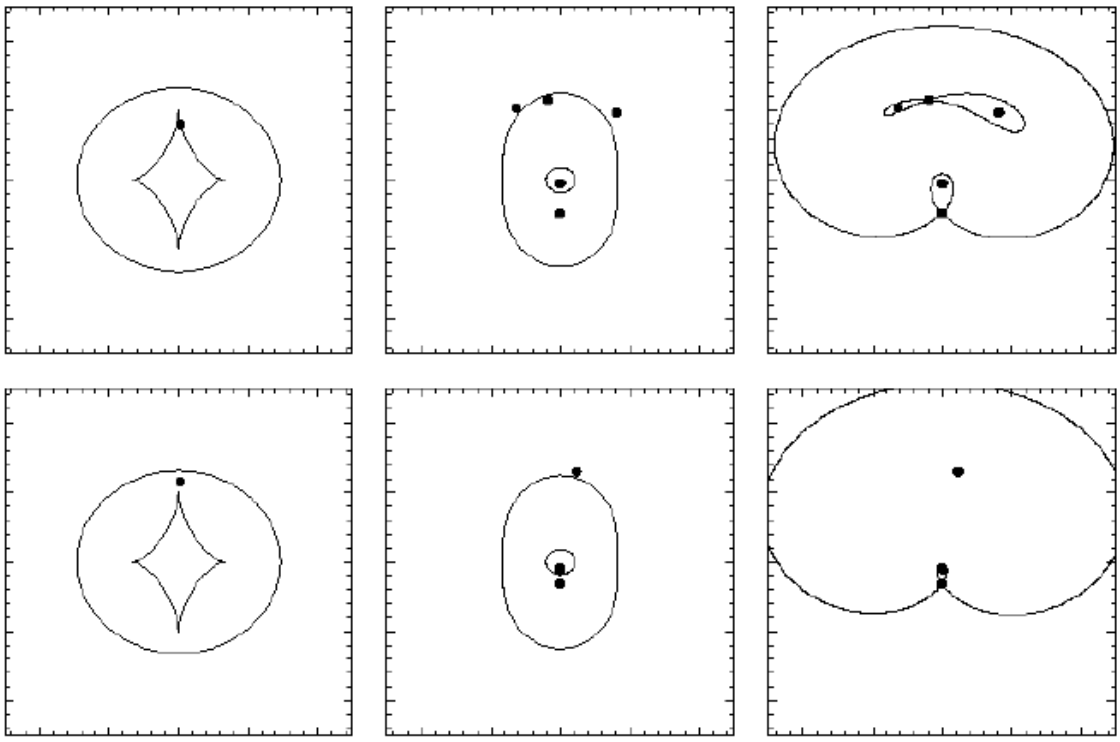


Figure 1.8: Image configurations for two lensing scenarios. The left column panels show the source position and caustics; the middle column panels show the image positions and critical curves; The right column panels show the image positions with contours of the time delay surface overlaid. Figure reproduced from Courbin et al. (2008).

The local magnification can take either positive or negative values, and so images formed at extrema of the time delay surface might have either positive or negative parity. Images outside the tangential critical curve will have positive parity, since both eigenvalues of the lensing Jacobian are positive. In the region between the tangential and radial critical curves, the eigenvalues have opposite signs, and thus the images produced have negative parity (i.e. they are inverted with respect to the source). Inside the radial critical curve, both the eigenvalues of the lensing Jacobian are negative, and so the resulting images have positive parity. Positive parity images form at minima of the time delay surface, whilst negative parity images form at the saddle points and a single positive parity image forms at the maximum (Blandford & Narayan, 1986).

1.4 Basic Lens Models

In general, the deflection angle for a mass distribution must be found through numerical integration. However, for certain mass models, it is possible to find analytic solutions, owing to the exploitation of some symmetries. These simple models are typically spherically symmetric, or rather axisymmetric, since we deal with the projection of the mass distribution onto a plane. In these situations, the lens equation can be recast as an equation in one dimension. We shall consider the properties of several basic lens models here, such as the point mass lens and the singular isothermal sphere.

1.4.1 Axially Symmetric Lenses

An axisymmetric lens has a matter distribution that depends only on the distance from the centre $\Sigma(\xi) = \Sigma(|\xi|)$, provided that we locate our origin at the centre. This implies that $\kappa(\theta) = \kappa(|\theta|)$ and by applying Equation 1.11 we see that

$$\alpha(\theta) = \frac{2\theta}{|\theta|^2} \int_0^{|\theta|} \mathrm{d}\theta' \theta' \kappa(\theta'), \tag{1.47}$$

or expressed in terms of the impact parameter ξ,

$$\begin{aligned} \hat{\alpha} &= \frac{\xi}{|\xi|^2} \frac{4G}{c^2} 2\pi \int_0^{|\xi|} \mathrm{d}\xi' \; \xi' \Sigma(\xi') \\ &\equiv \frac{4GM(|\xi|)}{c^2 |\xi|^2} \xi, \end{aligned} \tag{1.48}$$

where in the last step we have defined $M(|\xi|)$ to be the projected mass inside a circle of radius $|\xi|$. From this we can see that the deflection angle at a point ξ, due to an axisymmetric mass distribution, is simply the deflection for a point mass $M(|\xi|)$. This is analogous to Birkhoff's theorem that states the gravitational force due to a spherically symmetric shell is zero inside it.

Due to the symmetry of this situation, α is collinear with θ and β. Thus, if we describe a source position $\beta = \beta \hat{\mathbf{e}}$, then $\theta = \theta \hat{\mathbf{e}}$ too, where $\hat{\mathbf{e}}$ is a unit vector. This leads to the lens equation becoming one dimensional,

$$\beta = \theta - \alpha(\theta), \tag{1.49}$$

and we can write the lens equation in terms of the deflection angle as

$$\beta = \theta - \frac{D_{LS}}{D_L D_S} \frac{4GM(\theta)}{c^2 \theta}. \tag{1.50}$$

Rearranging equation 1.50, for a source lying on the optical axis (i.e. $\beta = 0$), leads to the expression of the Einstein radius,

$$R_{Ein} = \theta_E = \sqrt{\frac{D_{LS}}{D_L D_S} \frac{GM(\theta_E)}{c^2}}. \tag{1.51}$$

The Einstein radius is an important quantity in gravitational lensing, acting as a characteristic length scale for any strong lens. A source lying on the optic axis, lensed by a circularly symmetric lens will form a ring image of radius θ_E. Additionally, the typical image separation of multiply imaged background sources is twice the Einstein radius, Furthermore, sources within the Einstein radius are likely to be highly magnified, whereas those that lie outside it are not. For many lens models, the Einstein radius also acts as a rough indicator for whether a source will be multiply or singly imaged.

Defining the dimensionless mass $m(\theta)$ over a disc of angular radius θ, and the mean surface mass density $\bar{\kappa}$ inside this radius,

$$m(\theta) = 2 \int_0^{\theta} \mathrm{d}\theta' \; \theta' \kappa(\theta'); \quad \bar{\kappa} = \frac{m(\theta)}{\theta^2}, \tag{1.52}$$

we see that the deflection angle can be written as

$$\begin{aligned} \alpha(\theta) &= \frac{m(\theta)}{\theta} \\ &= \bar{\kappa}(\theta)\theta. \end{aligned} \tag{1.53}$$

Using these definitions, one can write the lens equation as

$$\beta = [1 - \bar{\kappa}(|\theta|)]\theta, \tag{1.54}$$

and computing the Jacobian yields

$$\mathbf{A}(\theta) = [1 - \bar{\kappa}(|\theta|)]\hat{I} - \frac{\bar{\kappa}'}{|\theta|}\begin{pmatrix} \theta_1^2 & \theta_1\theta_2 \\ \theta_1\theta_2 & \theta_2^2 \end{pmatrix}, \tag{1.55}$$

where $\hat{I}$ is the unit matrix and $\bar{\kappa}' = \dfrac{\mathrm{d}\bar{\kappa}}{\mathrm{d}\theta} = \dfrac{2[\kappa(\theta) - \bar{\kappa}(\theta)]}{\theta}$. The determinant of this Jacobian is,

$$\mathrm{det}\mathbf{A} = (1 - \bar{\kappa})(1 - \bar{\kappa} - \theta\bar{\kappa}') \tag{1.56}$$

and the simple form makes for easy exploration of the critical curves for such a system. The critical curves are circles that occur when $1 - \bar{\kappa} = 0$ or when $1 - \bar{\kappa} - \theta\bar{\kappa}' = 0$, i.e. when the magnification tensor diverges.

1.4.1.1 The Point Mass Lens

The point mass lens is characterised by a surface mass density of the form $\Sigma(\boldsymbol{\xi}) = M\delta_D(\boldsymbol{\xi})$ and from Equation 1.11 one finds the deflection angle to be

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \frac{4GM}{c^2}\frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|^2}, \tag{1.57}$$

where the direction of deflection matches what you would expect given the symmetry of this situation. Rewriting this in terms of the scaled deflection angle (Equation 1.13) and making use of the symmetry, we find that

$$\alpha(\theta) = \frac{4GM}{c^2\theta}\frac{D_{LS}}{D_L D_S}. \tag{1.58}$$

Defining the Einstein angle, θ_E,

$$\theta_E = \sqrt{\frac{4GM}{c^2}\frac{D_{LS}}{D_L D_S}} \tag{1.59}$$

and substituting into the lens equation yields,

$$\theta^2 - \beta\theta - \theta_E^2 = 0 \tag{1.60}$$

from which we can find the solutions

$$\theta = \frac{\beta}{2} \pm \sqrt{\frac{\beta^2}{4} + \theta_E^2}. \tag{1.61}$$

The two solutions to this equation correspond to two separate positions on the sky where images will form of the lensed object.

1.4.1.2 The Singular Isothermal Sphere

The singular isothermal sphere is of particular interest as a mass model, since it yields the flat rotation curves we observe in spiral galaxies. The density distribution is characterised by one number, its velocity dispersion σ_v,

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2}. \tag{1.62}$$

This model describes a distribution of self-gravitating particles whose velocities are Maxwellian. There are two key problems with this distribution: The density diverges at $r = 0$ and the total mass diverges as $r \to \infty$. Nevertheless, it is a useful model to study.

By projecting Equation 1.62 along the line of sight, we calculate the surface mass density $\Sigma(\xi)$ to be,

$$\begin{aligned} \Sigma(\xi) &= \int_{-\infty}^{\infty} \mathrm{d}r_3 \; \rho\left(\sqrt{\xi^2 + r_3^2}\right), \\ &= \frac{\sigma_v^2}{2G\xi}. \end{aligned} \tag{1.63}$$

Computing the deflection angle using Equation 1.48,

$$\begin{aligned}\hat{\alpha}(\xi) &= \frac{\xi}{|\xi|^2}\frac{4G}{c^2}2\pi\int_0^{\xi}\mathrm{d}\xi'\xi'\Sigma(\xi'),\\ &= \frac{\xi}{|\xi|^2}\frac{4G}{c^2}2\pi\int_0^{\xi}\mathrm{d}\xi'\xi'\frac{\sigma_v^2}{2G\xi'},\\ &= \frac{4\pi\sigma_v^2}{c^2}.\end{aligned} \tag{1.64}$$

Through application of the lens equation, one finds the Einstein radius, θ_E, for this system to be

$$\theta_E = \frac{4\pi\sigma_v^2}{c^2}\frac{D_{LS}}{D_S}. \tag{1.65}$$

From the definition of the dimensionless surface mass density in Equation 1.17, one finds that

$$\kappa(\theta) = \frac{\theta_E}{2|\theta|}; \quad \bar{\kappa}(\theta) = \frac{\theta_E}{|\theta|}, \tag{1.66}$$

which leads us to the conclusion that the critical curve for this system satisfies $\theta = \theta_{\mathrm{E}}$.

1.4.2 Elliptical lens models

A useful generalisation of the Singular Isothermal Sphere model described in Section 1.4.1.2 is the so-called Singular Isothermal Ellipsoid (SIE). Lens galaxies are typically not spherically symmetric, and so being able to account for that fact with our lens models is important. The SIE model is defined through transforming the radial coordinate, r of the SIS model to $r \to R = \sqrt{x^2 + y^2/q^2}$, where q is the axis ratio of the lens. The deflection angles in this case can be expressed as

$$\alpha_x = \frac{\theta_E q}{\sqrt{1-q^2}}\tan^{-1}\left[\frac{\sqrt{1-q^2}x}{R^2}\right], \tag{1.67}$$

$$\alpha_y = \frac{\theta_E q}{\sqrt{1-q^2}} \tanh^{-1}\left[\frac{\sqrt{1-q^2}y}{R^2}\right], \tag{1.68}$$

as given by Keeton (2001).

A further generalisation is to allow the slope of the mass profile to deviate from isothermal (allowing for the evolution of the mass density profile, as discussed in Section 1.2.2). Such models are referred to as elliptical power law models, where the isothermal profile is a special case. The parameterisation of the elliptical power law surface mass density used in this thesis is given by

$$\kappa(x,y) = \frac{3-\alpha}{1+q}\left(\frac{\theta_E}{R}\right)^{\alpha-1}, \tag{1.69}$$

and the resultant deflection angles are calculated numerically, using algorithms such as those presented in Barkana (1998). Extensive use of the SIE and elliptical power law mass model is made throughout this thesis.

1.5 Gravitational Lensing Science

1.5.1 The Einstein Mass

A characteristic mass scale for a strong lens, the Einstein mass $M_{Ein} = M(\theta_E)$, can be computed by rearranging Equation 1.51 for the mass term. Provided that one has spectroscopic redshifts for both the lens and source and a cosmological redshift-distance relation, the relevant distances can be measured, whilst the source's image plane positions provide θ_E. From this, we can obtain a measurement of the total projected mass inside the circular aperture of radius equal to the Einstein radius. The Einstein mass is model independent and thus does not rely upon any assumptions about the mass profile of the lens.

The robustness of the measurement of M_{Ein} makes it ideal for use as a constraint on other astrophysical measures. Stellar dynamics can also be used to obtain a measure of the total mass inside a particular aperture, but this is limited to relatively small scales, typically on the order of the half-light radius of the galaxy, and assumes that the galaxy is dynamically relaxed. Combining this measurement of mass, with that of the Einstein mass, provides two independent measures of the mass enclosed at two distinct scales that can be used to constrain the density profile of the lens galaxy.

Measurements of this kind, making use of the Sloan Lens ACS Survey (SLACS; Bolton et al., 2006), have shown that the total mass density profiles of massive elliptical galaxies are very close to isothermal on average (Koopmans et al., 2006b; Treu & Koopmans, 2004), with low intrinsic scatter. The physical mechanism responsible for this result is not yet fully understood. In the Λ Cold Dark Matter paradigm, the logarithmic slope of the halo mass profile α, is not isothermal (i.e. $\alpha = 2$) but rather $\alpha \lesssim 1.5$. If instead, we consider that mass follows light, then for a de Vaucouleurs profile $\alpha \sim 2.3$ (Shankar et al., 2017). Given that we observe $\alpha \approx 2$, this implies that both the baryons and dark matter contribute a significant fraction of the mass inside the effective radii of massive elliptical galaxies. The implication that the Baryons and dark matter must "know" about each other in order to produce total mass density slopes close to isothermal is sometimes referred to as the "Bulge-halo conspiracy". Recent works have aimed to gain a more detailed understanding of this interplay by employing decomposed mass-profiles in studies of stellar dynamics (Sonnenfeld et al., 2015; Poci et al., 2017; Nightingale et al., 2019).

1.5.2 Measuring The Hubble-Lemaître Constant

The Hubble constant is one of the most important cosmological parameters, with its value setting the age, size and critical density of the Universe. Extremely precise measurements of its value have been made, derived from observations of the Cosmic Microwave Background anisotropies (Planck Collaboration et al., 2020a), finding a value of $H_0 = 67.4 \pm 0.5 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$. Presenting a challenge to the extremely successful ΛCDM model, is the measurement of the Hubble constant using late universe probes, such as gravitational lensing of variable sources, which find values of H_0 that are significantly higher (Birrer et al., 2020; Wong et al., 2020).

In 1964, Sjur Refsdal proposed a method to determine the Hubble constant using the measured time delay of a gravitationally lensed supernova (Refsdal, 1964). As was discussed in Section 1.3.7, images form at the stationary points of light travel time. If we have a sufficiently varying background source, we can measure the difference in arrival times between two images as a time delay

$$\Delta t_{AB} = \frac{(1+z_L)}{c}\frac{D_L D_S}{D_{LS}}\Delta\tau_{AB}, \tag{1.70}$$

where $\Delta\tau_{AB}$ is the difference in the Fermat potential at the two image positions A and B. Given a mass model that predicts $\Delta\tau_{AB}$, and a measurement of the time delay Δt_{AB}, we can infer the time delay distance $D_{\Delta t}$, which is in fact a combination of angular diameter distances:

$$D_{\Delta t} = (1+z_L)\frac{D_L D_S}{D_{LS}}. \tag{1.71}$$

These angular diameter distances can be determined provided we have spectroscopic redshifts for the lens and source, and assume a cosmological model. This ratio of

angular diameter distances depends on the Hubble constant, $D_{\Delta t} \propto H_0^{-1}$, and so this quantity can be used to infer its value. At the time of Refsdal's suggestion, there were no known strong gravitationally lensed sources. Ultimately, the first time-delay measurements of the Hubble constant were made using lensed quasars, which are sufficiently variable, several decades after Refsdal's idea was proposed (Vanderriest et al., 1989). The first resolved multiply-imaged supernova, named SN Refsdal, was discovered in 2014, and lensing models were used to accurately predict when new image components would appear on the sky (Rodney et al., 2016a,b).

In practice, there are several key ingredients to a successful determination of H_0 using this method. An accurate measurement of the time delay value must be made from comparing the multiple light curves of the images. High resolution imaging and spectroscopy are required to constrain the lens galaxy's mass profile. A multi-plane lensing approach, taking into account the weak lensing effects of the surrounding environment, must be undertaken.

There is significant risk of systematic errors when determining the lens mass profile. Whilst image positions are invariant under the mass sheet transformation (see Section 1.3.6) (Falco et al., 1985; Schneider, Peter & Sluse, Dominique, 2013), and its generalisation, the source position transformation (Schneider & Sluse, 2014), the predicted time delays are not. In general, this has the effect of making the mass distribution $\kappa(\theta)$, and its transformed counterpart, $\kappa'(\theta)$ being equally likely given the data. Further information about the lensing system is required to break this degeneracy. A popular method for achieving this is to combine the constraints of independent measures of the mass distribution, such as stellar kinematics, with those of the lensing analysis (Koopmans & Treu, 2003).

1.6 Interferometry and Synthesis Imaging

The maximum achievable resolution in radians, θ, of a typical telescope is given by

$$\theta \approx 1.22\frac{\lambda}{D}, \tag{1.72}$$

where λ is the observed wavelength of light, and D is the diameter of the telescope aperture. Evidently, for a given λ, it is necessary to build larger and larger telescopes to improve this resolution. We can avoid this issue by employing the use of interferometers.

1.6.1 Interferometery

An interferometer is an array of antennae, whose resolving power depends upon the maximum separation between two antennae, or baseline distance B, such that $\theta \propto \frac{\lambda}{B}$. An interferometer does not observe a source's surface brightness $T(l, m)$ directly, where (l, m) are on-sky coordinates; rather, an interferometer measures the complex visibility function $V(u, v)$, which is its Fourier transform,

$$V(u, v) = \int\int dldmA(l, m)T(l, m)e^{-2\pi i(ul+vm)}, \tag{1.73}$$

where $A(l, m)$ is the primary beam profile and (u, v) are the baseline coordinates in the Fourier plane. Assuming that the source is small compared to the primary beam, equation 1.73 implies that a visibility sample $V(u, v)$ contains information on $T(l, m)$ everywhere, and the sample $V(0, 0)$ determines the total flux density of the source. Furthermore, due to the real nature of $T(l, m)$ we get two visibilities for each baseline measurement since $V(-u, -v) = V^*(u, v)$.

Recovery of the true source surface brightness is complicated by the fact that we can only measure $V(u,v)$ at discrete coordinates in Fourier space, determined by the baselines. The result of this is that we can recover a so-called dirty image $T^D(l,m)$,

$$T^D(l,m) = A^{-1}(l,m) \int\int dudv S(u,v) V(u,v) e^{2\pi i(ul+vm)}, \tag{1.74}$$

where $S(u,v)$ is the sampling function in the uv-plane. We can rewrite the dirty image as the convolution of the true source surface brightness with the dirty beam $B(l,m)$, such that,

$$T^D(l,m) = T(l,m) * B(l,m), \tag{1.75}$$

where the dirty beam is the inverse Fourier transform of the sampling function.

The effect of this incomplete sampling of the uv-plane is to produce a side-lobe structure in the dirty beam that depends on the physical configuration of the antennae as well as the weighting scheme (see Section 1.6.2 below) applied to the visibilities. This leads to spatially correlated noise in the image plane that must be dealt with through deconvolution techniques. Side lobe artefacts and correlated noise could in principle bias our ability to accurately recover lens model parameters, as we will explore in Chapter 5.

1.6.2 Synthesis Imaging

When creating images, we must choose a visibility weighting function $W(u,v)$, which modifies the sampling function and thus changes the shape of the dirty beam. A natural choice is to set the weighting function to the inverse of the noise variance of

that visibility, such that the imaging weight, w_i, is given by

$$w_i = \omega_i = \frac{1}{\sigma_k^2} \tag{1.76}$$

where the data weight is given by ω_i, determined during calibration. When gridding the data into the same uv-cell, the weights are summed, resulting cells containing more visibilities being given greater weight; this is known as natural weighting. This weighting scheme has the effect of maximising point source sensitivity and producing the lowest RMS noise in the image, but due to the greater weight given to short baselines the resulting angular resolution is lowered.

Another option for $W(u, v)$ is the so-called uniform weighting function, where the weight is set according to the inverse of the local density of points in the uv-plane (and zero for empty cells). Firstly, the data weights are calculated identically to the natural weighting scheme. The data is then gridded into uv-cells and re-weighted to achieve uniform imaging weights. Letting ω_i be the inverse variance of the RMS noise on the visibility data, the imaging weights are given by

$$w_i = \frac{\omega_i}{W_k}, \tag{1.77}$$

where W_k is the weight on the gridded data. This means that the sum of weights in a given occupied cell is constant, more weight is given to longer baselines, and thus angular resolution is enhanced. It does have the drawback however of degrading the point source sensitivity.

A third choice of $W(u, v)$ is Briggs weighting, which is a variant of uniform weighting capable of avoiding giving too much weight to uv-cells that would otherwise have

low natural weightings. The imaging weights are given by

$$w_i = \frac{\omega_i}{1 + W_k f^2}, \tag{1.78}$$

where W_k is defined in the same way as in the uniform weighting scheme, and

$$f^2 = (5 \times 10^{-R})^2 \frac{\sum_i \omega_i}{\sum_k W_k^2}. \tag{1.79}$$

Briggs weighting is able to smoothly vary between behaving like uniform weighting or natural weighting (i.e. between maximum resolution and maximum point source sensitivity), via adjusting the so-called Robustness parameter, R. A common choice that provides a good compromise between sensitivity and resolution is to set $R = 0$. Further processing of visibilities via uv-tapering can be performed, but we do not consider the effects of this in this work.

The dominant deconvolution algorithm in radio astronomy is CLEAN, originally developed by Högbom (Högbom, 1974). At its core lies the a priori assumption that the sky brightness distribution $T(l, m)$, is a collection of point sources. The CLEAN algorithm begins by initialising a residual map to the dirty image, as well as creating a CLEAN component list. The highest peak in the residual map is identified as a point source, and some fraction is subtracted by scaling the dirty beam by a particular gain. This point source is added to the CLEAN component list and the cycle repeats until a stopping criterion, typically a threshold value or number of iterations, has been reached. The last step in the process is to make a model image from all the point source CLEAN components. The point sources are convolved with an elliptical Gaussian, or CLEAN beam, that has been fitted to the main lobe of the dirty beam. The remaining noise in the residual map is then added back to the CLEAN image.

1.7 Semi-Linear Gravitational Lens Inversion

At the heart of much of the work in this thesis is the inference of lens model parameters and the reconstruction of the unlensed background source. In particular, Chapter 2, Chapter 3 and Chapter 5 deal with the modelling of image plane data, and we will discuss the methodology used to facilitate this in this section. The modelling of interferometer datasets is relevant to Chapter 4 and Chapter 5 and will be discussed in the following section. There exists a multitude of methods for inferring lens model parameters and performing source reconstruction, some of which rely on fitting parametric source light distributions and others that allow for much more flexibility in the source light distribution, such as shapelet and pixelised source plane methods (Birrer et al., 2015; Tagore & Jackson, 2016).

The method of Semi-Linear gravitational Lens Inversion (SLI) was introduced by Warren & Dye (2003). This method makes use of a pixelised source light distribution and a parametric mass distribution for the lens mass. The key advantage of Semi-Linear Inversion is that the step of inverting an image for a given mass distribution in order to obtain the deconvolved source light distribution and its uncertainties is a linear one. Finding the parameters to model the mass distribution is still a non-linear problem, hence the name semi-linear inversion. The increased efficiency of SLI makes the task of optimising the lens model simpler and faster.

We first consider the case where the figure of merit of choice is $G = \chi^2_{im}$, computed from a model image and the observed image, and there is no regularisation applied. The essence of the problem is to find the values of the source pixels that, given a mass distribution, minimise the merit function. We label pixels in the source plane $i = 1, \ldots, I$ and pixels in the image plane $j = 1, \ldots, J$. Note that there is no restriction on the pixelisation scheme used in the source plane. We assume that the counts in

the image pixels are purely due to the lensed image of the source and that the data in each image pixel is independent, i.e. no covariance between pixels.

The inversion process works as follows: for a given mass distribution, we form the image by ray tracing and convolution with the PSF, of each pixel in the source plane with unit counts. In other words, we compute the counts in the jth image pixel due to the ith source pixel f_{ij}. This results in I images, one for each source pixel. We must now combine these I images, scaled by a factor s_i in such a way as to minimise G. The scalings constitute the deconvolved intrinsic source surface brightness distribution. Writing the merit function as

$$G = \chi^2_{im} = \sum_{j=1}^{J} \left(\frac{\sum_{i=1}^{I} s_i f_{ij} - d_j}{\sigma_j} \right)^2 \tag{1.80}$$

where d_j and σ_j are the counts and 1-σ uncertainty of the jth observed image pixel. Minimising this with respect to the source terms yields a set of I simultaneous equations

$$\frac{1}{2}\frac{\partial G}{\partial s_i} = \sum_{j=1}^{J} \left(\frac{f_{ij} \sum_{k=1}^{I} s_k f_{kj} - f_{ij} d_j}{\sigma_j^2} \right) = 0 \tag{1.81}$$

where we have artificially inserted a fact or of $\frac{1}{2}$ for reasons that will become obvious shortly. Rewriting these equations on matrix form

$$\mathbf{FS} = \mathbf{D}. \tag{1.82}$$

$\mathbf{S}$ is a column matrix with I entries representing the scalings s_i, which we wish to solve for. $\mathbf{F}$ is a symmetric $I \times I$ matrix, whose elements are $F_{ik} = \sum_{j=1}^{J} \frac{f_{ij} f_{kj}}{\sigma_j^2}$. Lastly, $\mathbf{D}$ is a column matrix with I elements $D_i = \sum_{j=1}^{J} \frac{f_{ij} d_j}{\sigma_j^2}$. We can solve for the counts in

the source pixels by matrix inversion

$$\mathbf{S} = \mathbf{F}^{-1}\mathbf{D} \tag{1.83}$$

which has eliminated the need for variational optimisation of the source. We now wish to compute the covariance matrix for the source pixels in order to find the errors. We note that

$$F_{ik} = \frac{1}{2}\frac{\partial^2 G}{\partial s_i \partial s_k} \tag{1.84}$$

is the Hessian matrix multiplied by a factor of a half. This is the reason for multiplying Equation 1.81 by one half. We can show that $\mathbf{C} = \mathbf{F}^{-1}$ is the covariance matrix of $\mathbf{S}$.

Under our assumption of independent image pixels, the covariance between source pixels i and k is given by

$$\sigma_{ik}^2 = \sum_{j=1}^{J} \sigma_j^2 \frac{\partial s_i}{\partial d_j}\frac{\partial s_k}{\partial d_j} \tag{1.85}$$

which can be rewritten as

$$\sigma_{ik}^2 = \sum_{j=1}^{J} \sigma_j^2 \sum_{l=1}^{I} C_{il}\frac{f_{lj}}{\sigma_j^2} \sum_{m=1}^{I} C_{km}\frac{f_{mj}}{\sigma_j^2} = C_{ik}. \tag{1.86}$$

Since we have fixed the mass model parameters, the covariance matrix of the source pixels has been computed during the inversion process. It is not in fact the complete solution for the errors, since we have not varied the mass model, but the so-called "source covariance matrix" is useful in exploring different solutions.

One important difference between SLI solutions to this problem and the maximum entropy method, is that we place no constraints on the counts in the source pixels. In principle, they can be negative, due to being below the mean sky count, whereas in the maximum entropy solution, they must be positive.

To complete the solution, we must search the mass model parameter space, minimising χ^2_{im} at each point using SLI to find the global minimum.

The full solution to the errors is more complicated than before, due to the additional non-linear mass parameters, L (labelled m_l). We must form the $(I+L)\times(I+L)$ curvature matrix, but note that we have already computed the majority of terms from the computation of $\mathbf{F}$ at the global minimum. The remainder of the terms can be found through measurement of the shape of the χ^2_{im} surface at the global minimum. The full covariance matrix for the mass and source parameters is therefore the inverse of this curvature matrix.

1.7.1 Regularised Linear Inversion

This method can be readily extended to include a linear regularisation term that avoids the problem of over-fitting noisy solutions (Warren & Dye, 2003). We can preserve the linearity of the minimum χ^2 approach by forming our regularisation term G_L from a linear combination of the source scalings $s_i s_k$,

$$G_L = \sum_{i,k} a_{ik} s_i s_k. \tag{1.87}$$

We write the merit function containing the regularisation term as,

$$G = \chi^2_{im} + \lambda G_L, \tag{1.88}$$

where λ is a scalar weighting of the regularisation. This term acts like a prior, favouring solutions where the reconstructed source has a smooth distribution. The source pixel brightness is given by

$$\mathbf{S} = [\mathbf{F} + \lambda \mathbf{H}]^{-1}\mathbf{D}, \tag{1.89}$$

where $\mathbf{H}$ is termed the regularisation matrix and can be computed as follows

$$H_{ik} = \frac{1}{2}\frac{\partial^2 G_L}{\partial s_i \partial s_k}. \tag{1.90}$$

If we choose for example, $G_L = \sum_{i=1}^{I} = s_i^2$, then $\mathbf{H}$ is simply the identity matrix, which is termed "zeroth-order" regularisation. Other forms of regularisation include "gradient" and "curvature" forms, which take into account differences between neighbouring pixel counts. Details of these schemes can be found in (Warren & Dye, 2003). The computation of the covariance matrix is complicated in this case by that fact that it is no longer true that the inverse of the full curvature matrix is the full covariance matrix. More recently, the method of regularised Semi-Linear Gravitational Lens Inversion has been placed within a Bayesian framework by Suyu et al. (2006), and it is the maximisation of the Bayesian evidence that is used throughout this thesis. Finding the optimal lens model parameters, to maximise this Bayesian evidence, was achieved using MultiNest (Feroz et al., 2009) for the work in Chapter 2 and Chapter 3, whilst dynesty (Speagle, 2020) was used for the work in Chapter 4 and Chapter 5.

1.7.2 Adaptive Semi Linear Inversion

Previously, it was mentioned that SLI places no restriction on the pixelisation scheme used in the source plane. Adaptive semi-linear inversion makes use of an h-means clustering algorithm to produce a pixelisation that adapts to the magnification of the lens model (Nightingale & Dye, 2015). Adaptive SLI overcomes many of the disadvantages

associated with using a fixed source pixel grid, such as the ability to place smaller pixels in regions of the source plane with higher magnification. Additionally, there is no need to define the source plane grid in advance, and so the issue of unconstrained source pixels is avoided (Nightingale & Dye, 2015).

The essence of how adaptive-SLI works is the way in which it allocates traced image pixels to their source plane pixels. The spatial coordinates of the J traced image pixels are the input to the h-means clustering algorithm (Hartigan & Wong, 1979), giving us a set of h-clusters. These h-clusters are the regions in the source plane to which a subset of image pixels belong. These h-clusters are defined through their centre coordinates, found by minimising E,

$$E = \sum_{i=1}^{I} e_i = \sum_{i=1}^{I} \sum_{k=1}^{K} r_k^2 \tag{1.91}$$

where e_i are the cluster energies. The cluster energy is the sum in quadrature of the distances to each of it's associated image pixels from the cluster's centre. The algorithm works by first allocating traced image pixels to their nearest cluster centre, and then recomputing the set of cluster centres for these new assignments. This continues until the cluster centres no longer move, or a set number of iterations have occurred. The resultant pixelisation of the source plane then depends on the spatial distribution of the traced image pixels, and thus the magnification of the lens model, unlike a regular Cartesian grid. The shapes of these clusters are completely unrestricted, and thus do not suffer from the biasing effects on the lens model of some schemes.

In practice, it is the centres of sub-grid pixels that are fed into the h-means algorithm. For example, we may split a pixel into a 4×4 grid and use the centres of these pixels during clustering and inversion. Whilst this increases the computational workload, it helps to address certain aliasing issues that arise (Nightingale & Dye,

2015). Furthermore, the image is masked to remove the background noise from the image. This means that the source plane maps only to those points contained within the mask and helps to reduce the number of coordinates being fed into the clustering algorithm.

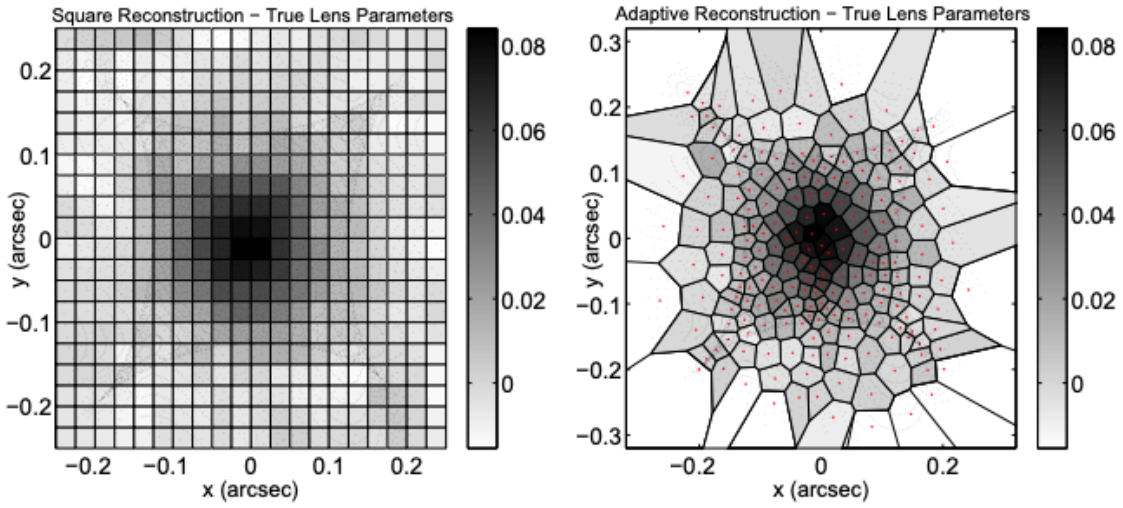


Figure 1.9: A comparison between square and adaptive SLI pixelisation schemes (Nightingale & Dye, 2015)

One of the key disadvantages with using the regular grid is that it is set up before the inversion takes place. This leads to large variations in magnification between source pixels and the possibility that pixels on the outer edge of the source not mapping to any of the image pixels. These unmapped pixels are then only constrained by regularisation.

1.7.3 The semi-linear inversion method in the uv-plane

The standard image plane approach of the semi-linear inversion method makes use of a pixelised source plane. For a given lens model, the image of each pixel is formed and the linear superposition of these images that best fits the data determines the source

surface brightness distribution. Analogously to the image plane version, when working with interferometer visibility data, a model set of visibilities is formed for an image of each source pixel. The linear combination of these model visibilities determines the source surface brightness distribution for a particular lens model.

In this thesis, we have used the source inversion method implemented within `PyAutoLens` (Nightingale et al., 2021), which is based on the operator approach described within Powell et al. (2020), for the modelling of interferometric datasets (see Chapter 4 and Chapter 5). An interferometer visibility dataset $\boldsymbol{d}$ is comprised of samples of complex visibilities. The surface brightness in the source plane is given by the vector $\boldsymbol{s}$, with each element corresponding to the surface brightness of a source plane pixel. The parameterised projected surface mass density of the lens model is given by the vector, $\boldsymbol{\eta}$ and the mapping of the source light $\boldsymbol{s}$ to the image plane is described by the operator $\mathbf{L}(\boldsymbol{\eta})$. The sky brightness is therefore simply $\mathbf{L}(\boldsymbol{\eta})\mathbf{s}$. The response of an interferometer is encoded into the operator $\mathbf{D}$, which performs the Fourier transforms to convert the pixelised sky brightness distribution into a set of complex visibilities. The observed data $\boldsymbol{d}$, can therefore be described by the combination of these effects:

$$\boldsymbol{d} = \mathbf{DL}(\boldsymbol{\eta})\mathbf{s} + \mathbf{n}. \tag{1.92}$$

Assuming uncorrelated Gaussian noise $\boldsymbol{n}$ in the observed visibility data, the noise covariance can be represented by the diagonal matrix $\mathbf{C}^{-1}$. The uncertainties on visibilities are often set to arbitrary values, or estimated from poorly understood antenna and receiver properties. This can lead to problems when accurate uncertainties are required, such as for performing visibility data lens modelling. Using a similar method to Dye et al. (2018) to determine the 1σ uncertainties on the visibilities, we used the `CASA` task `statwt` to empirically measure the visibility scatter computed over all

baselines, ensuring reasonable uncertainties on all visibilities.

Combining equation 1.92 with the set of model visibilities $\mathbf{DL}(\boldsymbol{\eta})\boldsymbol{s}$ allows us to write the χ^2 statistic as

$$\chi^2 = (\mathbf{DL}\boldsymbol{s} - d)^T \mathbf{C}^{-1} (\mathbf{DL}\boldsymbol{s} - d). \tag{1.93}$$

With the addition of a prior on the source denoted by the operator $\mathbf{R}$ and with the regularisation strength given by λ_s, Powell et al. (2020) show that we can write the regularised least-squares equation as

$$\left[(\mathbf{DL})^T \mathbf{C}^{-1} \mathbf{DL} + \lambda_s \mathbf{R}^T \mathbf{R}\right] \boldsymbol{s}_\mathrm{MP} = (\mathbf{DL})^T \mathbf{C}^{-1} \boldsymbol{d} \tag{1.94}$$

where the maximum a posteriori source inversion matrix (i.e., the solution matrix for $\boldsymbol{s}_\mathrm{MP}$) is given by the quantity in square brackets in equation 1.94. This linear system of equations is in principle straightforward to solve, but becomes extremely memory intensive for large numbers of visibilities and/or large numbers of source plane pixels. For this reason, a direct Fourier transform is replaced by a non-uniform fast Fourier transform, constructed out of a series of operators. This substitution results in a modified version of equation 1.94, where the solution for $\boldsymbol{s}_\mathrm{MP}$ is given by a series of operators evaluated by use of an iterative linear solver (see Powell et al., 2020; Nightingale et al., 2021, for more details on this methodology and the specific implementation used here).

To find the optimal lens model parameters with `PyAutoLens`, we used the nested sampling algorithm `dynesty` (Speagle, 2020) to maximise the Bayesian evidence as derived within Suyu et al. (2006). We adopted a gradient regularisation scheme, analogous to that which is described in Warren & Dye (2003), with a constant weight for the source plane due to its simplicity and to not add more computational costs

to an already expensive procedure. Other techniques, such as those that use multiple regularisation weights adapted to the surface brightness in the source plane, require introducing more free parameters to be fitted, and thus increase the complexity of the problem. When dealing with data sets containing many visibilities, such a process could add days of analysis time. We first reconstructed the background source using a source plane pixelisation adapted to the lens magnification. This magnification-based fit was then used to initialise a new search of parameter space with a source plane that adapted to the brightness of the reconstructed source (Nightingale & Dye, 2015; Nightingale et al., 2018). The mass model parameters, along with the source plane parameters (regularisation, number of source plane pixels), were fully optimised throughout the lens modelling procedure.

1.8 Machine Learning

Deep neural networks are being used widely in astronomy, in topics ranging from measurements of galaxy morphology (Tohill et al., 2021), detection of strong lensing events (Jacobs et al., 2017; Lanusse et al., 2017a; Cheng et al., 2020) and predictions of the merger history of high redshift galaxies (Whitney et al., 2021). The major benefits of deep neural networks are that they are naturally suited to the analysis of image-plane data, and that once a network has been trained, the computational cost of performing a classification or regression task is extremely low compared to other traditional methods.

1.8.1 Neural Networks

Fundamentally, neural networks rely on simple principles analogous to many other regression techniques. The desired output of the network is a prediction, $\vec{y}_{pred}$, given an input $\vec{x}$ and model parameters $\vec{\alpha}$. In order to optimise this non-linear model, $F(\vec{x}|\vec{\alpha}) = \vec{y}$ and find the best fitting set of model parameters $\hat{\alpha}$, we must minimise the error in $F(\vec{x}|\hat{\alpha}) = \vec{y}_{pred}$. This is achieved with the use of a loss function, $\mathcal{L}(\vec{y}, \vec{y}_{pred})$, and typically requires a large set of known measurements $(\vec{x}, \vec{y})$, known as the training set. Analogously, if we wanted to perform linear regression, our model F would represent a straight line, and our loss function could be represented as the χ^2 function.

We can represent a deep neural network as a graph of multiple layers, each consisting of many neurons (see Figure 1.10). In this case, the function F and the free model parameters are encapsulated by the interconnected nodes, called hidden layers. The inputs are passed from neuron to neuron through the layers of the network, through the connections, as shown in Figure 1.10. Popular networks in use today, so-called deep neural networks, are constructed from several layers, each consisting of hundreds or thousands of layers. Furthermore, these neurons are typically connected to each neuron in the following layer, creating a fully-connected network (Figure 1.10 is such an example).

The optimisation of the network is achieved by tuning the weights, $\vec{w}_{i,j}$, and biases, $b_{i,j}$, of each i,jth neuron. Applied to each neuron is an activation function f, which controls whether a specific neuron produces an output. Accordingly, we can describe the output of a given neuron as

$$\vec{x}' = f(\vec{w}_{i,j}.\vec{x} + b_{i,j}) \tag{1.95}$$

The chosen activation function may depend upon the specific application being con-

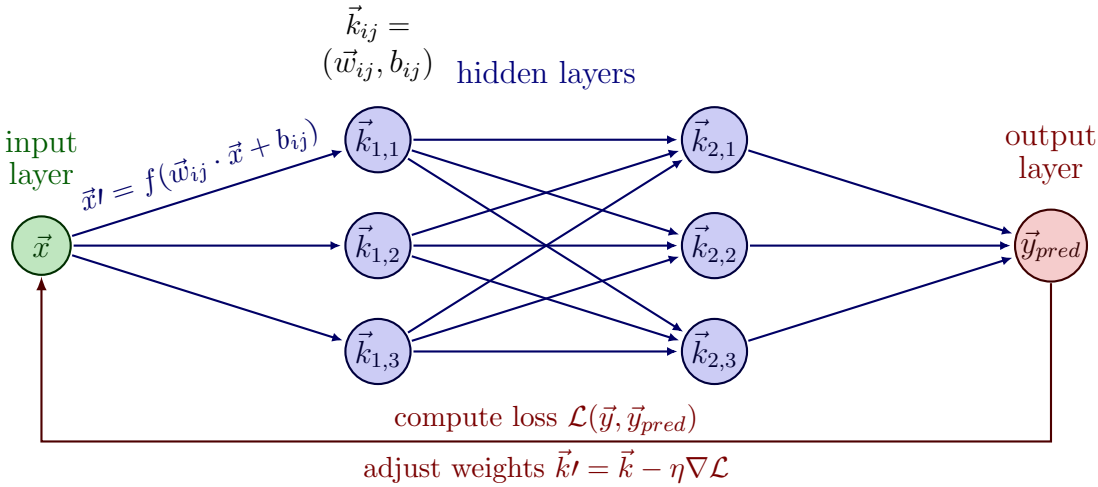


Figure 1.10: Schematic of a neural network model. The input $\vec{x}$ (green) is passed through a series of hidden layers (blue), each applying a function to $\vec{x}$. The output of the hidden layers is combined to predict the output, $\vec{y}$ (red).

sidered, but a commonly used example is the Rectified Linear Unit (ReLU; Nair & Hinton, 2010), defined as

$$f(x) = \text{ReLU(x)} = \max(0, x). \tag{1.96}$$

The popularity of ReLU is owed to its simplicity to compute and the fact that it does not suffer from the problem of vanishing gradients (Glorot & Bengio, 2010). Typically, the output later uses the softmax activation function (Goodfellow et al., 2016) when dealing with classification tasks.

Finally, the output of the model $\vec{y}_{pred}$ and the true values $\vec{y}$ are compared, with their difference encoded by the loss function $\mathcal{L}$. A common choice of loss function for classification tasks is that of the categorical cross entropy loss function. The neuron weights are then adjusted, from the last layer to the first, in a process called back-propagation, using the gradient of the loss function to minimise loss, such that

$$\vec{w}'_{i,j} = \vec{w}_{i,j} - \eta\vec{\nabla}\mathcal{L}(\vec{y}, \vec{y}_{pred}), \tag{1.97}$$

where η is the so-called learning rate. η is typically chosen to be small such that the training process is stable, but not so small that training takes an inordinate amount of time.

1.8.2 Convolutional neural networks

Image analysis is a particularly common task in astronomy, and one that deep learning methods excel at. Convolutional Neural Networks (CNNs; Fukushima, 1980) are a subset of deep neural networks that were developed specifically to analyse imaging data.

The key difference between CNNs and the more simplistic networks discussed in the previous section, is that in a CNN the neurons are represented by convolutional kernels, not simply weights. An input image is convolved with each kernel, producing its response image. This response represents the features picked out by a particular kernel and can then be used later in the tasks of classification or regression.

Each kernel has dimensions $N \times N$, with each element having its own weight, and a single bias is applied to the entire kernel. Resultantly, these filters add $N^2 + 1$ free parameters to the model that must be optimised. Every layer of the network can be made up from multiple kernels, and the number of these will depend on the CNN architecture and the specific application being considered. The filters produce varying responses to certain features in the image, such as sharp edges, patterns, or colour gradients. As the image is fed through the network, the features detected by filters build upon earlier features, capturing more complex information from the image.

Once the final convolutional filter has been applied to an input image, typically a flattened vector is formed and this is passed into a fully connected network (as described in the previous section) for the task of classification or regression.

CNNs are employed in both Chapter 2 and Chapter 3 of this thesis, for both the tasks of regression and classification; a detailed description of the relevant network architecture is provided in the respective chapters, and so we refrain from repeating such an illustration here.

1.9 Outline

In Chapter 2 we compare two different approaches to lens modelling, parametric fitting and machine learning, testing these techniques against a series of increasingly

complex and realistic simulated images. In addition to this, we propose two methods for combining these techniques to produce a hybrid, fully automated approach to lens modelling, that can be applied in the era of large numbers of discovered strong lensing systems with upcoming surveys. We show that this combined methodology is greater than the sum of its parts, achieving higher accuracies than either method alone, and receiving a speed boost from the CNN-assisted prior initialisation.

In Chapter 3 we discuss the existence of unphysical source reconstructions, that represent local minima in χ^2, and often serve to derail the lens modelling process. These solutions represent a challenge to be overcome in the quest for a truly automated lens modelling pipeline. We train a CNN to identify when our lens modelling has converged upon a solution with an unphysical source reconstruction, and develop a simple prescription to re-initialise the parameter priors in order to ensure convergence upon the true solution. We show that using this approach, we can greatly improve upon the fraction of lens modelling runs that converge to the global maximum solution. This represents a further step towards creating a fully automated lens modelling pipeline.

In Chapter 4 we present our modelling of high-resolution ALMA observations of strongly lensed dusty star forming galaxies detected by Herschel. We employ uv plane lens modelling to infer the mass profiles of the lensing galaxies, and by determining the magnification factors, we investigate the intrinsic properties and morphologies of the lensed sub-millimeter sources. We find that these sources all have ratios of star formation rate to dust mass that are consistent with, or in excess of, the mean ratio for high-redshift sub-millimetre galaxies and low redshift ultra-luminous infrared galaxies.

In Chapter 5 we investigate the differences between performing lens modelling in the uv plane versus the image plane. We simulate a set of ALMA observations, for

varying antenna configurations and on-source integration times, and for a variety of lensed image configurations. From each simulated measurement set, we produce two image plane measurements via the CLEAN algorithm, using natural and Briggs weighting. We then perform direct visibility lens modelling, and image plane modelling for each system, and compare our results to look for biases introduced by one method over another.

Finally, in Chapter 6 we summarise the results from this thesis and discuss the prospects for extending the work presented here in the future.

Chapter 2

Strong lens modelling: comparing and combining BNNs and parametric profile fitting

To date, a relatively small number of strong gravitational lensing systems have been discovered, on the order of hundreds, with most of these being low redshift observations. A variety of surveys have been undertaken with the goal of discovering lensing systems, such as the Sloan Lens ACS (SLACS) survey (Bolton et al., 2006), the CFHTLS Strong Lensing Legacy Survey (SL2S; Cabanac et al., 2007), the Sloan WFC Edge-on Late-type Lens Survey (SWELLS; Treu et al., 2011), and the BOSS Emission-Line Lens Survey (BELLS; Brownstein et al., 2011). This is all set to change with the advent of two upcoming surveys: Euclid (Laureijs et al., 2011a), the European Space Agency's telescope scheduled to launch in 2023 will cover 15,000 $\deg^2$ over 6 years and study the accelerated expansion of the universe out to a redshift of $z = 2$; Additionally, the Vera Rubin Legacy Survey of Space and Time (LSST; Ivezic et al., 2008), also focused on the study of dark energy and dark matter, will

commence science operations in 2024. LSST will cover 18,000 deg^2 over ten years in six different filters (u, g, r, i, z, y). It is expected that these surveys will discover many tens of thousands of lensing systems; 120,000 and 170,000 lenses for LSST and Euclid respectively (Collett, 2015a). For this reason, the development of fast, automated, and accurate pipelines for finding and modelling strong lenses is of great importance.

CNNs have been employed to great effect in rapidly identifying strong gravitational lenses (Lanusse et al., 2017a; Petrillo et al., 2017; Davies et al., 2019; Jacobs et al., 2019; Metcalf et al., 2019; He et al., 2020; Huang et al., 2021), and have been proved to be able to handle the vast amounts of data we are expecting due to their ability to rapidly classify images. The performance of a CNN hinges upon the quality of its training data, and since we have so few known strong lensing systems, we must turn to simulated images to provide the necessary volume of data. Citizen science and spectroscopic methods have also been leveraged in the search for strong lenses (More et al., 2015; Sonnenfeld et al., 2020; Holwerda et al., 2015). However, it has been shown that these methods tend to detect distinct populations of lensing systems, highlighting a need for multiple approaches to build up a complete sample (Knabel et al., 2020).

Once a lens system has been identified, parametric mass modelling techniques are typically used to study the lens galaxy and reconstruct the background source distribution. A variety of techniques are available for doing this, such as those that reconstruct the source on a pixelised grid (Warren & Dye, 2003; Vegetti & Koopmans, 2009; Nightingale et al., 2018), or those that make use of the shapelet decomposition (Birrer et al., 2015; Birrer & Amara, 2018). Such modelling is a time-consuming process, typically involving a great deal of trial and error to hone in on an approximate set of parameters around which we can perform a more exhaustive search. Additionally, fitting models to complex datasets is often done in an iterative procedure,

with relatively simple fits performed at the start of the process, and more complex models of mass and source light distribution being fitted as the model is refined. To compound the problem, preprocessing of the data is often required for point spread function estimation, lens light removal and image masking. Techniques that simultaneously fit the lens galaxy light and mass profile have been developed, but at the expense of adding more non-linear parameters to the problem, and thus increasing the computational cost of lens modelling further.

In addition to lens identification, CNNs have been used as an alternative means of lens modelling. This was demonstrated for the first time by Hezaveh et al. (2017a), who used a combination of neural networks to predict the mass model parameters of simulated and real lensed images. Approximate Bayesian Neural Networks (BNNs) have been used to predict the uncertainties on the mass model parameters (Levasseur et al., 2017), further extending the utility of machine learning methods in this area. Whilst the training of these networks requires multiple days of GPU time, once this up front cost has been paid, the lens modelling speed is increased by several orders of magnitude over traditional parametric methods. Furthermore, these techniques have been extended to the application of uv-plane interferometer modelling and the reconstruction of the background source (Morningstar et al., 2018; Morningstar et al., 2019).

Whilst a trained neural network is capable of predicting the mass model parameters of a lensing system in a fraction of the time of traditional methods, it is constrained by the quality of the training data; A comprehensive set of data is required to deal with the nuances of real observations, and shortcomings in this area can lead to CNNs producing unreliable predictions. This is a key area in which the conventional parameter fitting techniques retain their utility and justify the significantly longer modelling times.

Figure 2.1: Examples of the simulated images used to train the neural network. Top and middle rows contain Sérsic and Gaussian profile sources, respectively. The bottom row contains complex sources made up of multiple Sérsic profiles. The data set was simulated to have the expected imaging characteristics of the *Euclid* telescope's VIS filter. Figure and caption reproduced from Pearson et al. (2021)

In this chapter, we investigate the following questions: How well can CNNs recover the mass model parameters of increasingly realistic simulated images? How well do traditional parameter-fitting techniques perform in comparison? Can the combination of the two techniques provide a more powerful lens modelling tool than either on their own?

For this work, we focus on the imaging characteristics of the *Euclid* telescope, and compare the CNN to the semi-linear inversion technique of `PyAutoLens` (Nightingale et al., 2018) as well as a combination of the two. Each of these modelling methods is tested on a suite of increasingly realistic lensing setups, beginning with SIE lenses and parametric sources, then incorporating real Hubble Ultra-Deep Field (HUDF; Beckwith et al., 2006a) source galaxies and CosmoDC2 (Korytov et al., 2019) line-of-sight (LOS) structure. We then move to using elliptical power law mass profiles and galaxies from the Evolution and Assembly of Galaxies and their Environments (EAGLE) cosmological hydrodynamical simulation suite (Crain et al., 2015; Schaye et al., 2015).

This chapter is organised as follows: Section 2.1 details the simulations used to produce the image data sets used for training and testing. Section 2.2 provides an overview of the parameter-fitting methods used in this chapter: `PyAutoLens` and the convolutional neural network. Section 2.3 presents and compares the results from testing on image data sets of increasing complexity, from simple parametric lenses and sources to hydrodynamical EAGLE galaxy lenses with real HUDF sources and line-of-sight structure. The results are discussed in Section 2.4 along with a conclusion of this work in Section 2.5.

2.1 Gravitational Lens Simulation

Here, we detail the simulated gravitational lens images used throughout this work. The CNN was trained on data comprised of simulated parametric lenses and parametric sources, whilst the datasets used to test the network had varying levels of complexity, from purely parametric, to HUDF sources and lens galaxies extracted from the EAGLE simulation suite.

2.1.1 Training set

Pearson et al. (2019) describes the simulation method used to generate the 100,000 images needed to train our CNN. The majority of strong gravitational lenses are massive elliptical galaxies, and so we opted to use the SIE model (Kormann et al., 1994; Keeton, 2001) and its generalised form, the elliptical power law model (Tessore & Metcalf, 2015), to represent the mass profiles in our training simulation datasets. The SIE model has three key parameters: Einstein radius, orientation and axis ratio. The elliptical power law model has an additional parameter; the power law slope. Consequently, we produced two trained CNNs, one trained on a set of SIE lensed images, and another trained on the set of power law lensed images. Aside from this difference in training, the two CNNs were identical in architecture and the training procedures used. The distributions of the lens model parameters were chosen to be uniform to ensure sufficient training across the parameter space. A range of $0.2 - 3.2$ arcseconds, $0 - 180°$, $0.4 - 1.0$, and $1.5 - 2.5$ for Einstein radii, orientations, axis ratios and slopes, were chosen, respectively. The range of Einstein radii was chosen to reflect that which we expect to obtain from the *Euclid* survey (Collett, 2015a). We assumed that the lenses were correctly centred, with the lens centroid following a Gaussian distribution centred on the image centre, and a standard deviation of one

pixel.

We adopted the expected characteristics of the *Euclid* telescope's VIS filter, for all the simulated images in this work. Resultantly, we have used an image pixel scale of 0.1 arcsec pixel^{-1} (Racca et al., 2016) and a point spread function with full-width half-maximum of 0.17 arcseconds. Supplementing the parametric Sérsic profiles used in Pearson et al. (2019), our training data also contained Gaussian sources and complex sources constructed from multiple Sérsic sources to represent a wide range of source morphologies. The fraction of the training data taken up by the single Sérsic profiles, Gaussian profiles and multiple Sérsic profiles were 25, 25, and 50 per cent, respectively. The position of the sources were chosen at random from a uniform distribution within the Einstein radius of the lens, to ensure that a wide range of lens image configurations and magnifications were represented in the training data. Example training images are shown in Fig. 2.1.

The lens and source redshifts were sampled from uniform distributions with upper limits of $z_{\rm lens} = 2$ and $z_{\rm source} = 6$. A number of criteria were enforced to ensure that strong gravitational lenses were produced: the centre of the source must be multiply-imaged, with a resolved counter image, and a minimum magnification of $\mu_{\rm TOT} > 3$, tangential shearing ($\mu_{\rm TOT} R_{\rm e,source} > seeing$) and signal-to-noise ratio (SNR > 20).

The sky background for *Euclid* VIS (Collett, 2015a; Niemi, 2015) was added along with Poisson noise, the expected read noise (five electrons per readout) and dark current (two electrons per pixel per second; Radeka et al., 2009). The simulated images initially contained light from the lens galaxy, and the true profile was subtracted after convolution with the PSF, leaving only Poisson noise residuals.

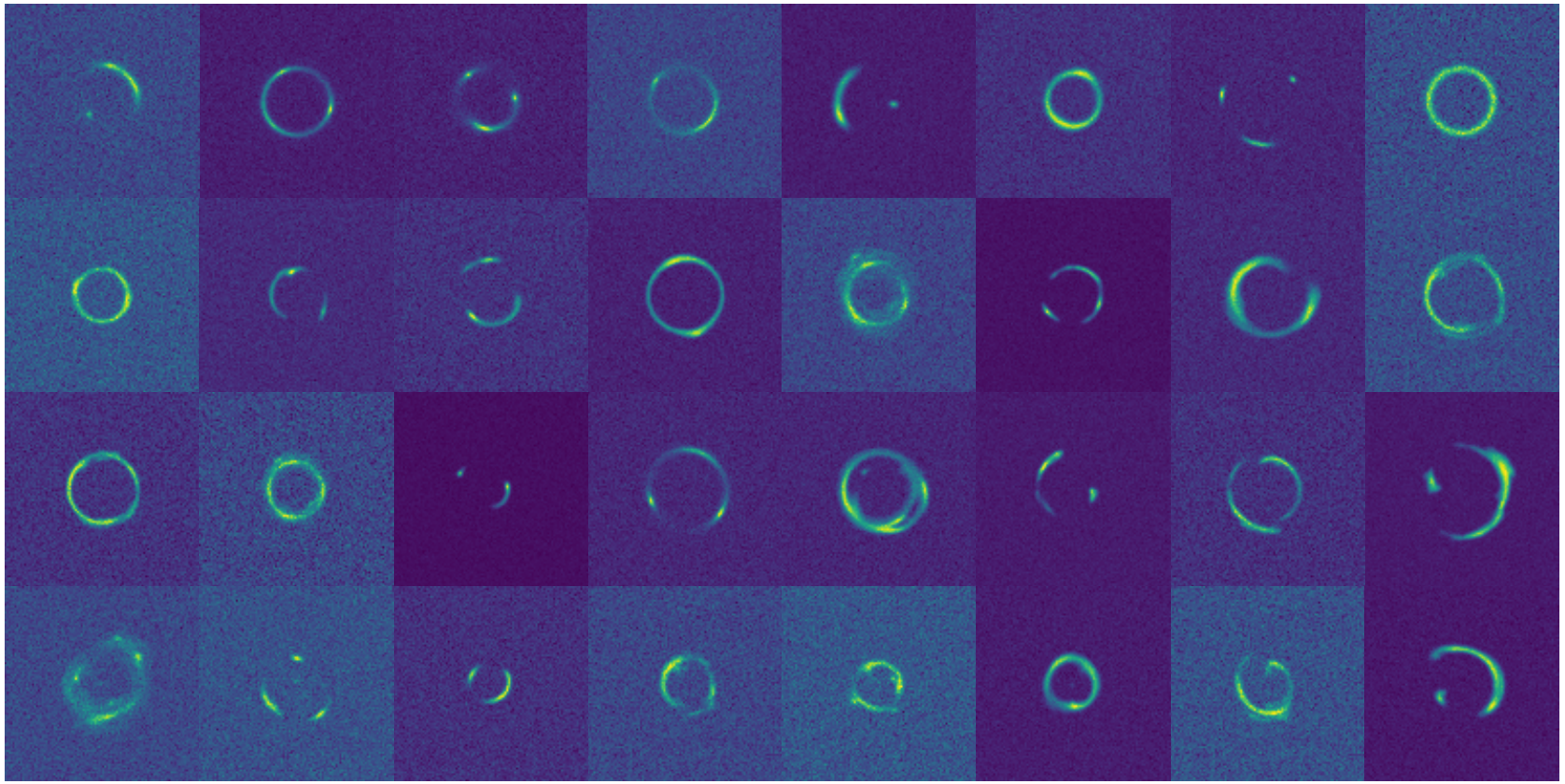


Figure 2.2: Examples of the simulated images produced from PICS software, used to test the different mass modelling methods. Top row: examples containing SIE lenses with parametric source profiles, with four Sérsic sources followed by four Gaussian sources. Second row: examples containing SIE lenses with HUDF sources, with the last four containing LOS structure. Third row: examples containing power law lenses with HUDF sources, with the last four containing LOS structure. Bottom row: examples containing EAGLE galaxy lenses with HUDF sources, with the last four containing LOS structure. The data sets were simulated to have the expected imaging characteristics of the *Euclid* telescope's VIS filter. Figure and caption reproduced from Pearson et al. (2021).

2.1.2 Test sets

In order to test the CNN and `PyAutoLens` we created a variety of test datasets, of 1000 images each, using the Pipeline for Images of Cosmological Strong Lensing (PICS; Li et al., 2016a). As a source for complex, high redshift and high resolution background galaxies, we made use of their selection of Hubble Ultra-Deep Field (HUDF; Beckwith et al., 2006a) galaxies. We did not attempt to deconvolve these sources in this work. The magnitudes and angular extents of these sources, along with the Sérsic and Gaussian profiles, were incorporated into the images. Changes in the lens and source redshifts therefore only had the effect of modifying the Einstein radii, and so we fixed the lens and source redshifts to be $z_{lens} = 0.5$ and $z_{source} = 2.0$, respectively, with Einstein radii set manually.

In addition to the main lensing galaxy, PICS facilitates the inclusion of line-of-sight (LOS) structure through the ray-tracing of small light cones covering lens caustics, extracted from the semi-analytic CosmoDC2 (Korytov et al., 2019) model, based on the Outer Rim cosmological N-body simulation (Heitmann et al., 2019). We selected one thousand of these light cones, with each containing a primary lens and up to one hundred LOS galaxies between the observer and background source. A range of test data sets were created for a variety of increasingly complex and realistic lensing systems, examples of which can be seen in Figure 2.2:

- SIE lenses + simple parametric source profiles
- SIE lenses + HUDF sources, with and without LOS structure
- Power law lenses + HUDF sources, with and without LOS structure
- EAGLE galaxy lenses + HUDF sources, with and without LOS structure

In the case of the SIE lens mass profiles, the parameter distributions followed those

determined by PICS to be the main halos in the CosmoDC2 light cones. The Einstein radii, orientations and axis ratios of the lenses were distributed in the ranges 1.3-2.6 arcseconds, 0-180° and 0.4-1.0, respectively. For the power law lenses, the same distribution of lens parameters was used, with the addition of the slope parameter being uniformly sampled from a range of 1.5-2.5, matching the CNN training data. On average, the main halo lenses selected in this manner, were more massive and thus produced larger Einstein radii than those expected to be found by the *Euclid* survey. For the purposes of this work, we did not change these values, as we wished to maintain the halo masses and positions of the LOS structure for realistic ray-tracing.

EAGLE galaxy lenses, selected from the EAGLE galaxy catalogue, were also used. The galaxies were selected by computing their Einstein radii from their halo masses, assuming $z_{lens} = 0.5$ and $z_{source} = 2.0$, and ensuring that they were within the range of 0.5-5.0 arcseconds. This resulted in a sample of lenses with Einstein radii distributed in the range of 0.5-3.2 arcseconds. This distribution of parameters differs from those selected from the CosmoDC2 halos, but we limit the axis ratios and orientations to the same range. These values were obtained by performing a least squares fit to the convergence maps of the EAGLE particle data, weighted by the square root of the convergence map, using the SIE and power law mass profiles as the fitting function. This process was repeated for the inclusion of LOS structure.

The inclusion of LOS structure serves to increase the observed Einstein radius of a lens when compared to the Einstein radius of the lens itself. As such, we measured the observed Einstein radius for each system, using the previously mentioned convergence map fitting process. This measured value was then taken as the true value of the Einstein radius of the system, to mitigate biases in the results due to the CNN and `PyAutoLens` examination of the observed images.

2.2 Lens modelling methodology

In this section, we detail the techniques used to model our simulated gravitational lenses. We describe the processes used within the context of the parametric fitting approach of `PyAutoLens`, the convolutional neural network and the means by which we combine these two techniques.

2.2.1 `PyAutoLens`

To represent the conventional parameter fitting approach to modelling strong gravitational lenses, we used the semi-linear inversion implementation of `PyAutoLens`[1] (Nightingale et al. (2018); see Section 1.7 for a detailed description). This allows for the simultaneous modelling of the foreground deflector and reconstruction of the background source in a Bayesian framework. `PyAutoLens` offers a wide range of configuration options and can be tailored to suit many lensing scenarios. For example, the background source can be reconstructed on a rectangular grid, or a variety of adaptive grids that place smaller pixels in regions of high magnification or where the source is brightest. During the modelling process, a χ^2 statistic for the observation and model is produced and used to calculate the Bayesian evidence. The parameter space of the mass model and source plane is explored, searching for the set of parameters that corresponds to the global maximum of the evidence (and correspondingly, the global minimum in χ^2). Due to the large volume of simulated lenses that we had to model, we opted for a relatively simple setup of `PyAutoLens`, in order to make the computational cost of the process more reasonable. We chose to reconstruct the background source on an adaptive, magnification based, Voronoi grid, making use of a gradient regularisation scheme, analogous to that which is described in Warren & Dye

[1] https://github.com/Jammy2211/PyAutoLens

(2003), with a constant weight for the source plane. The lens modelling process was initialised with a mass model prior, a regularisation prior and a prior on the number of pixels in the source plane (see below). Each image-plane pixel was sub-gridded into four sub-pixels for the purpose of ray-tracing calculations, and the images were masked according to the process described in Maresca et al. (2020). `PyAutoLens` offers a number of sampling algorithms for the optimisation of model parameters, but for this work we opted to use the Bayesian inference tool, MultiNest (Feroz et al., 2009), run in constant efficiency mode and a sampling efficiency of 0.4.

Setting up `PyAutoLens` requires choosing a set of priors on the free parameters of the model. The priors on the mass model are typically set through a process of human inspection, and trial and error. This process would be too time-consuming for this work, and so throughout this work, we set up an automated and standardised approach to choosing the mass-model priors. `PyAutoLens` was run either with blindly initialised priors, or with values informed by the predictions of the CNN. To facilitate the blind modelling approach with `PyAutoLens`, we approximated the Einstein radius by fitting a circle to the image pixels above a 3-sigma threshold, using uniform weights. We then constructed a prior on the Einstein radius, centred on this approximate value, with a fixed with of ± 0.2 arcseconds. The priors on the other lens model parameters were left uniform over the full range of values. This automated approach was necessary to avoid significant contamination to our results from unphysical source reconstructions (Maresca et al., 2020), and is representative of what one might do when faced with finding an unknown lens model, so remained a fair comparison to the other automated methods we consider here.

2.2.2 Convolutional Neural Networks

The following section describes the CNN used throughout this work. A description of the network architecture is given in Section 2.2.2.1, and a brief overview of how uncertainties can be obtained on the CNN's predictions is given in Section 2.2.2.2. Some details of how the network was trained are provided in Section 2.2.2.3.

2.2.2.1 Architecture

The network used in this work is composed of six convolutional layers, six dropout layers, four pooling layers and two fully connected layers, as shown in Figure 2.3. In the convolution layers, the inputs are convolved with a series of kernels, as described in Section 1.8.1. The weights and biases of these kernels are optimised during the training process. Some hyperparameters of the network, such as the number of convolutional layers, and the number and size of the kernels used, were manually tuned through a process of experimentation to provide the best efficiency for this work.

The final layer of the network is the output layer, whose output is the CNN predictions of the lens model parameters, given a lensed image. For an SIE lens, the predicted parameters are the Einstein radius and the two ellipticity components, given by

$$e_1 = \frac{1-q^2}{1+q^2}\cos 2\phi\,,$$
$$e_2 = \frac{1-q^2}{1+q^2}\sin 2\phi\,,$$

where ϕ and q are the orientation and axis ratio of the mass profile, respectively. Where the power law profile has been used, the network also predicts the slope parameter, n. We refer the reader to Pearson et al. (2021) for further details of the CNN

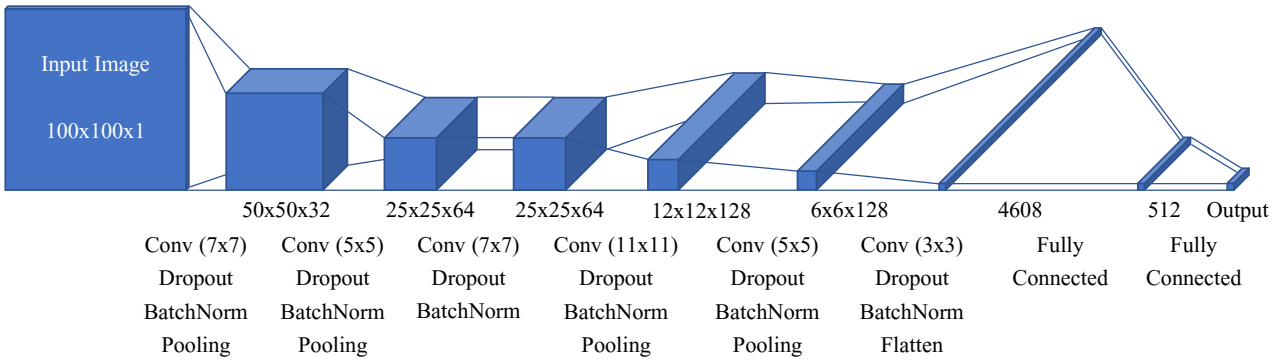


Figure 2.3: Layout of the CNN used in this work, detailing the six convolutional layers, with dropout and batch normalisation, four max-pooling layers and two fully connected layers. A flatten layer is used to connect the multidimensional data to the 1D fully connected layer. A ReLU activation function is used throughout. The kernel sizes (in pixels) are given underneath each block. Further details can be found at the end of Section 2.2.2.1. Figure reproduced from Pearson et al. (2021).

architecture.

2.2.2.2 Uncertainties

Convolutional neural networks applied to regression problems are trained to predict model parameters, but they do not offer a value of uncertainty on these values. In recent years, work by (Gal & Ghahramani, 2016a,b; Kendall & Gal, 2017) has incorporated Bayesian statistics into neural networks, allowing the network's weights to vary according to a probability distribution, and thus obtain a posterior probability and uncertainties on the predicted parameters. Work by Levasseur et al. (2017) applied these ideas to the problem of strong lens modelling, via the use of variational inference. A discussion of the theory underpinning this can be found in Pearson et al. (2021), with more detail being available within Levasseur et al. (2017).

2.2.2.3 Training

100,000 images were used to train the network, with a further 10,000 set aside for validation data. The images were 100×100 pixels, featuring a wide variety of source types in order to be able to cope with a wide array of realistic sources (see Section 2.1.1). The two CNNs (one for SIE predictions and the other for power law predictions), were trained in the same manner and featured identical architectures, except for the inclusion of an extra output for the slope parameter in the case of the power law model.

A GPU machine was employed, allowing for much faster processing than a typical CPU machine. The CNN was trained over 150 epochs and this was completed in under two hours.

2.2.3 `PyAutoLens` + CNN

Parametric fitting with `PyAutoLens` allows for accurate modelling of the lens profile and background source, with a great deal of flexibility in the chosen models. For example, it is trivial to add new components to the lens model, such as external shear, and there are no assumptions about the source morphology when using the semi-linear inversion method. However, this comes with a relatively high cost of CPU time and the need for the user to invest time in setting up the models. With our expectations of discovering hundreds of thousands of strong galaxy-galaxy lenses with upcoming surveys, we would like to retain the power and flexibility of parametric modelling, whilst reducing the time spent by the user in setting up and maintaining such large numbers of modelling runs. One can perform a kind of crude automation with `PyAutoLens` by setting the priors to cover parameter space without basing them on human inspection (i.e. "blind modelling"). In theory, this should allow for the exploration of parameter space and convergence upon the global best fit, however, in practice this leads to increased modelling times and an increase in the chances of converging upon a solution representing a local minimum in χ^2 (Maresca et al., 2020). The opposite problem also exists; setting the widths of the priors to be too narrow, when we don't know precisely where in parameter space we should search, won't allow for the sampling algorithm to converge upon the correct solution.

A trained CNN represents a significantly faster and more automated alternative to the approach of `PyAutoLens`, but the quality of the predictions from the CNN is dependent upon a sufficiently varied and realistic training set. Shortfalls in the training data, such as not including enough complexity in the background sources, could lead to the CNN not being able to reach the high degree of accuracy that `PyAutoLens` is capable of. However, we can leverage the strengths of both methods by combining the fast predictions of the CNN with the flexible and accurate modelling

of `PyAutoLens`.

Modelling a gravitational lens with `PyAutoLens` requires the user to set up prior distributions on the lens model parameters, which are typically either uniform or Gaussian distributions. The CNN can predict lens model parameters, along with uncertainties on those values, opening up the potential for using this as a means of initialising `PyAutoLens`. We propose two basic approaches to combining these methods: Firstly, we can use the CNN predictions to initialise the `PyAutoLens` lens model with uniform prior distributions of some predetermined width, but centred on the CNN predictions. Secondly, we can use the CNN predicted uncertainties to initialise the `PyAutoLens` lens model with Gaussian priors, centred on the predicted values and with the CNN predicted uncertainties defining the 1σ value of the distribution.

This combination of CNNs and parametric modelling has the potential to alleviate the shortcomings of each method. The CNN can vastly reduce the volume of parameter space necessary for `PyAutoLens` to search, via the automatic initialisation of sensible prior distributions, and `PyAutoLens` can provide a means of accurately modelling complex lensing systems that may not be properly accounted for in the training data of the CNN. In this manner, combining these two methods can lead to a fully-automated lens modelling pipeline, capable of providing accurate results on timescales that are acceptable when dealing with large numbers of newly discovered lensing systems.

2.3 Results

In this section, we compare the performance of blind modelling with `PyAutoLens` (PyAL (blind)) to that of the CNN for a range of test cases of varying complexity. Additionally, we compare with the two techniques we have used for combining the

CNN's predictions with the creation of priors in `PyAutoLens`: the first method (PyAL + CNN) uses uniform priors of a predetermined width, centred on the CNN predictions. The second method (PyAL + CNN (1σ)) uses Gaussian priors, centred on the CNN predictions with a width set by the CNN predicted uncertainty. We experimented with using twice the CNN predicted uncertainties as the 1-sigma widths of the Gaussian priors, but found this to be too generous, causing a negative impact on the results, and so have opted to not include those here. The prior distributions used by the PyAL (blind) method were uniform over the full range of parameter space, except for the Einstein radius, which used a uniform prior of width 0.4 arcseconds, centred on the approximate Einstein radius, as discussed in Section 2.2.1. For the PyAL + CNN method, we used uniform prior distributions, centred on the CNN predictions with bounds of ± 0.2 arcseconds, $\pm 40°$, ± 0.1, and ± 0.2 for Einstein radius, orientation, axis ratio and slope, respectively.

In Sections 2.3.1 and 2.3.2, the modelling techniques are applied to simulated images containing SIE lenses, with parametric (Sérsic and Gaussian) and real HUDF sources, respectively. In the case of HUDF sources, results are compared for lenses with and without line-of-sight structure, and this is repeated for the case of power law lenses in Section 2.3.3. In Section 2.3.4, the techniques are instead applied to simulated images with complex lenses in the form of galaxies obtained from the EAGLE hydrodynamical simulation, again with HUDF sources and for those with and without line-of-sight structure.

Throughout the remainder of this chapter, we quantify the error on predicted lens model parameters by the 68 per cent confidence interval computed from the distributions of differences between true and predicted parameter values across the test image set.

2.3.1 SIE Lenses + Parametric Sources

To begin with, we tested all of our methodologies on 1000 images with SIE lenses and parametric sources. Half of these were Sérsic sources, with the other half being Gaussian sources. We present the results in Figure 2.4 as distributions of the differences between predicted values and true values for each mass model parameter. Fractional differences are used for Einstein radius and axis ratio, whilst the absolute difference is used for the orientation. Table 2.1 gives the 68 per cent confidence intervals of these distributions. Scatter plots of predicted versus true values are given in Figure 2.4, including the CNN's predicted 1σ uncertainties. We can see from these results that the CNN's uncertainties are of a reasonable size and accurately capture its errors. The large uncertainties on some of the CNN's predictions of orientation are driven by rounder lenses having less well-defined orientations.

`PyAutoLens` by itself, working in the PyAL (blind) mode, achieves a higher accuracy than the CNN for both the Einstein radius and orientation parameters. PyAL (blind) suffers from significant numbers of outliers at $\pm 90°$ in the prediction of orientation (see Figure 2.4), which dramatically decrease its overall performance for this parameter. The CNN slightly outperforms PyAL (blind) in its predictions of the axis ratio. However, for all mass model parameters the combination of the two methods does significantly better than PyAL (blind), with PyAL + CNN (1σ) achieving 36-44 per cent lower errors than PyAL (blind). There is also a significant benefit to using the CNN's predicted 1σ uncertainties as the basis of a Gaussian prior for `PyAutoLens`, with 12-17 per cent lower errors for PyAL + CNN (1σ) compared to PyAL + CNN. There are some observed biases in the distributions of these predictions, but they are small enough to have negligible impact on modelling. The bias in Einstein radius is approximately 0.2 per cent, whilst the bias in axis ratio is < 1 per cent.

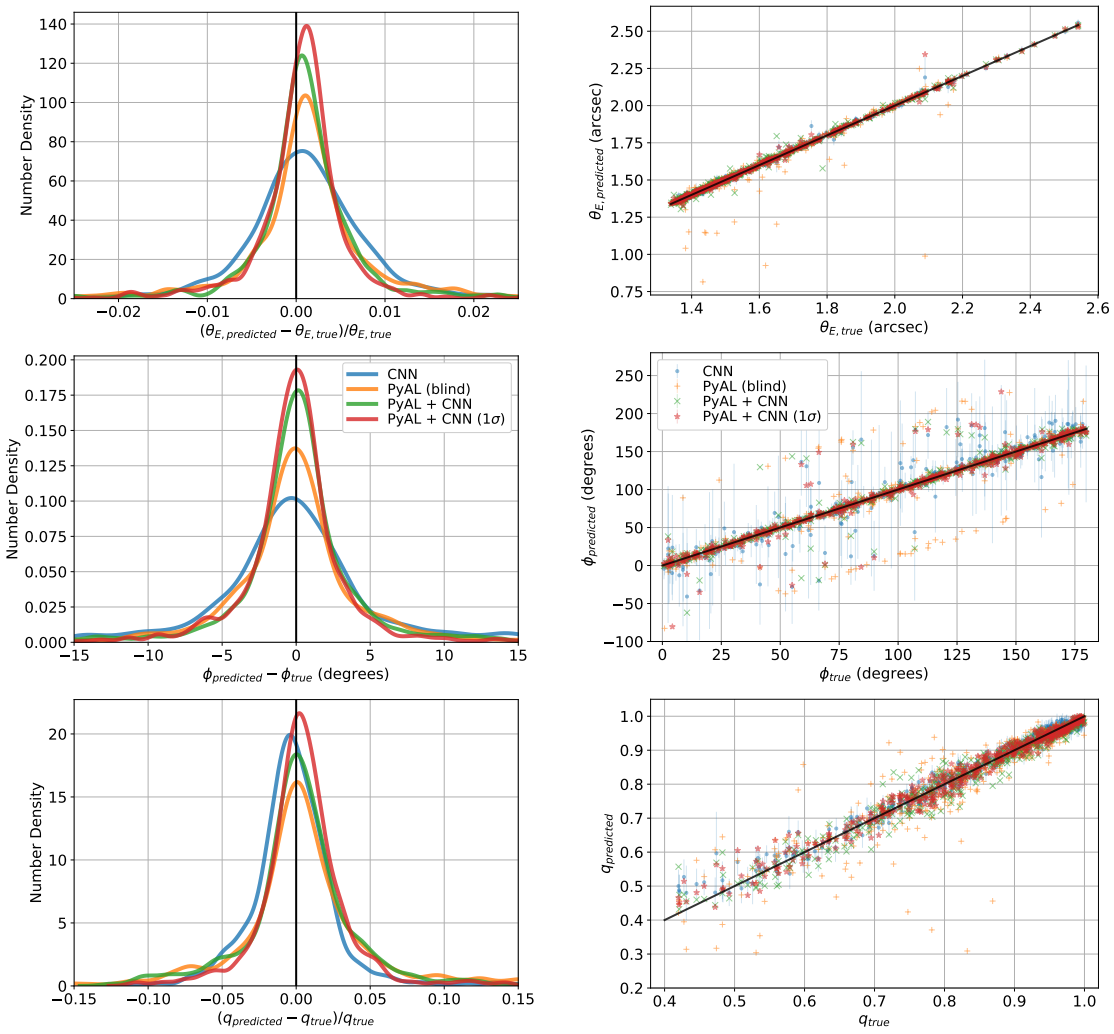


Figure 2.4: Left column: error distributions for the SIE mass model parameters for a test data set of 1000 images with parametric sources (500 Sérsic and 500 Gaussian). Einstein radius and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

Table 2.1: **SIE lenses + parametric sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images.

Method	θ_E (arcsec)	ϕ (°)	q
CNN	0.0090	4.92	0.018
PyAL (blind)	0.0089	4.44	0.027
PyAL + CNN	0.0065	2.81	0.021
PyAL + CNN (1σ)	0.0055	2.47	0.017

2.3.2 SIE Lenses + HUDF Sources

Our next series of tests involved replacing the simple parametric sources with real HUDF ones. We also investigated the impact of adding line-of-sight structure to the lensing simulations. The results from modelling images without LOS structure are shown in Figures 2.5. The results containing LOS structure are shown in Figures 2.6. The 68 per cent confidence intervals are given in Table 2.2.

Examining the results for images without line-of-sight structure, we see that in general, the accuracies for all methodologies have dropped in comparison to the images with parametric sources. As before, the CNN accuracy is lower for the Einstein radius and orientation, whilst performing better than PyAL (blind) when determining the axis ratio. Once again, we see that the two methods incorporating the CNN's predictions into `PyAutoLens` outperform either method in isolation. PyAL + CNN (1σ) errors are 31-39 per cent lower than PyAL (blind). Likewise, in Figure 2.5, we continue to see significant numbers of underpredicted axis ratios in the results of PyAL (blind); these results correlate with under-predicted Einstein radii, and are not dependent upon lensing configuration, but rather are likely due to MultiNest converging upon an unphysical source reconstruction, as discussed in Maresca et al. (2020); as such, the methods that restrict parameter space using priors informed by the CNN's predictions, do not suffer from this issue.

The inclusion of line of sight structure into the images has the effect of generally decreasing the accuracy for all results, with errors increasing by factors of 2.2, 8.2 and 2.3 on average for Einstein radius, orientation and axis ratio, respectively. This behaviour is expected, since the inclusion of extra mass along the LOS increases the complexity of the lensing system and this will not be properly accounted for by a single smooth parametric profile. As described in Section 2.1.2, we defined the "true"

parameters for these systems by fitting an SIE model to the convergence maps of the SIE profile and LOS structure. Of particular note is the fact that the CNN was not trained on any images containing HUDF sources or LOS structure, and yet it still performs comparatively well as an automated modelling method on its own or in conjunction with `PyAutoLens`.

Here, the CNN outperforms the other methods in its predictions for Einstein radius, achieving errors 10 per cent lower than PyAL + CNN (1σ), but this combined method continues to provide the best results for axis ratio and orientation. The overall difference in performance between each of the methods is less pronounced when compared to the results for images without LOS structure. Compared to PyAL (blind) the errors are reduced by 9-28 per cent, 11-17 per cent, and 16-21 per cent for the CNN, PyAL + CNN, and PyAL + CNN (1σ), respectively. The CNN's predicted uncertainties for images with LOS structure are not always reflective of its actual error, due to the lack of images with LOS structure in the CNN training set.

2.3.3 Power Law Lenses + HUDF Sources

We present the results of our lens modelling tests, but this time with power law lenses. Here, a nearly identical version of the CNN was used, with one extra output for predicting the power law slope, n, along with its 1σ uncertainty, and the training and testing data was composed of power law mass profiles. The 68 per cent confidence intervals are given in Table 2.3. Figure 2.7 presents results for images without LOS structure, while figure 2.8 presents those containing LOS structure.

Figures 2.7 and 2.8 highlight the difficulty of predicting the slope parameter of the power law model. PyAL + CNN (1σ) achieved the highest accuracy, but with significant scatter in its predictions. This extra degree of freedom, and the degeneracy

Table 2.2: **SIE lenses + HUDF sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images.

Method	θ_E (arcsec)	ϕ (°)	q
Without LOS structure			
CNN	0.013	6.10	0.029
PyAL (blind)	0.013	4.22	0.038
PyAL + CNN	0.010	3.02	0.030
PyAL + CNN (1σ)	0.009	2.57	0.026
With LOS structure			
CNN	0.021	30.7	0.067
PyAL (blind)	0.029	33.8	0.077
PyAL + CNN	0.026	28.0	0.068
PyAL + CNN (1σ)	0.023	27.1	0.065

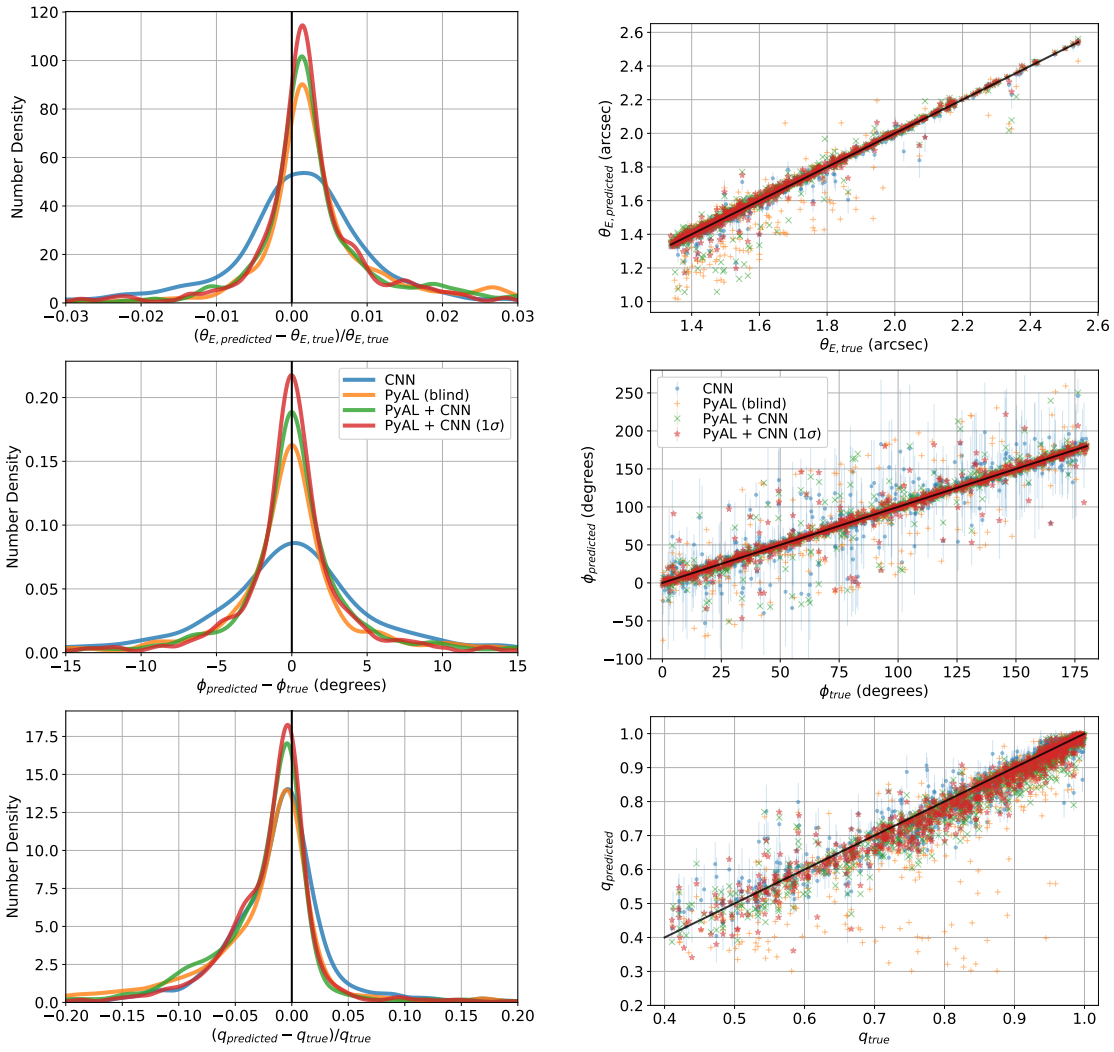


Figure 2.5: Left column: error distributions for the SIE mass model parameters for a test data set of 1000 images containing HUDF sources and no line-of-sight structure. Einstein radius and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

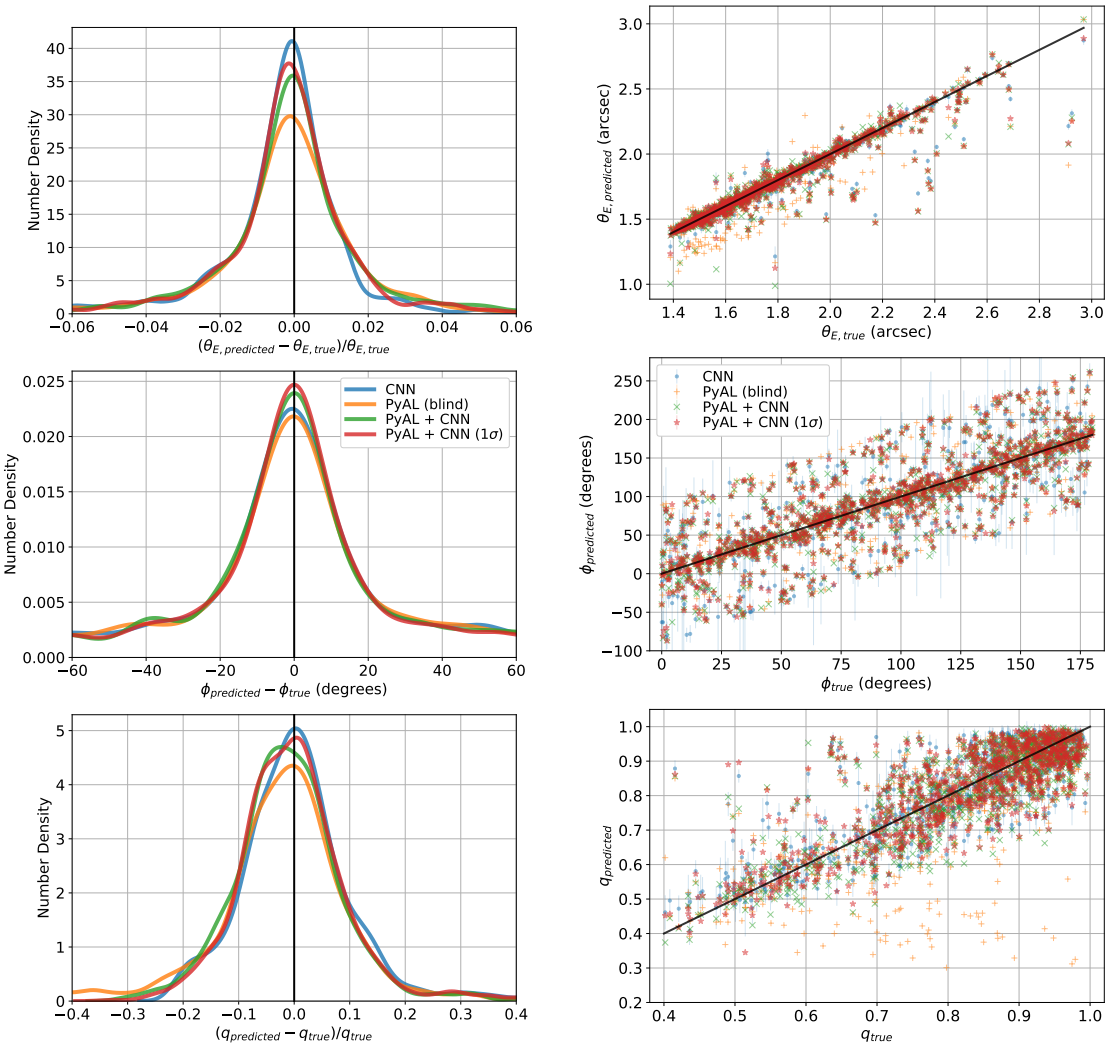


Figure 2.6: Left column: error distributions for the SIE mass model parameters for a test data set of 1000 images containing HUDF sources and line-of-sight structure. Einstein radius and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

between n, q, and θ_E, leads to generally worse performance at predicting the other lens model parameters compared to the SIE lenses; for increasing true slope values, predicted slopes in general go from over- to under-predicted, while the opposite occurs for axis ratio predictions. The errors increased by factors of 1.2-2.9 and 1.1-4.2 for test sets with and without LOS structure, respectively. While the scatter in the power law results is larger than that of SIE lenses, the CNN's predicted uncertainties remain a generally accurate reflection of its errors for both lenses with and without LOS structure.

Regarding the images without LOS structure, the errors of PyAL + CNN (1σ) are 5-39 per cent lower than those of PyAL + CNN, and 32-62 per cent lower than those of PyAL (blind). Meanwhile, the CNN alone achieves errors comparable to PyAL (blind), from 19 per cent lower to 13 per cent larger. We see from Figure 2.7 that a significant number of the predicted orientations from PyAL (blind) are offset by $\pm 90°$, whereas the frequency of this occurring is greatly reduced when incorporating the CNN predictions into `PyAutoLens`'s prior setup. These results once again highlight the power of combining machine learning methods with parametric modelling.

With LOS structure, the CNN and PyAL + CNN (1σ) methods reduce errors by 23-35 per cent and 27-36 per cent compared to PyAL (blind) respectively. As before, with the SIE lenses and LOS structure, the difference between the CNN and PyAL + CNN (1σ) methods is less pronounced than for the images not containing LOS structure; the PyAL + CNN (1σ) method obtains only 0.3 - 14 per cent lower errors than the CNN, compared with the 28 - 66 per cent improvement obtained for images without LOS structure. Similarly, the incorporation of 1σ uncertainties into the priors of `PyAutoLens` only serves to reduce the errors by 5-12 per cent over the PyAL + CNN method. The errors on the lens model parameters, when compared to the images without LOS structures, increase by factors of 1.6, 9.2, and 1.3 for Einstein radius,

orientation and axis ration, respectively, whilst the errors on the slope parameter actually decreased by 18 per cent. Generally speaking, the modelling methods tend to over-estimate the slope values (i.e. more centrally concentrated lenses), so the inclusion of LOS structure away from the central lens may be acting to reduce the inferred value and thus decrease the overall error in these results.

2.3.4 EAGLE Lenses + HUDF Sources

Our final set of test images were created with complex lenses, that do not follow an analytic profile, taken from the EAGLE hydrodynamical simulation. Once again, we created two sets of 1000 test images, which were identical apart from the inclusion of LOS structure, as was previously done for the parametric lenses. The "true" parameter values for these lenses were set by fitting the known convergence maps with an SIE mass profile. Figures 2.9 and 2.10 display the accuracies for each modelling method, with corresponding scatter plots in Figs 2.9 and 2.10, in the same format as previous results.

It is worth noting that the EAGLE galaxy lenses modelled here have, on average, smaller Einstein radii than those lenses discussed in previous sections (see Section 2.1.2), resulting in more pixelated lensed arcs. It is therefore expected that all of our modelling methods will produce higher errors on the parameters, in addition to the increased complexity of modelling the non-parametric mass distributions. However, when analysing the results of lenses with $\theta_E > 1.3$, as is the case for our other test sets, we see much larger errors than those seen previously, suggesting that the dominant source of error is the complexity of the lens distribution rather than the smaller Einstein radii.

With no LOS structure, the CNN outperforms PyAL (blind), achieving 36-55 per

Table 2.3: **Power law lenses + HUDF sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images. The final column contains the confidence intervals for the power law slope parameter.

Method	θ_E (arcsec)	ϕ (°)	q	n
Without LOS structure				
CNN	0.050	7.37	0.083	0.27
PyAL (blind)	0.065	6.53	0.106	0.29
PyAL + CNN	0.049	2.63	0.080	0.25
PyAL + CNN (1σ)	0.029	2.50	0.062	0.20
With LOS structure				
CNN	0.070	32.2	0.093	0.19
PyAL (blind)	0.091	44.1	0.143	0.26
PyAL + CNN	0.068	34.1	0.096	0.19
PyAL + CNN (1σ)	0.060	32.1	0.091	0.18

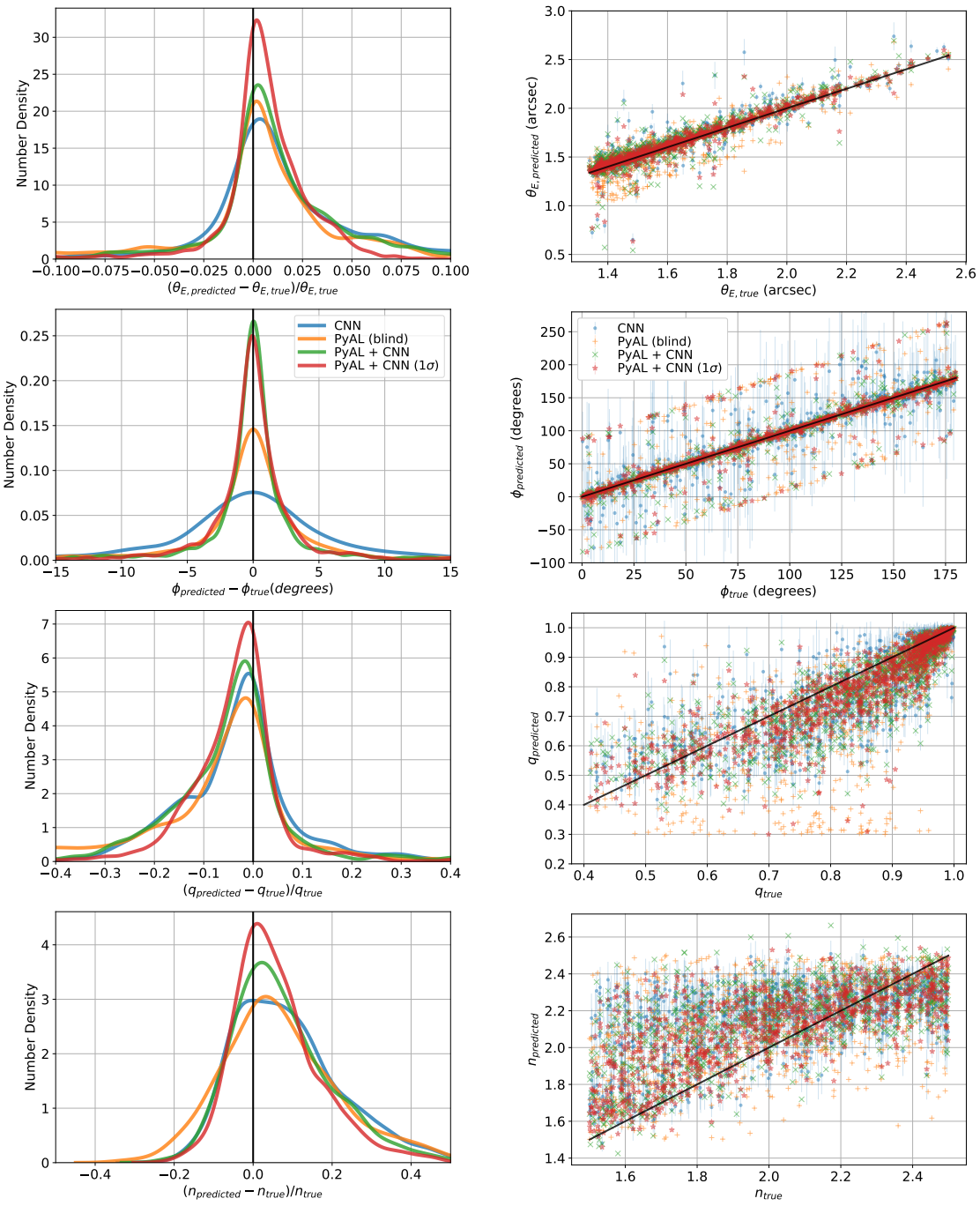


Figure 2.7: Left column: error distributions for the power law mass model parameters for a test data set of 1000 images containing HUDF sources and no line-of-sight structure. Einstein radius, slope, and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

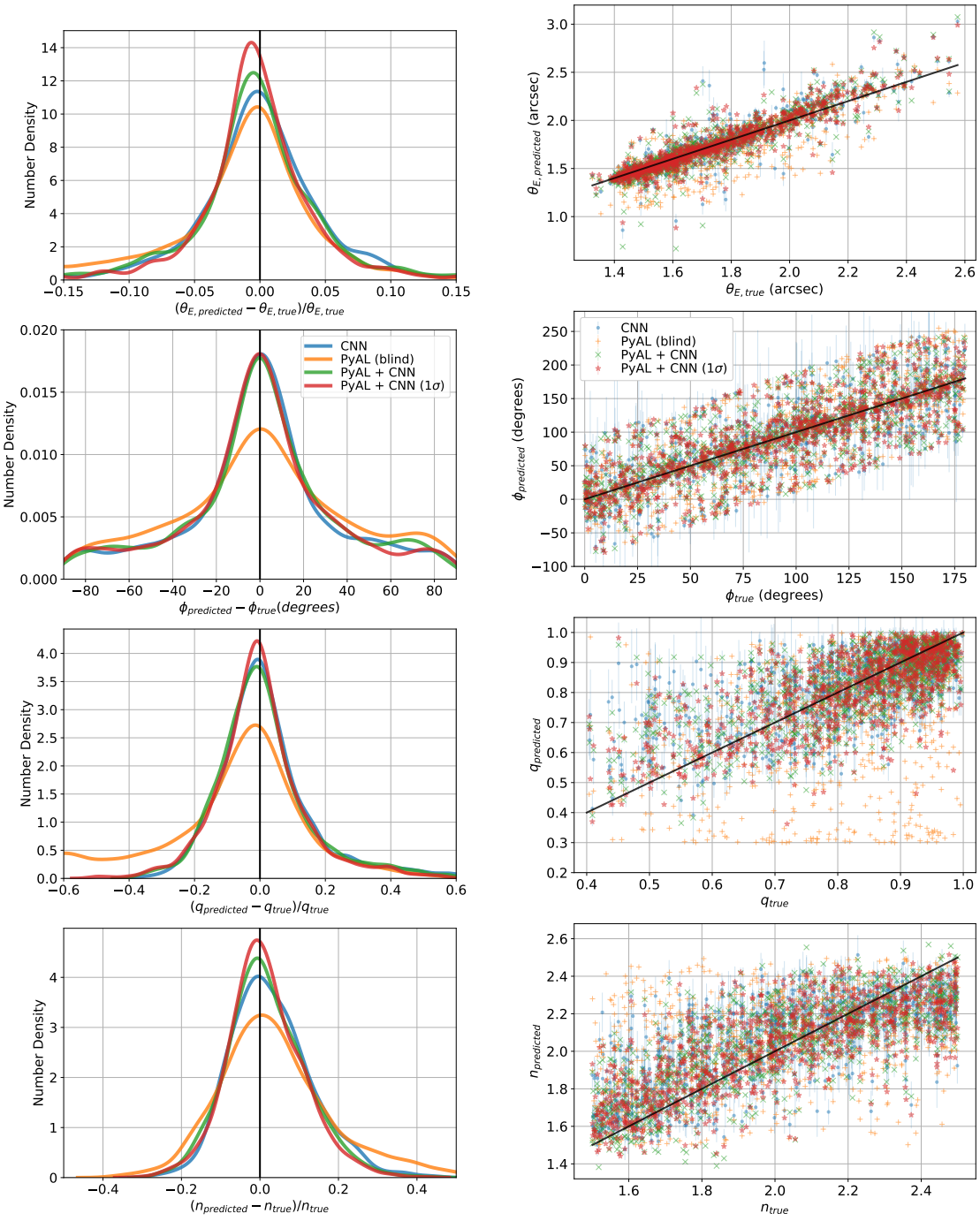


Figure 2.8: Left column: error distributions for the power law mass model parameters for a test data set of 1000 images containing HUDF sources and line-of-sight structure. Einstein radius, slope, and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

cent lower errors than the latter method. Here, the performance of PyAL + CNN (1σ) achieves similar results to the CNN, giving 30-58 per cent lower errors than PyAL (blind). However, this method does not outperform all the CNN based methods for all parameters: For Einstein radius and orientation, PyAL + CNN (1σ) achieves 11-22 per cent and 2-7 per cent lower errors compared to PyAL + CNN and the CNN, respectively, but for axis ratio the PyAL + CNN and CNN methods instead achieve 2.7 and 9.6 per cent lower errors than PyAL + CNN (1σ). This behaviour may be explained in part by the under-prediction of axis ratio (see Figure 2.9) by our modelling methods. This is likely due to a combination of the fact that these complex foreground galaxies can have axis ratios that are a function of radius, and that fewer smooth arcs are produced than the data on which the CNN was trained. It is also clear from both Figures 2.9 and 2.10, that the CNN has underpredicted its uncertainties, particularly in the case of the Einstein radius. This will have a knock-on effect to the achieved accuracy of the PyAL + CNN (1σ) method.

For our final test set of images, containing LOS structure, PyAL + CNN (1σ) achieves the lowest errors for all parameters, but is broadly comparable to the performance of the CNN's results, with only an improvement of 0.4-7 per cent over the latter method. Consequently, their improvements over PyAL (blind) are near identical, at 28-48 per cent and 28-47 per cent, for PyAL + CNN (1σ) and the CNN respectively. We continue to see an improvement in our results by incorporating the CNN's predicted uncertainties into our `PyAutoLens` prior models, which reduce errors by 6-28 per cent compared with PyAL + CNN. The inclusion of LOS structure in these images has a much smaller impact than the complex nature of the EAGLE galaxy lenses themselves, with errors a factor of 1.3, 1.5 and 1.0 larger than for images without LOS structure for Einstein radius, orientation and axis ratio, respectively. Including LOS structure, when compared to SIE lenses with LOS structure, the errors are factors of

2.2, 0.5 and 1.8 larger for Einstein radius, orientation and axis ratio; this is contrasted with factors of 3.9, 3.1, and 4.2, for Einstein radius, orientation, and axis ratio when there is no LOS structure.

2.3.5 Modelling Speed

One of the key benefits to incorporating the CNN into our parametric modelling approach is the reduction in CPU time required to model a strong gravitational lens system. The distributions of the time taken by each modelling approach are shown in Figure 2.11, for the SIE lenses with HUDF sources, SIE lenses with HUDF sources and LOS structure, and EAGLE galaxy lenses with HUDF sources and LOS structure. We do not include distributions for the CNN modelling time, as these are almost instantaneous.

Figure 2.11 clearly demonstrates the computational benefits of providing the CNN predictions to `PyAutoLens` for all three test data sets. PyAL + CNN achieves modelling speeds that are factors of 1.34, 1.08 and 1.14 times faster than PyAL (blind), whilst PyAL + CNN (1σ) achieves speeds 1.56, 1.37, and 2.25 times faster. The most significant improvements are obtained by using the CNN's predicted 1σ uncertainties to construct Gaussian priors in `PyAutoLens`, with particular benefits coming to the modelling of EAGLE galaxy lenses. Each of the modelling methods suffer from some outliers that take significantly longer to complete, however, PyAL + CNN (1σ) has far fewer and less severe outliers when compared to PyAL (blind). By applying priors that better constrain parameter space, `PyAutoLens` is able to converge upon a solution faster, as it spends less time exploring unhelpful regions of parameter space and is less likely to converge upon solutions corresponding to local minima in χ^2. This boost to computational efficiency relies on the selection of priors that accurately constrain

Table 2.4: **EAGLE galaxy lenses + HUDF sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images.

Method	θ_E (arcsec)	ϕ (°)	q
Without LOS structure			
CNN	0.035	8.70	0.108
PyAL (blind)	0.064	19.56	0.170
PyAL + CNN	0.044	9.12	0.116
PyAL + CNN (1σ)	0.034	8.12	0.119
With LOS structure			
CNN	0.046	12.6	0.114
PyAL (blind)	0.070	23.8	0.157
PyAL + CNN	0.060	15.4	0.121
PyAL + CNN (1σ)	0.043	12.4	0.114

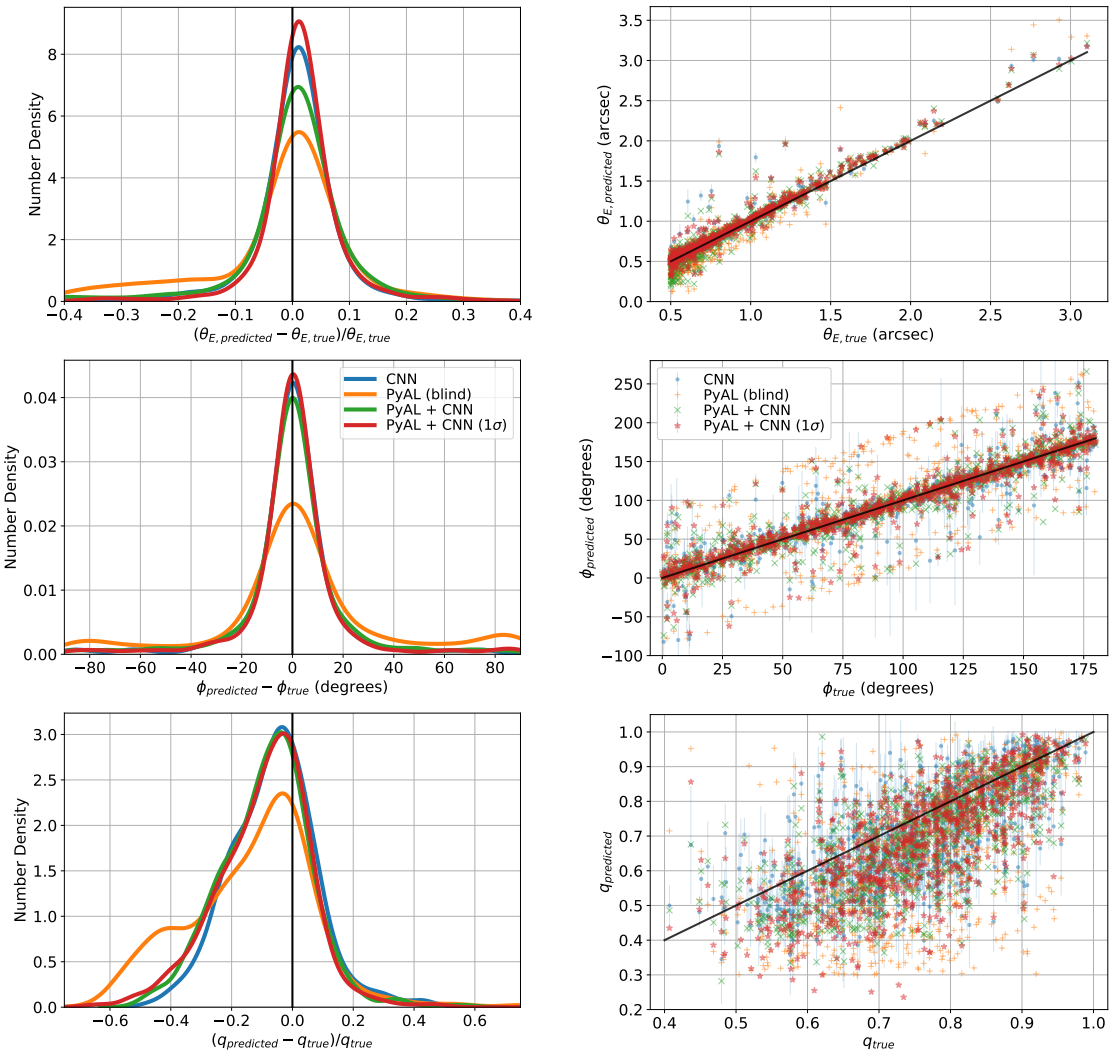


Figure 2.9: Left column: error distributions for the SIE mass model parameters for a test data set of 1000 images containing EAGLE galaxy lenses, HUDF sources and no line-of-sight structure. Einstein radius, and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

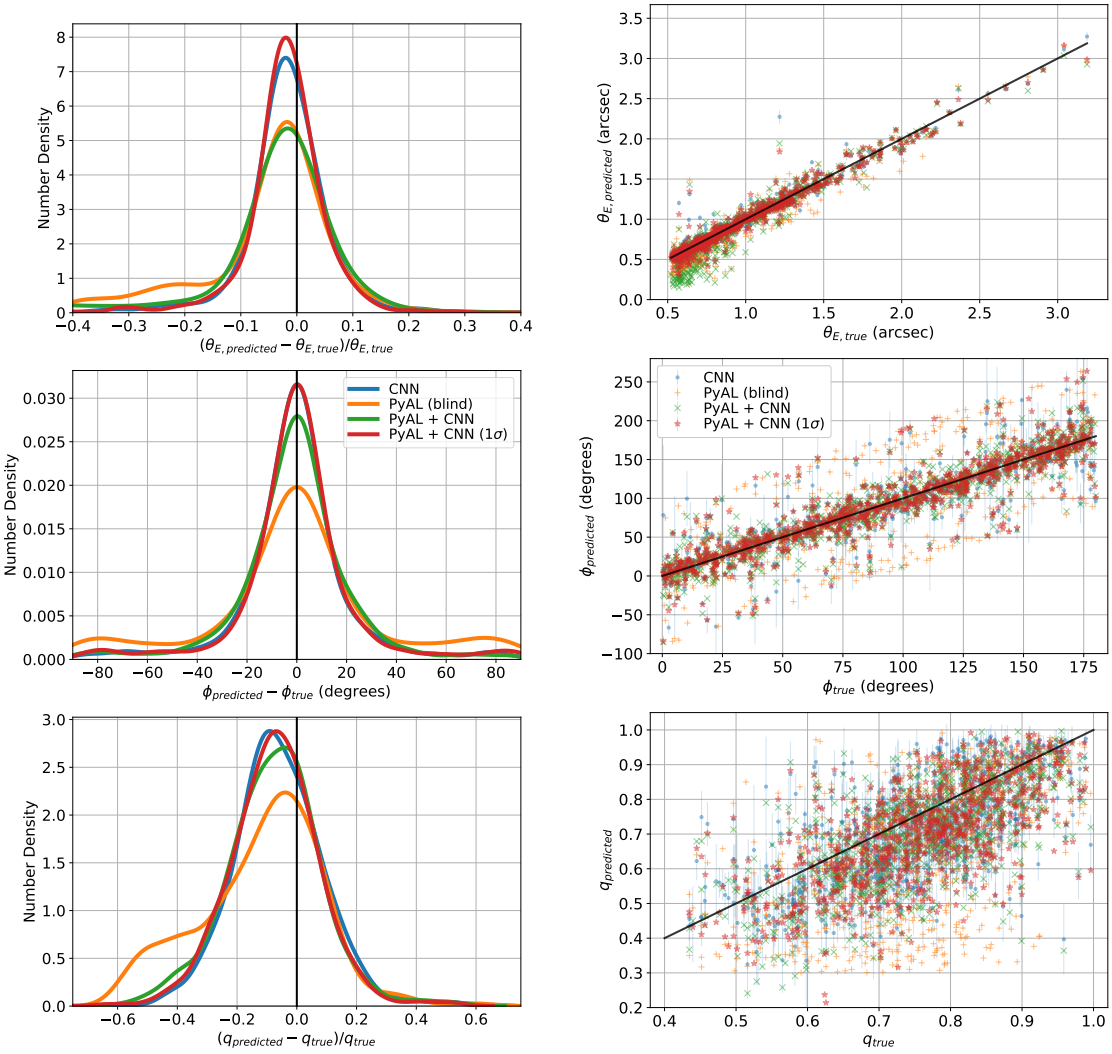


Figure 2.10: Left column: error distributions for the SIE mass model parameters for a test data set of 1000 images containing EAGLE galaxy lenses, HUDF sources and line-of-sight structure. Einstein radius, and axis ratio are given as fractional errors, whilst orientation is given as an absolute error. Right column: Comparison of the predicted vs true value for the same test data set, for each parameter. Figure reproduced from Pearson et al. (2021).

parameter space; in the case of the EAGLE galaxy lenses, the CNN under-predicted its 1σ uncertainties and so the priors applied within `PyAutoLens` were likely too restrictive, making it difficult for the proper exploration of parameter space around the correct solution. This problem is compounded by having too narrow priors that are not necessarily centred on the correct solution. Had this not been the case, we would expect to see a greater improvement to the modelling times and more accurate modelling results.

2.4 Discussion

We have compared the accuracy of a variety of automated approaches for modelling many thousands of simulated images of galaxy-galaxy strong lenses, using both SIE and power law mass models. The test sets had steadily increasing complexity and realism, beginning with parametric lenses and sources, and graduated to HUDF sources, and EAGLE galaxy lenses with line-of-sight-structure. We explored the feasibility of combining our CNN approach with that of `PyAutoLens`' semi-linear inversion implementation to produce a truly automated lens modelling pipeline that is both faster and more accurate than `PyAutoLens` alone. We examined two approaches to achieving this: One that uses the CNN predicted parameter values to centre uniform priors of a fixed width and pass these onto `PyAutoLens`, and another that also incorporates the CNN predicted 1σ uncertainties to define a Gaussian prior that can be applied to the `PyAutoLens` model.

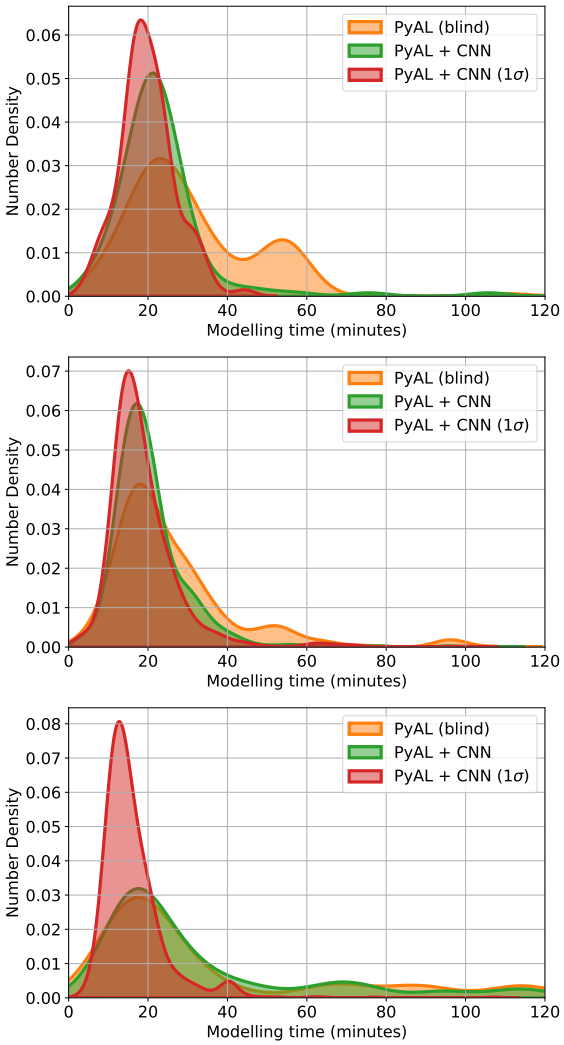


Figure 2.11: Distributions of the time taken by `PyAutoLens` to model the set of lenses making up a given test case. The distributions for PyAL (blind), PyAL + CNN and PyAL + CNN (1σ) are shown, but not the distribution of the CNN itself as this is almost instantaneous. From top to bottom: SIE lenses with HUDF sources, SIE lenses with HUDF source and LOS structure, and EAGLE galaxy lenses with HUDF source and LOS structure. Figure reproduced from Pearson et al. (2021).

2.4.1 Modelling Accuracy

In the simplest case of SIE lenses and parametric sources, all the methods we have considered perform extremely well, with errors below 0.01 arcsecond, 5° and 0.03 for Einstein radius, orientation and axis ratio, respectively. The story remains the same when we replace the background source with HUDF galaxies, with errors below 0.015 arcseconds, 7° and 0.04, respectively. In these test cases, the CNN's performance is comparable to that of PyAL (blind), however, we see a dramatic reduction in error when we combine our techniques, with PyAL + CNN (1σ) errors 37 per cent lower than PyAL (blind) on average. Incorporating the CNN's uncertainties into `PyAutoLens`'s priors leads to an improvement of 15 per cent over the errors of PyAL + CNN.

The modelling of power law profiles is significantly more challenging, and as such, we see a decrease in accuracy for all methods, with errors from 0.03 to 0.065 arcseconds, 2.5 to 7.4°, 0.06 to 0.11, and 0.2 to 0.3 for Einstein radius, orientation, axis ratio and slope, respectively; these correspond to factors of 1.1 to 4.2 times larger than when modelling an SIE profile. The scatter in predicted slope values is large for all our modelling methods, but we see significant improvements when combining the CNN with `PyAutoLens`, with PyAL + CNN (1σ) errors decreasing by 5-39, 28-66, and 32-62 per cent compared to PyAL + CNN, the CNN, and PyAL (blind), respectively.

The introduction of significant LOS structure complicates the results, increasing errors by factors of 2.2-8.2 for SIE lenses and by 1.3-9.2 for power law lenses. PyAL + CNN (1σ) and the CNN have comparable performance here, but both show significant improvement over PyAL (blind), with 27-36 per cent and 23-35 per cent lower errors, respectively. No accommodation for the LOS structure, such as allowing for an external shear or convergence field, was added into the model fitted by `PyAutoLens`,

and so it is no surprise that the performance was negatively impacted. Likewise, the CNN was not trained on images containing LOS structure, nor to predict external shear or convergence.

We see similar trends between each of the methods for the data set composed of EAGLE galaxy lenses. With no LOS structure, PyAL + CNN (1σ) achieves 30-58 per cent lower errors than PyAL (blind), whilst the CNN's errors are 36-55 per cent lower than PyAL (blind). With the inclusion of LOS structure, PyAL + CNN (1σ) and the CNN have almost identical improvements over PyAL (blind), with 28-48 per cent and 28-47 per cent lower errors. Even though the CNN was not trained on EAGLE galaxy lenses, or images containing LOS structure, it continues to be a powerful tool that can be used to enhance the performance of parametric modelling. The error in lens model parameters is seemingly dominated by the complexity of the lens galaxy, as the inclusion of LOS structure has a less pronounced effect, increasing errors on average by factors of 1.0 to 1.5. Further improvements to the performance of PyAL + CNN (1σ) would likely be made if the issue of the CNN underpredicting its uncertainties was addressed, through a more complete training set.

2.4.2 Modelling Speed

Not only does the combination of the CNN and `PyAutoLens` lead to an increase in accuracy, but it also results in faster modelling times than `PyAutoLens` alone. PyAL + CNN and PyAL + CNN (1σ) have modelling speeds factors of 1.19 and 1.73 times faster compared to PyAL (blind). The incorporation of more well-defined priors helps to reduce the number of outlier lenses that take large amounts of time to model; additionally, the problem of `PyAutoLens` converging upon local minima solutions is mitigated by the more restrictive priors. In this sense, the CNN can effectively replace

the need for significant human inspection and intervention in lens modelling, which is especially important in light of dealing with the expected large volume of data with upcoming surveys.

2.4.3 Modelling Difficulties and Limitations

As our results show, modelling power law profiles is more challenging than SIE profiles. The difficulty in accurately predicting slope values led to slight biases in our results, with over-predicted Einstein radii and under-predicted axis ratios. Our SIE results are, in general, unbiased, however, there exist a significant number of lenses where PyAL (blind) has under-predicted the Einstein radii and axis ratios. This behaviour is linked to `PyAutoLens` converging upon local minima corresponding to unphysical source reconstructions. These outliers are not seen in our other modelling methods, exemplifying the power of the CNN to constrain the parameter space.

The inclusion of LOS structure can simulate a lens residing within a galaxy group, with many tens of galaxies along the line-of-sight. This added complexity made lens modelling challenging with both the SIE and power law profile. The replacement of the parametric main lensing galaxy with hydrodynamical EAGLE galaxies further increased the difficulty of lens modelling. These lensing systems had fewer smooth, well-defined arcs, which can worsen the degeneracy between axis ratio and orientation (Mukherjee et al., 2018). Furthermore, these lenses had generally smaller Einstein radii than the parametric lenses, and thus produced smaller, more pixelated arcs, which provide less information for lens modelling. A key difficulty in accurately recovering the axis ratio of EAGLE galaxies is due to the fact that the axis ratio is able to vary as a function of radius of the galaxy. Additionally, many of these lenses contained notable substructure in their convergence maps, and as such, obtaining the

"true" parameter values via fitting to the convergence map proved difficult.

In this work, we have used a lower limit on magnification in order to ensure that strongly-lensed images are produced, focusing on sources in regions of higher magnification, and producing only around 30 - 40 per cent of double image lens-source configurations. It is, however, expected that doubles will dominate the discoveries of lensed systems in future surveys. We investigated the performance of the lens modelling methods for 100 doubles and 100 non-doubles, in the SIE lens and HUDF source test set. Errors in Einstein radii and axis ratios for doubles, were factors of 2.1 and 1.9 times larger, respectively, than non-doubles. Orientation errors were on average 0.8 times that of non-doubles, which were more likely to be near-complete Einstein rings with high axis ratios. Whilst the modelling errors depended upon whether a double-image was being examined, the relative performance of each methodology considered here remained much the same.

An important consideration for producing a fully automated lens modelling pipeline is the potential contamination from incorrectly identified lensing systems. `PyAutoLens` would likely produce extremely low evidence values if it attempted to fit a lens model to a non-lensed image, which could be used to indicate an issue with the process. The CNN, however, was not trained on such images and so its reaction to them is unknown. It may be that the predicted errors from the CNN would drastically increase, and this could be used to flag an issue with the data, but such investigation is left to further work.

2.5 Summary

Strong gravitational lensing is a unique and powerful tool for studying galaxy evolution, probing high-redshift source populations, and constraining cosmological models.

Our current catalogue of strong lensing systems is relatively small, but we expect it to grow to many tens of thousands of systems, thanks to upcoming large-scale surveys. Current lens modelling techniques, such as the parametric fitting of `PyAutoLens` require significant time investment from the user to setup and avoid common pitfalls during the modelling process, in addition to being relatively computationally slow. As a result, convolutional neural networks have been used to recover lens mass model parameters extremely efficiently, once the CNN has been properly trained.

In this work, we have compared the performance of an approximate Bayesian convolutional neural network with the semi-linear inversion technique of `PyAutoLens` for an array of increasingly complex lens simulations. These simulations included SIE and power law profiles, EAGLE galaxy lenses and line-of-sight structure. In addition to comparing these two methods, we explored the possibilities of combining them to leverage both of their strengths. We investigated two ways of doing this, one which made direct use of the CNN's predicted uncertainties to form priors and another that used priors of a fixed width.

Our results show that on average, the CNN's performance is comparable to that of PyAL (blind) for the majority of tests. This reflects the fact that `PyAutoLens` requires manual inspection of the images to properly constrain the lens model parameters and perform at its best. Combining these two methods, using only the CNN's predicted value and a fixed width prior, PyAL + CNN outperforms PyAL (blind), with errors 27 ± 11 per cent lower across the parameters. Additionally, the combination of these two modelling approaches leads to a speed-up of modelling times by a factor of 1.19. This method, however, is often matched in performance by the CNN.

The most significant improvement is attained when the CNN's predicted uncertainties are folded into the construction of `PyAutoLens`'s priors, PyAL + CNN (1σ). The CNN's predicted uncertainties are able to constrain the parameter space that

`PyAutoLens` must search, and this helps to ensure convergence upon the correct solution. This combined methodology achieves significantly higher accuracies than either method alone. The errors on PyAL + CNN (1σ) are 37 ± 11 per cent lower across the parameters compared to PyAL (blind), 17 ± 21 per cent lower compared to the CNN, and 13 ± 9 per cent lower compared to PyAL + CNN. This combination of the two techniques outperforms PyAL (blind) in all cases, as well as the CNN in the majority of testing scenarios. In addition to achieving the highest accuracy, modelling speeds are improved by a factor 1.73 on average across all test sets, compared to those of PyAL (blind). Particularly in the case of EAGLE galaxy lenses, the CNN tended to under-predict its uncertainties, leading to narrower than optimal priors. Addressing this issue would likely see even greater improvements to these results be realised.

Whilst the CNN's modelling speed makes it an attractive option for dealing with large numbers of newly discovered lenses, the fundamental limitation to its success lies in the quality of the training data. The complications and artefacts likely to be present in real data represent a significant challenge to overcome due to the difficulty in identifying poor fits to the data. However, the CNN still performed reasonable well when applied to the EAGLE lenses in this work, whilst having not been trained on such lenses. The PyAL + CNN (1σ) combination represents a promising automated modelling method that retains the flexibility and robustness of parametric modelling, whilst benefiting from the speed and accuracy of the CNN predictions. Rather than thinking of parametric modelling and machine learning techniques as adversarial techniques, this work highlights the potential for combining the two.

Chapter 3

Auto-identification of unphysical source reconstructions

As was discussed in Chapter 2, the number of strong lensing systems we know of is still relatively small, measuring in the hundreds. This is set to change in the coming years, with two significant surveys coming online: Euclid (Laureijs et al., 2011b); and the Vera Rubin Legacy Survey of Space and Time (LSST; Ivezic et al., 2008). It is expected that these surveys will discover many tens of thousands of lensing systems (Collett, 2015b); for this reason, the development of fast, automated, and accurate pipelines for finding and modelling strong lenses is of great importance. In Chapter 2, we explored the idea of combining machine learning approaches to lens modelling with those of parametric modelling; we used an approximate Bayesian neural network to predict lens model parameters and their 1σ uncertainties, which could then be used to initialise a subsequent round of modelling with `PyAutoLens`. We showed that the combination of these two techniques leveraged each of their strengths and formed a tool more accurate than either acting on their own. This represents a promising approach to fast, automated lens modelling. In this chapter, we aim to

address a common issue with parametric modelling, again through the application of machine learning methods, to further enhance the capabilities and robustness of a truly automated lens modelling pipeline.

When dealing with the lensing of an extended source, we wish to reconstruct both the source's intrinsic brightness distribution and model the mass distribution of the lens galaxy. One such method is that of semi-linear inversion (Warren & Dye, 2003), a technique that reconstructs the pixelised source in a linear step for a given lens model. This technique has been placed within a Bayesian framework for optimising the model evidence (Suyu et al., 2006), and more recent implementations reconstruct the source on an irregular grid of pixels that can adapt to the lens magnification or the source surface-brightness (Nightingale & Dye (2015); Nightingale et al. (2018); see Section 1.7 for a detailed description of the method). Another method for reconstructing the intrinsic source makes use of the family of polynomials known as shapelets (Birrer et al., 2015). An analytical reconstruction of the source can be formed using a small subset of these polynomials, leading to a reduced number of source parameters (Tagore & Jackson, 2016). Convolutional Neural Networks (CNNs) have been used to reliably and automatically recover the mass-model parameters of galaxy-galaxy strong lenses in orders of magnitude less time than traditional parametric techniques (Hezaveh et al., 2017b; Pearson et al., 2019, 2021). Furthermore, advancements have been made in the application of neural networks for reconstructing the background source of a strongly lensed system (Morningstar et al., 2019).

Techniques that model the lens mass with a parametric density profile remain a necessary and indispensable tool. There are significant difficulties involved in creating unbiased and sufficiently varied training sets for CNNs to learn from. This is a particular problem in the case of lensed high-redshift sources, where the source light is likely to be highly irregular. In addition, contamination of a lens data-set by objects

such as galaxy mergers and ring galaxies poses a problem for CNN based methods. In these circumstances, a CNN will produce a set of lens model parameters without any indication of failure, whilst parametric modelling techniques will fail to fit the data since they operate within the context of physically-motivated density profiles and are bound by the multiple imaging constraints of a real lens. Typically, it has been necessary to rely upon parametric techniques to obtain a robust measure of uncertainties on the lens parameters. Recently, however, methods for obtaining the uncertainties on CNN predicted parameters have been developed (Levasseur et al., 2017; Park et al., 2021).

A particular issue for methods based on the pixelised source reconstruction is the existence of unphysical solutions (see Section 3.1 for details). Such solutions are perfectly valid, providing excellent fits to the data, and can be challenging for sampling algorithms to avoid. An unsupervised modelling run can spend large amounts of time exploring the parameter space around these solutions and never converge towards the true parameter values. These solutions can be avoided with careful tuning of the model parameters, but this represents a significant investment of time for each system being modelled. For this reason, we have developed a CNN based approach to recognise these unwanted solutions and a simple prescription for updating the priors in our model to aid convergence towards the true solution. In this manner, we can iteratively improve our lens model by identifying and avoiding reconstructions that correspond to under and over-magnified solutions.

This chapter is organised as follows: Section 3.1 describes the occurrence of unphysical solutions in the modelling process and their properties. Section 3.2 discusses the methodology for simulating the required images of strongly lensed galaxies, and the processes involved to create source reconstructions from these images. Additionally, an overview of the CNN architecture is provided, with details on how the network

was trained and the manner in which the CNN was used in conjunction with our modelling process. The results of applying this technique to our testing set of data are presented in Section 3.3. Finally, the results in this work are discussed along with our conclusions in Section 3.4. Throughout this chapter, we assume a flat ΛCDM cosmology using the 2015 Planck results (Planck Collaboration et al., 2016), with Hubble parameter $h = 0.677$ and matter density parameter $\Omega_m = 0.307$.

3.1 Erroneous Solutions and Their Inversions

One of the key motivations for using the Semi-linear inversion method is the reduced computational complexity of the lens modelling process. Using analytic profiles to model the complex source light of a lensed galaxy can require exploring a highly multidimensional parameter space. Not only does this increase the likelihood of inferring a solution corresponding to a local maximum in evidence, but can also lead to biasing of the lens model due to constraints on the light profiles. Semi-linear inversion allows us to reconstruct the source light distribution in a linear step and since this distribution is pixelised, it is not constrained by an analytic profile. It does however introduce a new set of problems for the modelling process, namely under and over-magnified solutions.

These so-called under-magnified and over-magnified solutions can be understood in terms of the inferred amount of mass in the model lens galaxy. Here, we use the Einstein radius as a proxy for the mass of a galaxy. Ideally, the modelling process will converge upon the true value of the Einstein radius, along with the other model parameters, and the reconstructed source will reproduce the unlensed features of the source galaxy. If, however, the modelling process converges upon a solution with too small an Einstein radius, the resultant deflection angles will also be under-estimated.

This leads to the formation of an under-magnified image of the observation itself. Similarly, a model with too large an Einstein radius will over-estimate the deflection due to the lens. This will lead to an over-magnified, but this time, parity inverted image of the source. Fig. 3.1 illustrates this point with stylised ray diagrams for each class of source reconstruction we are considering.

Whilst these erroneous source reconstructions are obviously not the physical solution we are looking for, they exist nevertheless and can provide excellent fits to the data, thus posing a challenge for sampling algorithms to avoid them. Fig. 3.2 shows an example of another set of source reconstructions for a simulated observation. Here, we also show the residual and chi-squared maps for each reconstruction, showing the quality of the fit to the data. The Bayesian evidence is comparable for the under-magnified solution and the correct solution, whilst it is significantly lower for the over-magnified case. Generally speaking, we find that the under-magnified solution is much more probable to occur than the over-magnified one. This is likely due to the regularisation employed in the semi-linear inversion process. Regularisation serves to penalise overly complex solutions, which is certainly a characteristic of the over-magnified solution. In addition to regularisation reducing the likelihood of this solution, it can often be excluded by sufficiently accurate masking of the lens system. Provided the mask used when modelling the system does not extend considerably farther than the image separation, it can be used to set the upper-bound on the Einstein radius prior.

It is usually clear to the experienced modeller when something has gone wrong and an erroneous source reconstruction has been produced. It is not however so easy to discriminate between these solutions programmatically. Multiple techniques can be employed to avoid these solutions; careful tuning of the prior distributions on the lens model parameters is effective but time-consuming. For this reason, it is not a suitable

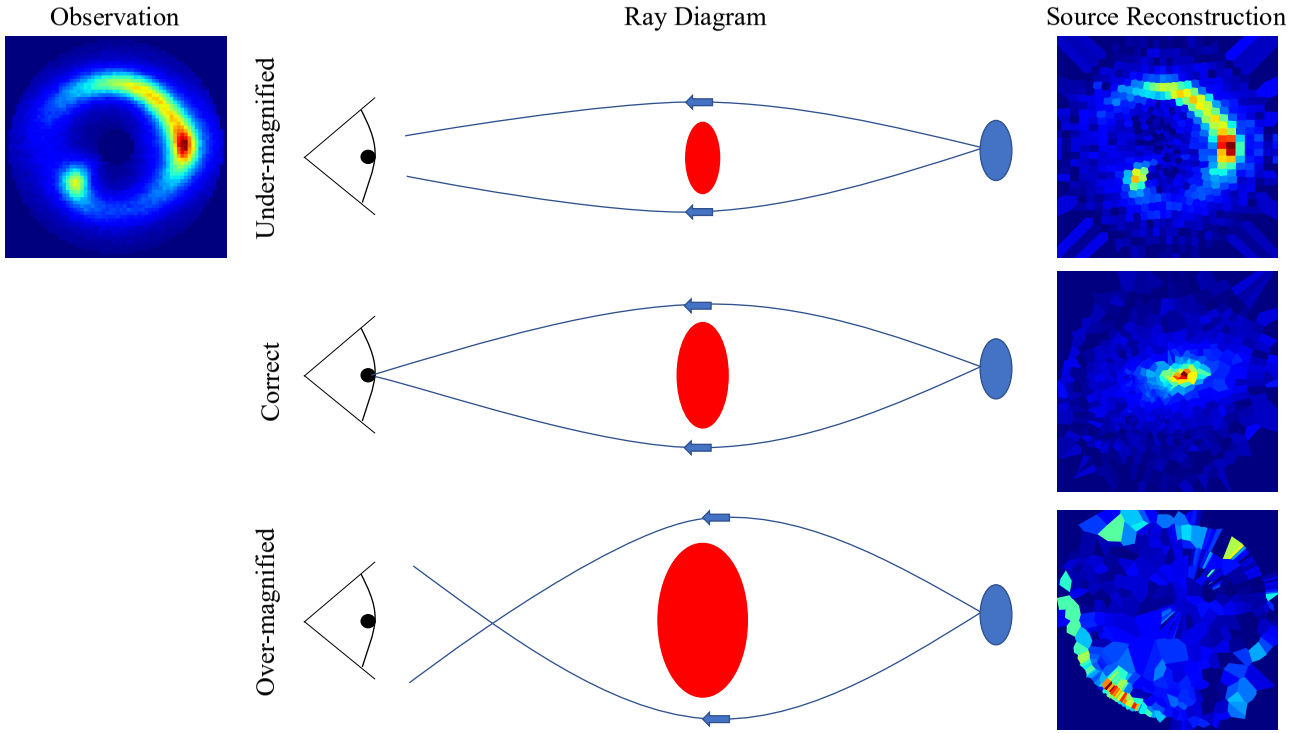


Figure 3.1: From left to right: Original observation, observer, lens galaxy, source galaxy, reconstructed source. From top to bottom: under-magnified solution, correct solution, over-magnified solution. Note the similarities in structure between the under-magnified solution and the original observation. Both images have similar morphology, but different angular extents. It is more challenging to see the similarities between the over-magnified solution and the original observation, but it is clear that the extended arc in the north of the observation is being reproduced in the south of the reconstruction, illustrating the parity inverted nature of the solution. The lensed image of each of the reconstructed sources considered here is almost indistinguishable from the observed image.

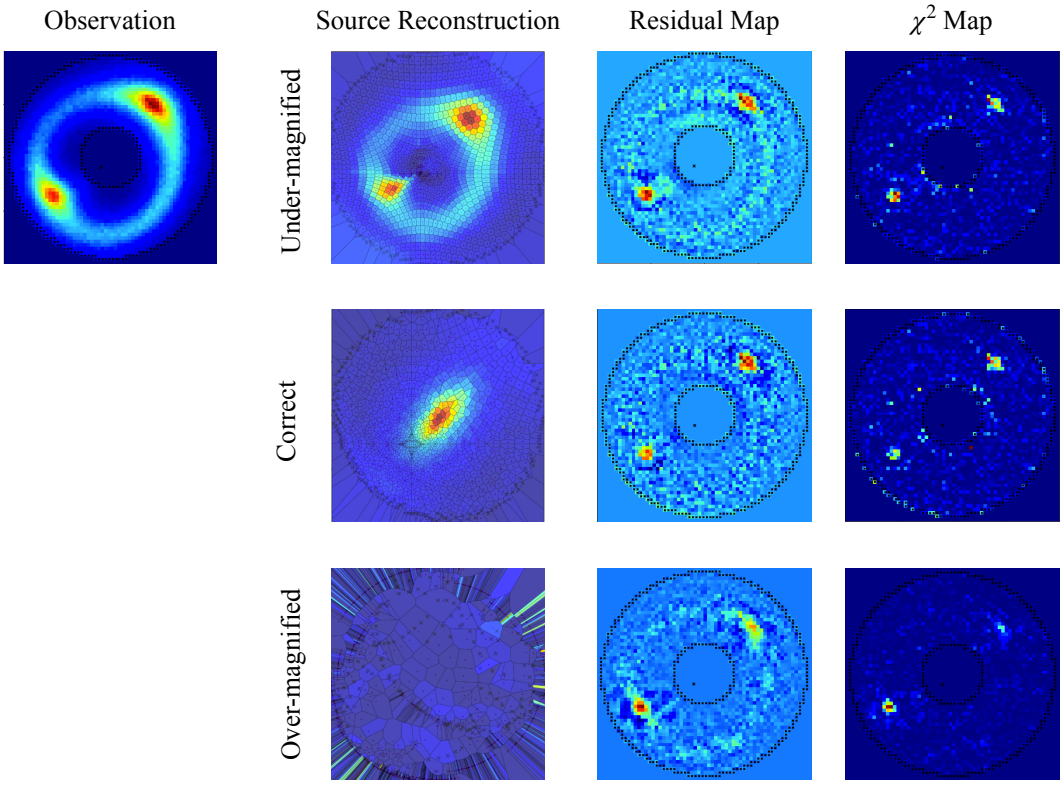


Figure 3.2: Column 1 contains a simulated observation of a lensed galaxy. Column 2 contains the source reconstructions for three classes of solution (under-magnified: row 1, correct: row 2, over-magnified: row 3). Column 3 contains the residual maps (model image - observation) for each class of solution. Column 4 contains the chi-squared maps (the squared significance of the residuals) for each class of source reconstruction. The lensed images of each class of source reconstruction demonstrated here are almost indistinguishable from the observed image.

method for dealing with the large numbers of lensed galaxies we expect to encounter in the coming years. Another possibility, which has the benefit of being automatic, is to create a pipeline of models that first fits an analytic light profile to the source galaxy and then uses the results of this fit to initialise the priors in the inversion process (Nightingale et al., 2018). By requiring a compact source in the initial phase of modelling, the aim is to infer a lens model sufficiently accurately to effectively rule out regions of parameter space that would correspond to under or over-magnified solutions. This lens model, along with new priors on its parameters, is then used in the inversion process to refine the lens model and more accurately fit the source galaxy's light. The complex morphology of high-redshift sources poses a challenge for fitting the data with an analytic light profile, which can lead to a poorly constrained or entirely wrong lens model. If the inferred lens parameters used to initialise the model in the inversion process are of poor quality, then the modelling can once again fail at this step. Even if a fit with an analytic source profile provides a reasonable initialisation for the inversion process, it is challenging to constrain the width of the subsequent prior distributions such that erroneous source reconstructions are ruled out but feasible lens models that fit the more complex source are not. Our approach to this challenge is to use a CNN that can accurately classify source reconstructions as successful or under/over-magnified. In this way, we completely remove the need to assume an analytic light profile for the source, since we can throw away unwanted solutions in the inversion process that do not correspond to a compact reconstructed source. Furthermore, we have developed a simple method for updating the model to move away from these unwanted solutions towards the correct parameters. This technique requires no human intervention and the CNN classification step is extremely fast (<1s).

3.2 Methodology

The CNN described in this work requires training data consisting of labelled source reconstructions and residual images. To produce these data, it was first necessary to create a large number of simulated strong gravitational lens images. We used the lens modelling software `PyAutoLens` [1] (Nightingale & Hayes, 2020; Nightingale & Dye, 2015; Nightingale et al., 2018) to produce our simulated images and to perform the source reconstruction. MultiNest (Feroz et al., 2009) was used for the exploration of parameter space where a full analysis of the data was carried out. The modelling process produces the residual images between the simulated observations and the reconstructed model image that we need for training the CNN. In Section 3.2.1 we describe our procedures for generating the simulated strongly lensed images. Section 3.2.2 details our method for generating the source reconstructions and residual images required for training our neural network. We then describe the CNN architecture used in this work in Section 3.2.4. The process used to update the prior distributions on the model, based on the CNN predictions is then detailed in Section 3.2.5.

3.2.1 Lensing Simulations

In this work, we have assumed that all the foreground deflectors are early-type galaxies, and so we have adopted the Singular Isothermal Ellipsoid (SIE) mass profile (Keeton, 2001). For the light profile of the background lensed galaxies, we have opted to use the Sérsic profile since it can represent a wide variety of galaxy morphologies.

The data sets generated for this work were simulated to have distributions of parameters similar to those observed in the Sloan Lens ACS (SLACS) survey (Bolton et al., 2006). The Einstein radius θ_E and axis ratio q of our lensing galaxies were drawn

[1] https://github.com/Jammy2211/PyAutoLens

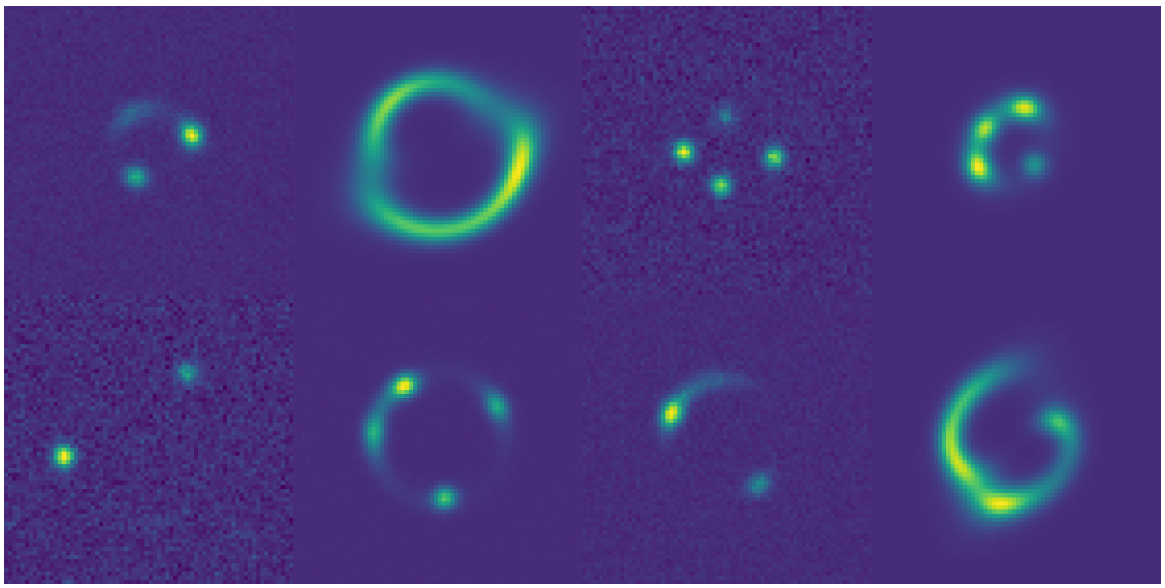


Figure 3.3: A selection of simulated images produced for this work, used for creating pixelised source reconstructions to train a CNN. All images have a pixel scale of 0.1 arcsec pixel^{-1} and each image's color scale has been normalised to the peak signal of the image.

from distributions fitted to the measurements of 131 strongly lensed galaxies observed in the SLACS survey (Bolton et al., 2008), whilst the orientation ϕ was allowed to vary uniformly over the full range. The Einstein radii of our lenses were drawn from a normal distribution with mean $\mu = 1.16$ and a standard deviation $\sigma = 0.42$. The axis ratios of our SIE profiles were randomly sampled from a normal distribution with mean $\mu = 0.80$ and standard deviation $\sigma = 0.16$, in close agreement with empirical studies (Koopmans et al., 2006b). In all cases, the centroid of the lens was placed in the centre of the image. In this work, we did not include light from the lens galaxies in the simulations.

As with the lenses, the parameters describing our source galaxy Sérsic profiles were randomly sampled from fitted distributions. In this case, we used the inferred Sérsic parameters from the parametric source reconstructions of a subset of the SLACS lenses (Newton et al., 2011). The Sérsic indices n, of our sources, were randomly drawn from an exponentially modified Gaussian distribution, with a probability density function of the form

$$f(x;\mu,\sigma,\lambda) = \frac{\lambda}{2}\mathrm{Exp}\left(\frac{\lambda}{2}(2\mu+\lambda\sigma^2-2x)\right)\mathrm{erfc}\left(\frac{\mu+\lambda\sigma^2-x}{\sqrt{2}\sigma}\right) \tag{3.1}$$

with scale parameter $\lambda = 0.723$, mean $\mu = 0.71$, standard deviation $\sigma = 0.97$, and erfc representing the complementary error function. The effective radii r_{eff} of our sources were randomly sampled from an exponential distribution with scale parameter $\lambda = 6.64$. We allowed the axis ratio of the sources q_s to vary uniformly over the range [0.3, 1]. The overall intensity normalisation I of the sources was drawn from a uniform distribution $I \sim U[10, 20]$ electrons s^{-1}, allowing for a wide variety of signal-to-noise ratios in our training data. The centroid of each source was uniformly distributed in the source plane, with the requirement that it lay inside the Einstein radius of the

lens (i.e. that there are multiple images).

In the production of our simulated images, we opted to use the pixel scale of the VIS instrument for Euclid (0.1 arcsec pixel^{-1}) and the characteristic exposure time of 565 seconds (Cropper et al., 2016). The lensed image was then convolved with a Gaussian point spread function with a full-width at half-maximum of 0.17 arcseconds. A background sky of 1 electron s^{-1} and Poisson noise due to the background sky and source light photon counts were added to the images, thus completing the simulation procedure. Some examples of our simulated images are shown in Fig. 3.3.

3.2.2 Training Data

The CNN was not trained directly on the simulated images, but rather the pixelised source reconstructions and residual images obtained from the modelling process. Before the modelling began, each simulated image needed to be masked to ensure that only the area of interest was reconstructed in the source plane and to reduce the computational load. Due to the large number of simulated images, an automated masking scheme was used. Firstly, the images were thresholded using the minimum cross-entropy approach (Li & Lee, 1993). Then, the centroid of this thresholded image was found through calculating its moments. A circular annular mask, centred on the centroid of the image, was then fitted to the thresholded pixels. For the inner radius of the annulus, the largest radius circle that did not contain any unmasked pixels was found, and 90 per cent of this value was used. Similarly, for the outer radius, the smallest circle containing all the unmasked pixels was computed, and 110 per cent of this value was used. These adjusted values for the inner and outer radii of the mask were used to minimise the chances of masking out faint emission from the source.

These masked images were then modelled using `PyAutoLens` to produce the pix-

elised source reconstructions and residual images that we need for training our CNN. In all cases, we adopted the SIE mass profile to model the lens galaxy. We reconstructed the background source on a pixelised grid that adapts to the magnification of the system. For each simulated lensed image, we created three source reconstructions and three residual images, corresponding to the under-magnified, over-magnified, and correct solutions. This resulted in approximately 300,000 images to be used as training data for our network. To deal with such a large computational task, it was necessary to employ some approximate methods in the source reconstruction/lens modelling process.

For 250 of our simulated images, we performed a full analysis of the data, optimising the lens model and source parameters in the inversion process. In each case, the analysis had to be repeated three times, to produce the under-magnified, correct and over-magnified source reconstructions. When modelling each of these systems, we allowed the mass-model parameters to vary uniformly over the full range of parameter space with the exception of the Einstein radius. To produce an under-magnified source reconstruction, we set a uniform prior distribution on the Einstein radius with an upper limit of 0.9 times the true value for the system, thus forcing `PyAutoLens` to find the under-magnified solution. To produce source reconstructions corresponding to the correct model, we allowed the Einstein radius to vary over a small range centred on its true value, guaranteeing that a sensible source reconstruction is produced. Finally, to produce over-magnified source reconstructions, we allowed the Einstein radius to vary over a range of 1.1 times the true value up to 3 times this value, again forcing `PyAutoLens` to find the over-magnified solution. In this manner, we built up an understanding of the properties of each class of source reconstruction.

In these tests, we observed that the mean fractional error in Einstein radius when producing an under-magnified source reconstruction is $\hat{f}_{\theta_E} \approx -0.5$. As expected, we

observed no significant bias in the Einstein radius, or any of the other parameters, when using a model with priors accurately centred on the true parameter values. The mean fractional error in Einstein radius when producing over-magnified reconstructions was $\hat{f}_{\theta_E} \approx 2$. A scatter plot of the true value of Einstein radius versus the inferred value for each class of source reconstruction is shown in Fig. 3.4 along with the coefficients of a linear fit to the data. These fitted parameters allowed us to define an approximate transformation of the Einstein radius, taking us from one class of source reconstruction to another. We found that the Einstein radius was the key parameter in controlling which class of source reconstruction was obtained. Fig. 3.5 shows that in both cases of erroneous source reconstructions, the axis ratio of the lens is most often under-estimated, but it does not follow an easily predictable pattern in the same way as the Einstein radius. Fig. 3.6 shows that there is no apparent relationship between the inferred orientation of the mass profile and its true value when either the under-magnified or over-magnified solution is found.

The relationship between the unphysical reconstructions and the correct solution allowed us to rapidly generate source reconstructions without the need for a full optimisation of the lens model. Fig. 3.4 shows how the predicted value of the Einstein radius relates to the true value in each of the three classes of source reconstruction we are considering here. The coefficients of a linear fit to the data allow us to construct an approximate transformation of the predicted Einstein radius to the true value for a given system. As expected, in the case of successful source reconstructions, the inferred value for the Einstein radius very closely matches the true value. The under-magnified solutions have inferred Einstein radii, $\hat{\theta}_U$ that can be approximated as $\hat{\theta}_{\mathrm{U}} \approx 0.46\theta_{\mathrm{E}} - 0.08$, where θ_E is the true value for the system. Similarly, in the case of over-magnified solutions, the inferred Einstein radii $\hat{\theta}_O$ can be approximated as $\hat{\theta}_0 \approx 2.11\theta_{\mathrm{E}} + 0.16$. Using these approximate transformations, along with the true

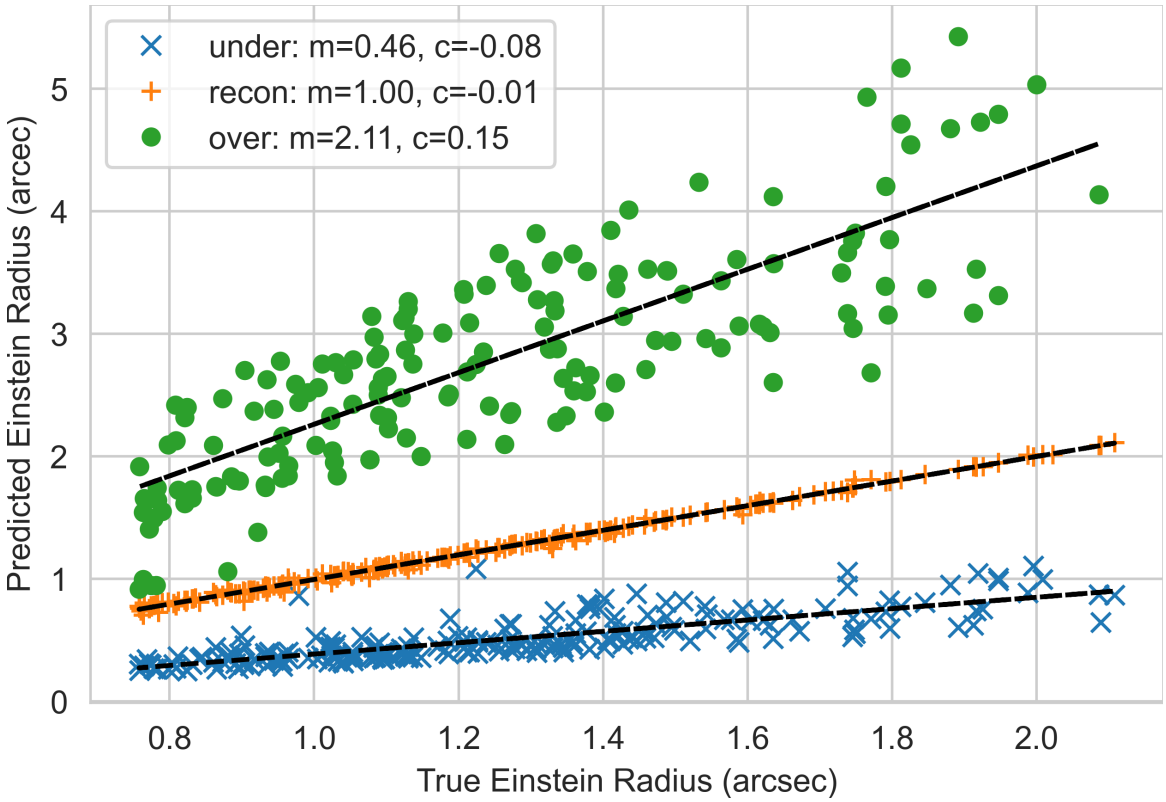


Figure 3.4: The relationship between the true value of θ_E and the predicted value $\hat{\theta}_E$ follows a predictable pattern for each class of source reconstruction. In each case, the coefficients of a linear fit are shown; These are used to construct the transformation of the prior distribution on θ_E to converge upon the correct solution.

Table 3.1: Summary of the approximate transformations of Einstein radius, linking the inferred values in a given class of solution to the true value.

CNN Prediction	Transformation	Updated Prior
Under	$\theta = \frac{1}{0.46}(\hat{\theta}_U + 0.08)$	$\theta_E \sim U[\theta \pm 0.25]$
Recon	$\theta = \theta_R$	$\theta_E \sim U[\theta \pm 0.25]$
Over	$\theta = \frac{1}{2.11}(\hat{\theta}_O - 0.15)$	$\theta_E \sim U[\theta \pm 0.5]$

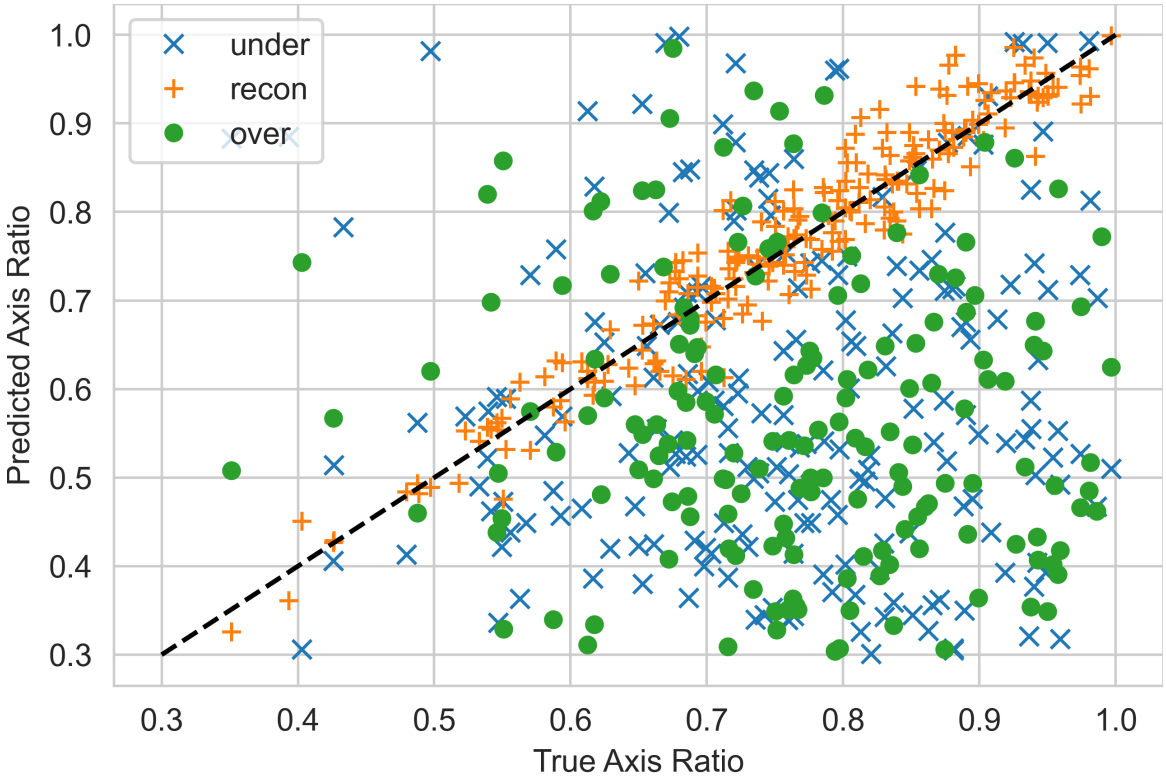


Figure 3.5: The relationship between the true axis ratio of the lens q and the inferred value $\hat{q}$ for each class of source reconstruction. For successful source reconstructions, q and $\hat{q}$ are directly proportional to one another, but in the case of under/over-magnified solutions there is no obvious trend aside from a tendency for $\hat{q}$ to under-estimate q.

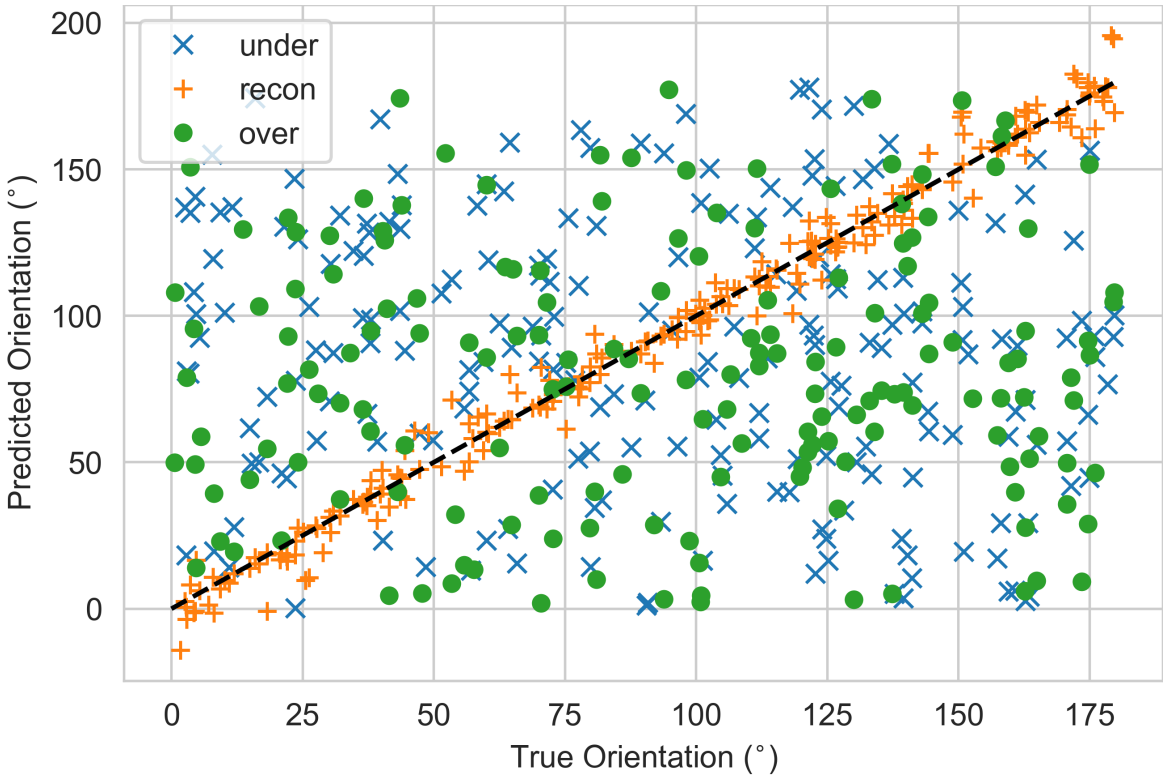


Figure 3.6: Successful source reconstructions accurately recover the lens orientation, ϕ, whilst under/over-magnified solutions infer a value $\hat{\phi}$ that appears to have no relationship to ϕ.

parameters describing the lens, we identified the regions of parameter space where we expect each class of source reconstruction to occur. Keeping the position, axis ratio and orientation of the lens fixed to the truth, we varied the Einstein radius around its expected value and computed the linear inversion for each sample. The inversion achieving the highest evidence is considered to be the solution, and we record the source reconstruction and residual image for our catalogue of training data.

3.2.3 Testing Data

A portion of the training data, produced as described in Section 3.2.2, was set aside for evaluating the CNN's performance after training. These simple source reconstructions allowed us to test the network on a set of images with similar properties to the training data. In addition, to explore whether our CNN trained on reconstructions of simple parametric sources would be capable of classifying the reconstructions of more complex lensed sources, we produced SIE-lensed images of high redshift galaxies extracted from the Hubble Ultra Deep Field (HUDF; Beckwith et al., 2006b). For this, we used the Pipeline for Images of Cosmological Strong lensing (PICS; Li et al., 2016b), simulating images to have the expected properties of Euclid VIS data (Cropper et al., 2016; Niemi, 2015). A sample of these simulated images is displayed in Fig. 3.8. For each of these simulated images, we produced a source reconstruction corresponding to the under-magnified, over-magnified and accurate solution, following the same full analysis procedure described in Section 3.2.2. These source reconstructions, along with the residual images of the models that produced them, were used to test the CNN's classification ability on significantly more complex images than it was trained on. A sample of the accurate HUDF source reconstructions is shown in Fig. 3.9.

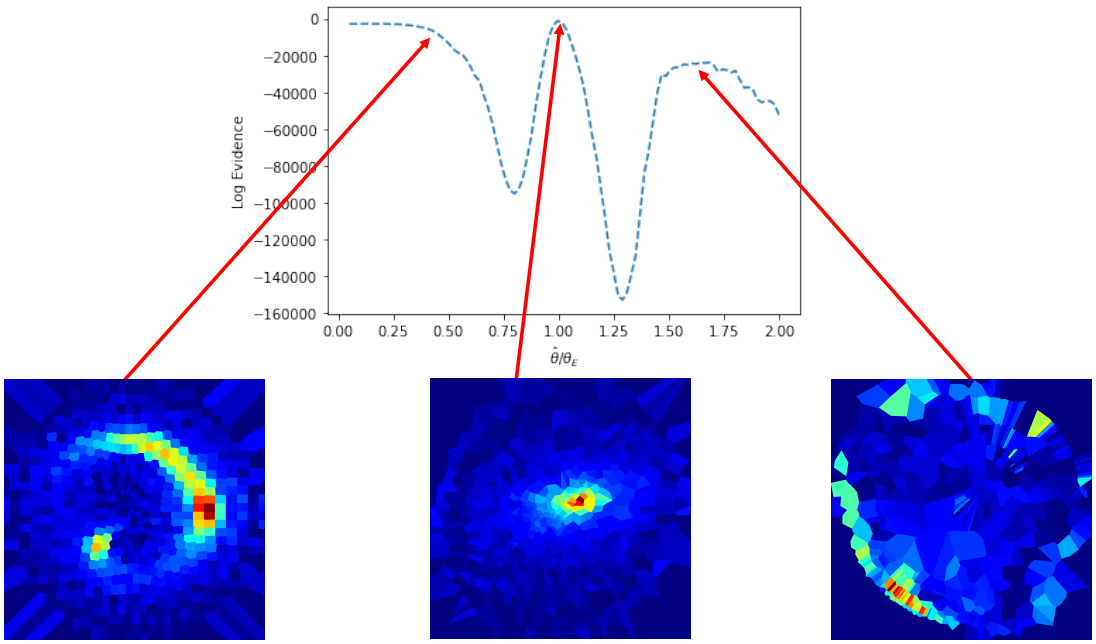


Figure 3.7: A one dimensional slice through Einstein radii values (normalised by the true Einstein radius), with all other mass model parameters fixed to their true values. Shown below from left to right: Under-magnified source reconstruction, successful source reconstruction, over-magnified source reconstruction.

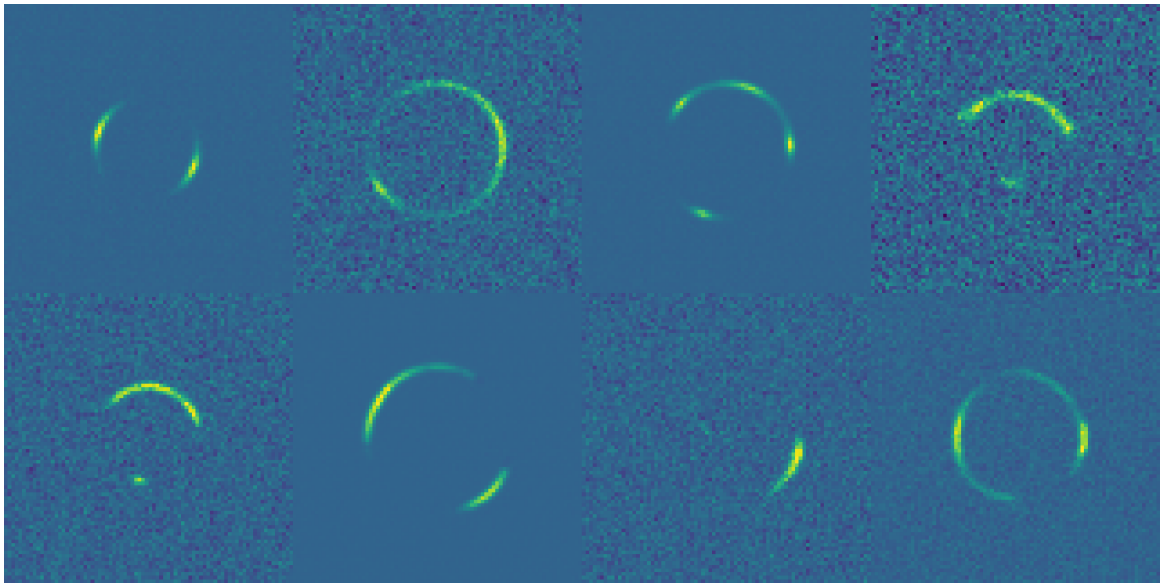


Figure 3.8: A selection of simulated SIE lenses with HUDF background sources, used to test the performance of the CNN on complex reconstructions. All images have a pixel scale of 0.1 arcsec pixel^{-1} and each image's colour scale has been normalised to the peak signal of the image.

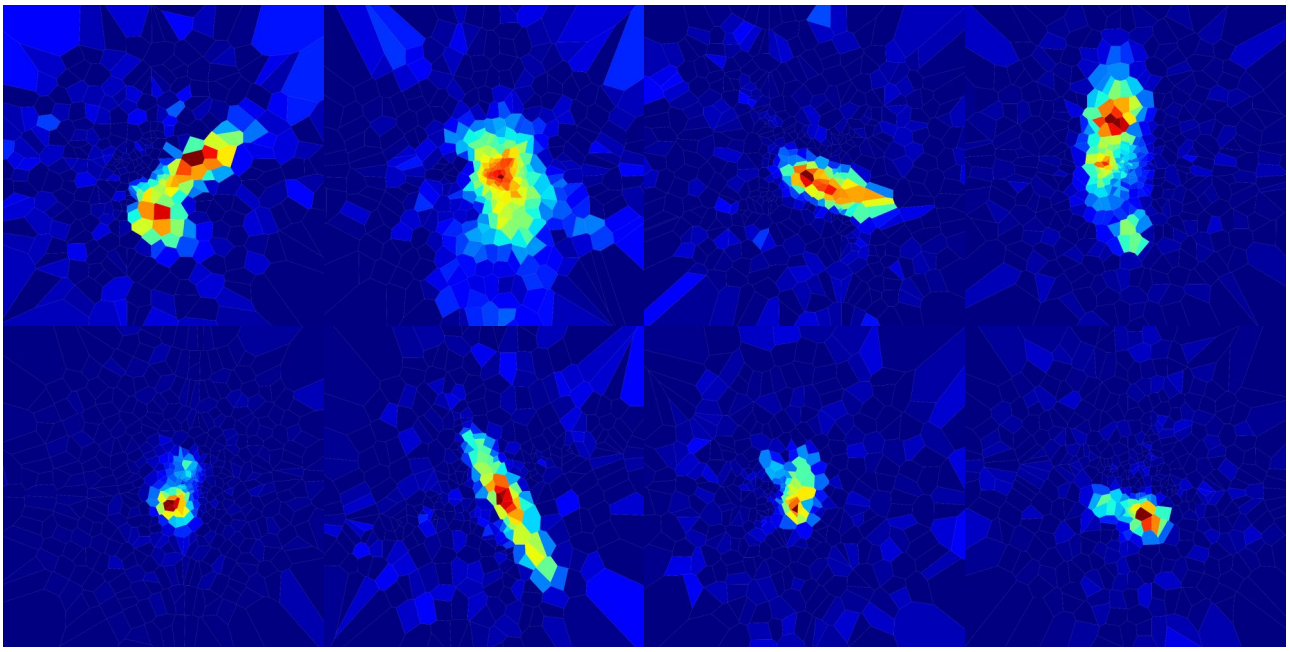


Figure 3.9: A selection of accurate source reconstructions of SIE-lensed HUDF galaxies. These reconstructions were used to test CNN performance on more complicated sources than the simple parametric sources used to create its training sample.

3.2.4 CNN Architecture

For a general discussion of CNNs and their operation, see Section 1.8.1. In this section, we describe in detail the architecture of the CNN used to carry out the work described in this chapter.

The network we built to classify our source reconstructions has a forked design, with two input paths. Each path consists of three convolutional layers and three max-pooling layers. The outputs of both paths are concatenated, before being flattened and fed into two fully connected layers. Dropout is employed between each layer to improve the network's resistance to over-fitting and the Leaky Rectified Linear Unit activation (Leaky ReLU; Nair & Hinton, 2010) function is used everywhere except for the final layer which employs the Sigmoid activation function. The Leaky ReLU activation function allows a small positive gradient for negative input values.

The tuneable hyperparameters for our network, such as the number of convolutional layers, the size of the kernels and the dropout rates were set by a process of hyperparameter optimisation. We opted to use Talos (Autonomio, 2019) to automate the evaluation of model performance. In order to explore the very large parameter space, it was necessary to down-sample and look at a small fraction of combinations of parameter values. Once a rough estimate of hyperparameters had been obtained, a more thorough search was carried out in a smaller region of parameter space.

The network aims to predict the category of source reconstruction that a given input belongs to. To train the network, pairs of source reconstructions and the corresponding residual images are fed into the network in batches. The error on a prediction is determined via the categorical cross-entropy loss function (LOSS),

$$\text{LOSS} = -\sum_{i=1}^{3} y_i \cdot \log \hat{y}_i \tag{3.2}$$

where y_i is the target value and $\hat{y}_i$ the predicted value. The network optimisation used the Nadam optimiser, which is a combination of stochastic gradient descent and Nesterov momentum (Dozat, 2015). The CNN was trained and tested on a GPU machine, vastly improving the time taken to process large numbers of images. The training took place over 50 epochs, using 120,000 pairs of images.

The weights and biases of the network are summarised as follows:

- Convolutional layer: For an input image of height x_1 and width x_2, the input is an (x_1, x_2, 1) matrix. The output of a convolutional layer is an (x_1, x_2, N) matrix, where N is the number of output filters applied in the convolution. Training adjusts the biases and weights for each filter, but their values remain fixed during each iteration. Each kernel of dimension (k_1, k_2) has an associated bias, giving a total of $k_1 \times k_2 \times N$ weights and N biases for each convolutional layer. The exact dimensions of each kernel are given in Fig. 3.10.

- Max-pooling layer: Pooling applies a 2×2 kernel with a stride of two, resulting in an output of dimension ($\lfloor x_1/2 \rfloor$, $\lfloor x_2/2 \rfloor$) for an input of (x_1, x_2).

- Concatenate: After the three convolutional layers in each input path of the network, the outputs are concatenated to form a tensor with dimensions (13, 13, 256).

- First fully connected layer: The input is a flattened 43,264-node array, whilst the output is a 512-node array. Accordingly, there are $43,264 \times 512$ weights and 512 biases.

- Final layer: The input is an array with 512 nodes, whilst the output is a 3-node array (one node for each class of source reconstruction), hence there are 512×3 weights and 3 biases.

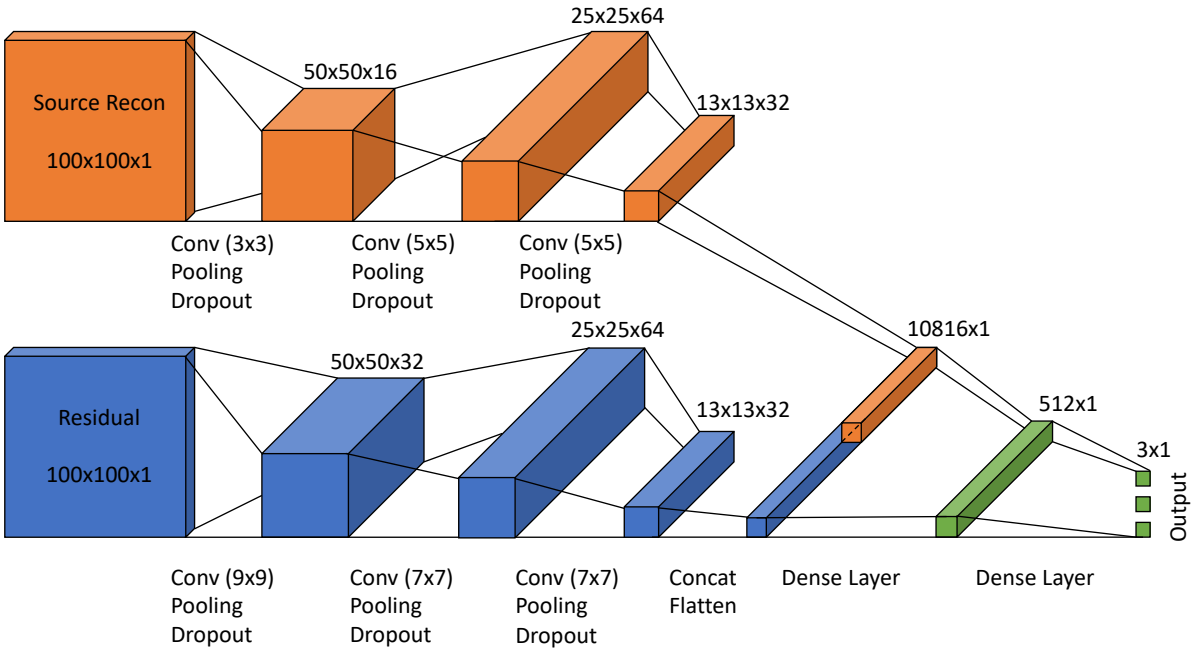


Figure 3.10: Structure of the CNN used in this work, showing the two input images and their respective paths in the network. There are six convolutional layers, each with max-pooling and dropout. A concatenation and flatten layer is included to join the outputs of the dual convolution pathways and connect this tensor with a 1D dense layer. LeakyReLU is used throughout the network, except for the activation of the final layer, which uses the Sigmoid activation function. The types of layers in the network at each step are given, along with the size of the kernel in pixels. The output dimensions are indicated above each block. A more detailed description can be found at the end of Section 3.2.4.

- There are a total of 5,820,323 trainable parameters.

3.2.5 Combining CNN and Lens Modelling

The trained CNN is capable of taking a source reconstruction and a residual image, both of which are outputted in the lens modelling process, and returning an accurate prediction of whether the correct lens model has been found, or whether an under/over magnified solution has been identified. This prediction, along with the knowledge of how the inferred Einstein radius relates to each class of solution, allows us to automatically correct the modelling process when erroneous solutions are found. Using the approximate transformations given in Table 3.1 we can update the model's prior distribution on θ_E for subsequent modelling. In this way, we aim to improve the robustness of our modelling process against unwanted solutions and reduce the amount of human intervention required to produce accurate lens-models and source reconstructions. When considering the predictions of our CNN, we will use the abbreviation UM to refer to a predicted under-magnified solution, OM for a predicted over-magnified solution and C for when the network predicts a correct reconstruction.

To test this hybrid approach to lens modelling, we simulated a new set of 100 lensed images, following the approach detailed in Section 3.2.1. We used `PyAutoLens` to model each system, conducting a full analysis, allowing all the SIE mass-model parameters to vary and reconstructing the background source on a magnification based Voronoi grid (see Section 1.7). For all the mass-model parameters, as well as the source plane pixelisation parameters, we opted to use uniform distributions covering a suitable range of parameter space. We chose a uniform prior distribution for the position of the lens centroid, centred on the true value with a width of 0.6 arcseconds. In the case of the orientation ϕ of the lens, we allowed the full range of

values $\phi \sim U[0, \pi]$ radians. The axis ratio of the lens, q was able to vary over the full range of values included in the simulations $q \sim U[0.25, 0.999]$. Again, the prior distribution of the Einstein radius θ_E followed a uniform distribution constrained only by the dimensions of the annular mask (computed according to the criteria detailed in 3.2.2, $\theta_E \sim U[r_{min}, r_{max}]$. Such an approach to modelling the data was taken to show the extremes of how things can go wrong without some tuning of the priors before modelling begins. Furthermore, this serves to illustrate the problems experienced by sampling algorithms when exploring large and complex parameter spaces.

Once this initial round of modelling was completed, the source reconstruction and residual images were fed into our CNN to obtain a prediction on whether the modelling had been successful or not. The next step in the process depends on the prediction of the CNN, as follows:

- UM prediction: The modelling process is repeated with an updated prior distribution on the Einstein radius. This new prior is defined in Table 3.1. The prior distributions on the other free parameters were left unchanged.

- C prediction: In this instance, we choose to repeat the modelling process with a decreased evidence tolerance and a narrowed uniform prior distribution centred on the inferred values from the previous modelling run. The goal of this repeated run is to more thoroughly explore the parameter space around the accepted solution and improve the accuracy of the model.

- OM prediction: The modelling process is repeated with an updated prior distribution on the Einstein radius, whilst leaving everything else unchanged. This new prior is defined in Table 3.1.

After this additional stage of modelling, the updated source reconstructions and residual images were fed into the CNN once more, providing a new prediction for each

system. With this information, we proceeded similarly to before, but now we take into account the history of results for each system.

- UM prediction:
  - If the previous prediction was also UM, then the system is flagged for manual intervention at a later time. This indicates that the process for updating the priors was unable to move the model away from this solution, or that the CNN has misclassified a reconstruction.
  - If the previous prediction was OM, this indicates that the prior update has 'overshot' the C solution, and so a uniform prior on the Einstein radius is chosen to lie between the two previous values. The width of the prior was set such that it excludes the regions of parameter space that corresponded to the previous under and over-magnified solutions.
- C prediction:
  - If the previous prediction was UM, as before, we chose to repeat the modelling process with a decreased evidence tolerance and use narrowed uniform prior distributions centred on the inferred values from the previous modelling run.
  - if the previous prediction was C, no further action required.
  - If the previous prediction was OM, again, we choose to repeat the modelling process with a decreased evidence tolerance and use narrowed uniform prior distributions centred on the inferred values from the previous modelling run.
- OM prediction:

- If the previous prediction was also OM, then the system is flagged for manual intervention at a later time. This indicates that the process for updating the priors was unable to move the model away from this solution, or that the CNN has misclassified a reconstruction.
- If the previous prediction was UM, this indicates that the prior update 'overshot' the correct solution and so a uniform prior on the Einstein radius is chosen lying between the two previous values. The width of the prior is set such that it excludes the regions of parameter space that corresponded to the previous UM and OM solutions.

This process can be repeated many times until an acceptable fraction of the CNN's predictions are that the correct model has been found. In practice, due to the crude nature of the prior-updating routine, there are diminishing returns on repeated cycles. The systems that become manually flagged during this process will need human intervention to guide the modelling to a suitable solution, but the overall load on the modeller is greatly reduced.

3.3 Results

In this section, we present the results of testing our CNN on the reserved data-set, evaluating its performance on a per-class basis. We show that the CNN performs exceptionally well at the task of classifying source reconstructions. Additionally, we show the result of modelling 100 simulated observations using the procedure outlined in Section 3.2.5. This set of images were simulated according to the procedures outlined in subsection 3.2.1. Here, we opted to apply our iterative approach three times, observing good progress towards a complete sample of successfully modelled lenses with each step.

3.3.1 CNN performance

The CNN was trained on 130,000 pairs of source reconstructions and residual images, for 50 epochs. 10,000 pairs of source reconstructions and residual images were used as validation data throughout the training process. To further increase the variety in the training data, augmentation techniques were employed. Each pair of images was randomly reflected horizontally, vertically or rotated through an angle. The remaining 6,928 pairs of images were reserved as a testing set to evaluate the performance of the network on never before seen images once training had completed.

Fig. 3.11 shows the confusion matrix C for the CNN evaluated on the testing data set. The elements of this matrix are defined such that $C_{i,j}$ contains the number of true objects of class i predicted to be in class j. Thus, the diagonal elements of C represent the correctly labelled instances and the off-diagonals where the network has incorrectly labelled an observation. The values displayed in C are normalised over the rows. The CNN's recall or ability to find all samples of a particular class is above 99.9 per cent in all cases, and performed perfectly on our test set for both under and over-magnified source reconstructions. Similarly, our CNN's precision, or ability to not label a sample of Y as X is greater than 99.9 per cent in all cases, with a perfect score in the case of successful source reconstructions. i.e, only successful source reconstructions were labelled as such. These results are summarised in Table 3.2.

As a further test of the CNN's ability to accurately classify source reconstructions, we applied it to the more complex HUDF source reconstructions described in Section 3.2.3. Here, the CNN gave predictions on 100 each of under-magnified, over-magnified and accurately reconstructed sources. We found that our CNN correctly classified 87 per cent of the under-magnified reconstructions, whilst misclassifying them as

Table 3.2: Summary of key performance statistics for our CNN applied to testing data.

	Precision	Recall	F_1-score	Support
UM	0.9996	1.000	0.9998	2277
C	1.0000	0.9991	0.9996	2341
OM	0.9996	1.0000	0.9998	2310

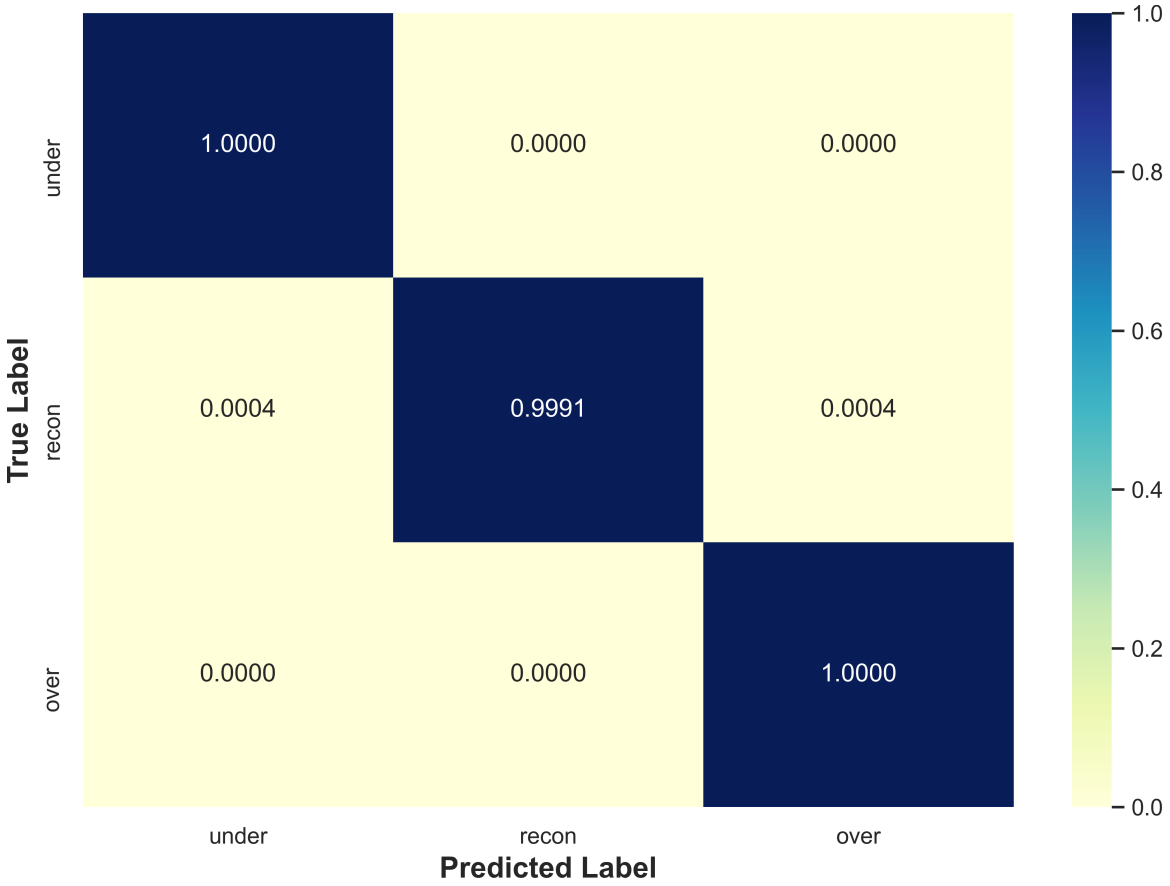


Figure 3.11: Confusion matrix for the CNN when tested on 6,928 never seen before pairs of source reconstructions and residual images for a simple Sérsic source. The confusion matrix has been normalised over its rows.

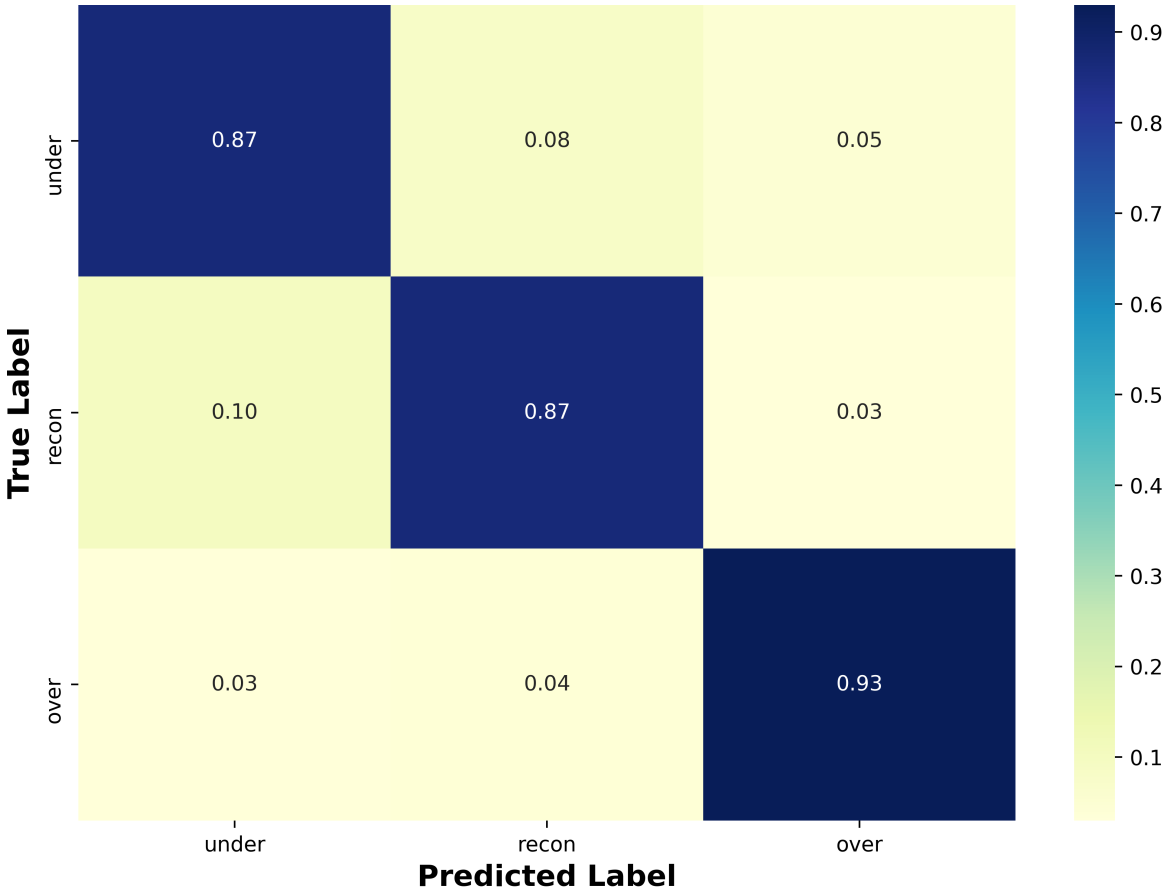


Figure 3.12: Confusion matrix for the CNN when trained on reconstructions of Sérsic sources and tested on reconstructions of HUDF sources. The confusion matrix has been normalised over its rows.

correctly reconstructed 8 per cent of the time, and incorrectly classifying 5 per cent of them as over-magnified. The CNN gave accurate predictions for 87 per cent of the correctly reconstructed sources, whilst incorrectly labelling 10 per cent as under-magnified and 3 per cent as over-magnified. Finally, the CNN correctly labelled 93 per cent of the over-magnified reconstructions, with just 3 per cent incorrectly labelled as under-magnified and 4 per cent mislabelled as accurate reconstructions. These results are summarised in Fig. 3.12. The performance of the CNN on this complex dataset is remarkably good, given the simplicity of the reconstructed sources in the training data.

3.3.2 Performance of `PyAutoLens` combined with CNN

Here, we describe the results of applying our CNN to blindly modelled data. For this, we have used our simulated images of Sérsic sources. We describe the process of using our CNN predictions to automatically adjust the prior distributions on the Einstein radius in three subsequent rounds of modelling.

The results of this are presented in Fig. 3.13. The initial modelling of this set of 100 lenses was carried out with no prior information on the lens model parameters and as such, under-magnified solutions have dominated the output. The bottom-right histogram in Fig. 3.13 shows how the proportion of different source reconstructions changes with each iteration of modelling according to our CNN predictions. Initially, our CNN identifies 88 models as UM, 11 as OM, and just 1 is identified as C. This is reflected in the error distributions for the key SIE mass model parameters. The top-left distribution in Fig. 3.13 shows the fractional error in Einstein radius f_θ for all 100 systems. There is a very significant peak in the initial data at $f_\theta = -0.45$, representing the large number of under-magnified solutions, and thus under-estimated Einstein

radii. We also see in the top-right distribution of Fig. 3.13 the significant bias towards under-estimating the axis ratio of the lens. The bottom-left distribution, showing the absolute error on the inferred orientation of the lens, reflects the seemingly random relationship between the erroneous models and the true lens orientation. Labelled as rerun 1, rerun 2, and rerun 3, we show that the application of our CNN and prior updating routine to these results leads to a huge improvement in recovering the true lens parameters for the sample. After rerun 1 has been completed, much of the bias in the Einstein radii fractional error distributions is removed, though there is still significant density in regions indicating under and overestimation of its value. Similarly, in the case of the axis ratio, a clear peak around $f_q = 0$ has been formed, removing much of the probability mass in the under-estimate region of before. The inference of the orientation of the lens has also been greatly improved, as we would expect by increasing the number of successfully modelled systems. These results are reflected in the bottom right histogram of Fig. 3.13, showing that the proportion of successful source reconstructions has increased from 1 to 52, according to our CNN predictions. The number of under-magnified reconstructions has been decreased by 68, down to just 20. The frequency of over-magnified solutions has increased, however, suggesting that our scheme for updating the Einstein radius prior has 'overshot' the correct solution in some cases. Rerun 2 increases the number of successful reconstructions by a small margin, but mostly results in moving solutions from the over-magnified category into the under-magnified category. Significant improvements are made in the final round of modelling, rerun 3, by considering the history of models for each case. For a system that has models that have previously been classified as under-magnified and over-magnified, we can search parameter space between the inferred Einstein radii values and hopefully converge upon the correct solution. In all the error distributions for the mass-model parameters, we see improvements, i.e. taller, narrower peaks centred on

zero. After the final round of modelling, we achieved a decrease of 69 per cent in the occurrence overall of unphysical source reconstructions. The final count of successful source reconstructions stands at 70, with 17 under-magnified and 13 over-magnified solutions. In principle, we could continue with this process until we no longer see any improvement in the number of successful source reconstructions being identified by the CNN, or all systems that have not been labelled as C become flagged for manual inspection.

3.4 Conclusions

Strong gravitational lensing allows us to probe the mass distributions of the lensing galaxy as well as the properties of the background sources. Upcoming surveys such as LSST and Euclid are expected to observe in excess of one hundred thousand strong gravitational lenses. To deal with this huge amount of data, it is necessary to develop fast, robust and automatic lens modelling pipelines that do not require significant time investment from humans for each system. For this reason, we constructed a CNN to detect when the modelling process has gone awry and developed a simple scheme for automatically adjusting the prior distribution on the Einstein radius to guide the sampler to the correct solution. Simulated images with the resolution and expected seeing characteristics of the Euclid VIS instrument were created, to be used as inputs for the production of source reconstructions. We chose to simulate all of our lenses as SIEs and used Sérsic profiles for our sources. In both cases, we used realistic distributions of parameters that matched those observed in the SLACS survey. From these simulated images, we produced three source reconstructions for each observation corresponding to the under/over magnified solution and the correct solution. These source reconstructions, along with the residual images for the model were used to

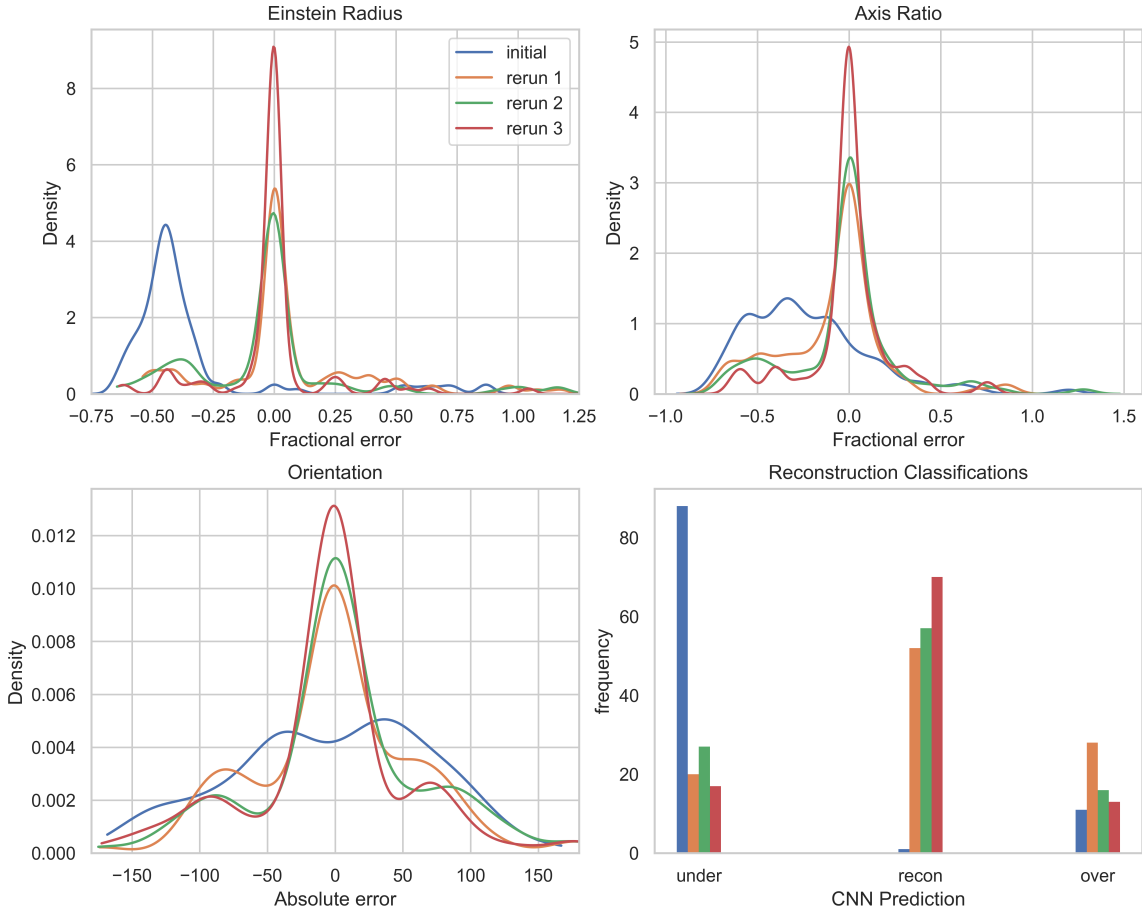


Figure 3.13: Top row: Fractional error distributions for the SIE mass model parameters, θ_E and q after successive rounds of modelling. Bottom left: Absolute error distribution for the SIE mass model parameter ϕ after successive rounds of modelling. Bottom right: Histogram of the CNN's predictions for how the fraction of successful source reconstructions changes with successive rounds of modelling.

train a CNN to classify source reconstructions. We then blindly modelled 100 strong lenses, reconstructing the background sources on a Voronoi grid. The CNN was used to detect the kind of source reconstruction that had been produced, and this information coupled with a simple scheme for updating the prior distribution on the Einstein radius was used to improve upon the fraction of successfully modelled systems.

We find that our CNN is capable of extremely accurate identification of under-magnified, successful and over-magnified reconstructed sources. The network achieves a precision and recall over 99.9 per cent, as well as an f_1-score, or harmonic mean of the precision and recall, greater than 0.99 across all classes of source reconstruction. In addition to identifying the class of solution that has been found, we have shown that a simple procedure for updating the model based on its predicted class can lead to significant improvements in the outcomes of blind modelling without the need for human intervention throughout the process.

The success of our CNN in this task suggests that our procedure for generating the source reconstructions, omitting the full exploration of parameter space, has not negatively impacted its ability to perform the task. The axis ratio of the SIE mass-model corresponding to an erroneous solution tends to be under-estimated. Our network is trained on source reconstructions produced by fixing the axis ratio to its true value. This leads to the network being trained on images produced by less elliptical lens models than it might encounter when being tested upon a freely varied model.

It is possible that incorporating the information regarding erroneous source reconstructions tendency to have an under-estimated lens axis ratio could lead to improvements in our procedure for updating the model priors. An approach that uses a Gaussian prior to bias towards higher values of q, but with a standard deviation large enough to easily allow the exploration of the lower end of parameter space is something that could be investigated.

We have also tested our CNN, trained on reconstructions simple Sérsic sources, on reconstructions of images generated using real sources extracted from the HUDF. The CNN continued to perform well, showing that it can generalise to a more complex dataset without any retraining. There is however an obvious detriment to the performance of the network, and so the construction of a more complex training set would likely be beneficial. Before this technique can be applied to real data, further investigations into how our simplifications affect the network's performance are needed. One such simplification that we made was to omit lens light from our simulated images. Even in the best possible scenario of lens light removal, its presence will affect the noise characteristics of the image, which can impact the source reconstruction. Realistic features in our simulated images such as cosmic rays and hot pixels were not considered. Increased complexity of the sources in our training data would be required to deal with the variety of real images that might be observed and to minimise the performance decrease due to an overly simplified training set. Furthermore, a wider variety of mass models, the inclusion of substructure and deflectors along the line of sight would be required to create more realistic lensing scenarios. The question of how well this method of applying CNN predictions to parametric models generalises to real data requires further investigation.

Chapter 4

Modelling high-resolution ALMA observations of strongly lensed dusty star forming galaxies detected by Herschel

As already discussed in Chapter 1, SMGs play a key role in our overall picture of galaxy evolution. Similarities in their size, number density and clustering properties with quiescent galaxies at low redshifts (e.g., Simpson et al., 2014; Toft et al., 2014) is suggestive of an evolutionary connection (An et al., 2019; Dudzevičiūtė et al., 2020). An explanation for how the population of SMGs became the red and dead population of massive elliptical galaxies observed in the local Universe is underpinned by the process of gas quenching and the mechanism of gas-poor mergers (Oogi & Habe, 2012; Guo & White, 2008; Lapi et al., 2018). Although considerably less abundant, Ultra Luminous Infrared Galaxies (ULIRGs) found in the local Universe are often seen as analogues to high-redshift SMGs due to their strongly dust-obscured UV luminosities,

high infrared luminosities, and comparable bolometric luminosities (Alaghband-Zadeh et al., 2012; Rowlands et al., 2014), and as such ULIRGs can provide useful insight into the processes at play in SMGs.

The study of high-redshift sub-mm galaxies has benefited greatly from the advent of large interferometric arrays such as the Atacama Large Millimeter/Sub-millimeter Array (ALMA) (Wootten & Thompson, 2009), which allow observations to reach resolutions of < 0.1 arcsec and thus probe previously unreachable physical scales. Strong gravitational lensing provides a further boost in spatial resolution due to the magnification of the background source, which is typically within the range of 5-10 for SMGs (Spilker et al., 2016; Bussmann et al., 2013; Bussmann et al., 2015). In addition, there exists a strong lensing bias in the sub-mm regime (see Chapter 1 for a more detailed explanation). Using this technique, follow-up ALMA observations of strongly lensed sub-mm galaxies detected in wide area surveys, such as the Herschel Astrophysical Terahertz Large Area Survey (H-ATLAS; Eales et al., 2010), have contributed to a rapidly increasing understanding of galaxy formation in its early stages (Dye et al., 2018; Sun et al., 2021; Harrington et al., 2021; Cañameras et al., 2017a,b). Most existing observations have mainly targeted extremely luminous sources, but with the increased sensitivity of instruments such as ALMA, it has become possible to investigate more typical main-sequence star-forming galaxies which are responsible for the bulk of the Universe's star formation at $z \sim 3$.

High-resolution sub-mm follow-up observations of lensed SMGs not only allow for more precise determination of the lens mass distribution, but also allow intrinsic properties of the source – such as its luminosity, SFR, and gas and dust mass – to be better characterised. Indeed, investigating properties such as these will be the main motivation of this chapter. A classic example of this approach is the lensed system SDP.81, first discovered within the H-ATLAS sample and then observed with ALMA

(ALMA Partnership et al., 2015; Dye et al., 2015; Rybak et al., 2015a; Rybak et al., 2015b; Negrello et al., 2010). Studies of the molecular gas and dust in lensed sub-mm galaxies have been used to constrain models of star formation in the early Universe (Dye et al., 2022a; Cava et al., 2018; Dessauges-Zavadsky et al., 2019). Recent studies have begun to reveal the compact nature of dust in SMGs (Puglisi et al., 2019; Tadaki et al., 2020) which indicate a disparity in size when compared with local ULIRGs.

Reconstruction of the background lensed source from interferometer observations can be achieved with two distinct approaches. There are methods that model the visibilities directly in the uv-plane (Bussmann et al., 2012; Bussmann et al., 2013; Dye et al., 2018), and those that model the cleaned image plane data (Dye et al., 2015; Inoue et al., 2016; Yang et al., 2019) (see Chapter 1 for a discussion of lens modelling in both the image and uv plane, along with a description of how images are formed from interferometer datasets). The benefit of working in the image plane is that the task is significantly less computationally expensive; however, due to the incomplete coverage of the uv-plane, spatially correlated noise is introduced which can in principle bias the inferred lens model, as we will discuss in Chapter 5. Working directly with the visibility data avoids this problem, at the cost of longer modelling times.

In this work, we carry out lens modelling in the uv-plane of seven galaxies observed with ALMA. Four of these systems are galaxy-galaxy scale, and the remaining three are group-scale lenses. These systems were originally detected by Herschel within H-ATLAS and the extension to the HerMES field, HerMES Large Mode Survey (HeLMS) (Asboth et al., 2016; Nayyeri et al., 2016). We investigate the intrinsic source properties, namely, their luminosities, SFRs, and gas and dust masses by determining the magnification factors. Additionally, we investigate the morphologies of the reconstructed sources.

The layout of this chapter is as follows: Section 4.1 describes the ALMA observations and other sources that were drawn upon for this work. Section 4.2 details the methodology of our lens modelling, and Section 4.3 presents the results of our work. In Section 4.4 we compare our results to other similar studies. Finally, in Section 4.5 we summarise our findings and discuss their interpretation. Throughout this chapter, we assume a flat ΛCDM cosmology using the 2015 Planck results (Planck Collaboration et al., 2016), with Hubble constant $H_0 = 67.7\,\mathrm{kms^{-1}Mpc^{-1}}$ and matter density parameter $\Omega_m = 0.307$.

4.1 Data

The seven ALMA observations modelled in this work are from the ALMA programme 2013.1.00358.S (PI: Stephen Eales) and are described in detail within Amvrosiadis et al. (2018). The observation targets for the original ALMA programme were selected from the H-ATLAS and HeLMS surveys for having the brightest 500 μm flux densities ($> \sim$130 mJy) and with spectroscopic redshifts > 1, with the exception of HeLMS J235331.9+031718. Originally, 42 sources were identified in this way, but only 16 had follow-up ALMA observations. Of the 16 sources observed during ALMA cycle 2, 14 exhibit obvious lensing features. Six of these remaining 14 sources have already been modelled in the work of Dye et al. (2018), and are all galaxy-galaxy scale lenses; one of them we leave for future work, and the remaining seven are modelled in this work. Due to the timings of when the ALMA observations were carried out, it was natural to split the sample, leading to all the group scale lenses being dealt with in this work. Of these seven sources, one was identified in H-ATLAS, whilst the remaining six are from HeLMS.

The spectral setup employed by ALMA was identical for each of the lensing systems

Table 4.1: A summary of the targeted position of each source, the number of antennae used in each observation, the continuum sensitivity, the size of the beam, and the physical scale probed at the source redshift.

Short name	Targeted position (RA DEC)	$N_{\rm ant}$	Sensitivity (mJy/beam)	$\theta_{\rm maj} \times \theta_{\rm min}$	Physical scale (kpc)
G09v1.97	08:30:51.040 +01:32:25.000	35	0.1044	$0.229'' \times 0.154''$	1.33
HELMS18	00:51:59.450 +06:22:40.400	42	0.1007	$0.187'' \times 0.158''$	1.48
HELMS5	23:40:51.520 -04:19:38.400	46	0.1034	$0.229'' \times 0.151''$	1.37
HELMS7	23:24:39.420 -04:39:34.500	46	0.1034	$0.218'' \times 0.151''$	1.51
HELMS2	23:32:55.500 -03:11:36.400	46	0.1034	$0.227'' \times 0.151''$	1.48
HELMS15	23:32:55.510 -05:34:26.500	46	0.1034	$0.223'' \times 0.151''$	1.52
HELMS40	23:53:31.990 +03:17:18.400	42	0.1147	$0.260'' \times 0.150''$	1.55 – 1.33

observed. The band 7 continuum observations were comprised of four spectral windows, each with a width of 1875 MHz and centred on the frequencies 336.5 GHz, 338.5 GHz, 348.5 GHz and 350.5 GHz. The central frequency of 343.404 GHz corresponds to a wavelength of 873 μm. Each spectral window is comprised of 128 frequency channels, resulting in a spectral resolution of 15.6 MHz. The ALMA configuration utilised between thirty-five and forty-two 12 m antennae, with an on-source integration time of approximately 125 s. Upon combining all four spectral windows, this achieves an angular resolution of 0.12 arcsec and Root Mean Square (RMS) values of approximately 230 μJy/beam and 130 μJy/beam for the H-ATLAS and HeLMS sources, respectively. This setup was chosen to optimise band 7 continuum observations, as well as to achieve the angular resolutions necessary to probe the interstellar medium on the scales of giant molecular clouds with high signal-to-noise. The synthesised beam shape for HeLMS J005159.4+062240 is the least elliptical of our sample, with a ratio of major to minor axis of $\sim$1.2. For the observation of HeLMS J235331.9+031718, the ratio of major to minor axis of the synthesised beam is the most elliptical of our sample at $\sim$1.7, with the remaining observations having ratios of $\sim$1.5. The maximum recoverable scales for our observations are between 1.32 arcsec and 1.46 arcsec. The details of the number of antennae, achieved sensitivity, and beam sizes are summarised in Table 4.1.

In this work, we used the visibility data provided by the ALMA science archive, and re-calibrated them using Common Astronomy Software Applications version (`CASA`) 4.3.1 (McMullin et al., 2007), and the scripts provided by the archive. Baselines flagged as bad by the ALMA data reduction pipeline were excluded from the analysis. The `CASA` task `tclean` was used to create images in order to measure the flux density of the sources at 873 μm. The images were constructed using a natural weighting scheme and were primary beam corrected. In order to ensure the minor axis of the

primary beam was well-sampled, image pixel scales of 0.02 arcsec and 0.03 arcsec were used for the H-ATLAS and HeLMS sources respectively. For the creation of these images, we used `CASA` version 6.1.2.7. The data reduction follows closely the method of Dye et al. (2018), with the exception of using natural weighting in favour of Briggs weighting when creating images with `tclean`.

For the calculation of intrinsic source properties, photometry from our ALMA data was used in combination with a number of other datasets. Sub-mm photometry, obtained by the Herschel Space Observatory, was employed. This photometry made use of two different instruments: the Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al., 2010) at wavelengths of 250 μm, 350 μm, and 500 μm; and the Photoconductor Array Camera and Spectrometer (PACS; Poglitsch et al., 2010) at wavelengths of 100 μm and 160 μm. SPIRE and PACS photometry for the H-ATLAS and HeLMS sources was taken from Zhang et al. (2018) and Nayyeri et al. (2016). Where possible, we have also used 850 μm Sub-millimeter Common User Bolometer Array 2 (SCUBA-2) flux densities (14 arcsec resolution; Mackenzie et al., 2011) taken from Bakx et al. (2017), 880 μm photometry taken from the Sub-millimeter Array (SMA) (0.6 arcsec resolution) as described in Bussmann et al. (2013) and ALMA/ACA rest-frame 290 μm and 310 μm photometry taken from Riechers et al. (2021). Finally, the ALMA band 4 (1940 μm, 0.3 arcsec resolution) flux density for H-ATLAS J083051.0+01322 was taken from Yang et al. (2019).

The available lens and source redshifts for the seven systems modelled in this work can be found in Table 4.2. The observed photometry can be found in Table 4.3.

Table 4.2: The list of the seven lensing systems modelled in this work, along with their lens galaxy redshifts, $z_{\rm l}$, and their background source redshifts $z_{\rm s}$. Where appropriate, the redshift of both lensing galaxies have been provided, distinguished by the numbered subscript. The references from which the lens and source redshifts were obtained are as follows: [a]Bussmann et al. (2013); [b]Negrello et al. (2017); [c]Nayyeri et al. (2016); [d]Okido et al. (2020); and [e]Riechers et al. (2021). The range of redshifts given for HeLMS J235331.9+031718 is an assumption based on the range of source redshifts in this thesis, since there is no available redshift measurement for this source. A dash indicates missing redshift information for a lens, whilst 'N/A' is used to indicate a second mass profile was not used in our modelling procedure. A list of short names is provided that may be used throughout this thesis; these names are consistent with Nayyeri et al. (2016) and Bussmann et al. (2013). A list of alternate names that have been used to describe these sources (e.g. in Riechers et al., 2021) is also provided to emphasise that these are indeed the same sources.

ID	Short name	Alternative name	$z_{\rm l,1}$	$z_{\rm l,2}$	$z_{\rm s}$
H-ATLAS J083051.0+013225	G09v1.97	-	0.626[b]	1.002[b]	3.634[a]
HeLMS J005159.4+062240	HELMS18	HeLMS-9	0.602[d]	0.599[d]	2.3934[e]
HeLMS J234051.5-041938	HELMS5	HeLMS-62	-	N/A	3.5027[e]
HeLMS J232439.5-043935	HELMS7	HeLMS-45	-	-	2.4726[e]
HeLMS J233255.4-031134	HELMS2	HeLMS-44	0.426[c]	N/A	2.6895[e]
HeLMS J233255.6-053426	HELMS15	HeLMS-61	0.976[c]	N/A	2.4022[e]
HeLMS J235331.9+031718	HELMS40	-	0.821[c]	N/A	2.0 – 3.7

Table 4.3: Observed (i.e. lensed) source flux densities in mJy. The passband central wavelength in μm is indicated by the subscripts. For both the H-ATLAS and HeLMS sources, the flux densities f_{100} to f_{500} are taken from Zhang et al. (2018) and Nayyeri et al. (2016). The flux densities f_{850}, f_{873}, f_{880} and f_{1940} are taken from Bakx et al. (2017), this work, Bussmann et al. (2013) and Yang et al. (2019), respectively. The flux densities $f_{290}^{\rm rest}$ and $f_{310}^{\rm rest}$ are rest frame quantities and are taken from Riechers et al. (2021)

ID	f_{100}	f_{160}	f_{250}	f_{350}	f_{500}	f_{850}	f_{873}	f_{880}	f_{1940}	$f_{290}^{\rm rest}$	$f_{310}^{\rm rest}$
G09v1.97	53 ± 3	198 ± 10	260 ± 7	321 ± 8	269 ± 9	121 ± 9	92 ± 9	86 ± 4	8.8 ± 0.5	-	-
HELMS18	31 ± 3	91 ± 15	166 ± 6	195 ± 6	135 ± 7	-	41 ± 4	-	-	29.6 ± 1.7	-
HELMS5	7 ± 3	68 ± 7	151 ± 6	209 ± 6	205 ± 8	-	94 ± 9	-	-	32.15 ± 0.43	21.79 ± 0.90
HELMS7	33 ± 4	129 ± 7	214 ± 7	218 ± 7	172 ± 9	-	36 ± 4	-	-	22.3 ± 1.2	-
HELMS2	25 ± 4	146 ± 14	271 ± 6	336 ± 6	263 ± 8	-	75 ± 8	-	-	39.3 ± 1.1	-
HELMS15	14 ± 3	44 ± 8	148 ± 6	187 ± 6	147 ± 9	-	48 ± 5	-	-	32.89 ± 0.78	-
HELMS40	-	-	102 ± 6	123 ± 7	111 ± 7	-	26 ± 3	-	-	-	-

4.2 Methodology

4.2.1 The semi-linear inversion method in the uv-plane

The modelling of the systems in this work was carried out in the uv-plane, and as such, we used the source inversion method implemented within `PyAutoLens` (Nightingale et al., 2021), which is based on the operator approach described within Powell et al. (2020). For a description of this methodology, see Section 1.7.3. We first reconstructed the background source on a Voronoi grid adapted to the lens magnification, before using this to initialise a fit that reconstructed the source on a grid adapted to the source light (Nightingale & Dye, 2015; Nightingale et al., 2018). A gradient regularisation scheme, analogous to that which is described in Warren & Dye (2003), with a constant weight for the source plane, was used throughout. The nested sampling algorithm DYNESTY (Speagle, 2020) was used to find the optimal lens model parameters that maximised the Bayesian evidence (Suyu et al., 2006).

4.2.2 Lens Model

We have used the elliptical power-law density profile, which is a generalised form of the singular isothermal ellipsoid that is commonly used to fit strong lens profiles (Keeton, 2001). When it improves the Bayesian evidence, an external shear component is included to compensate for the influence of line of sight galaxies that may be outside our field of view. Where indicated by additional imaging or unusual image configurations, two elliptical power-law profiles have been used to model the group-scale lenses present in the sample. We find that in all cases this is sufficient to provide acceptable fits to the data and, through the use of optical/near-infrared imaging, we see no indication of more than two lenses being required to produce an accurate model. The

surface mass density, κ, of this profile is given by:

$$\kappa(R) = \frac{3-\alpha}{1+q}\left(\frac{\theta_{\rm E}}{R}\right)^{\alpha-1}, \tag{4.1}$$

where $\theta_{\rm E}$ is the model Einstein radius in arc seconds, α the power-law index and R the elliptical radius defined as $R = \sqrt{x^2 + y^2/q^2}$, where q is the axis ratio (Suyu, 2012). The lens orientation ϕ is measured as the angle counterclockwise from the east axis to the semi-major axis of the elliptical lens profile. The centre of the lens profile is given by the coordinates in the image-plane $(x_{\rm c}, y_{\rm c})$. The external shear field is parameterised by a shear strength γ and shear direction ϕ_γ, measured counterclockwise from east. The shear direction is defined to be perpendicular to the direction of the resulting image stretch. For single lens systems, there are six lens model parameters (eight when including external shear) and 12 lens model parameters (14 when including external shear) in the case of group-scale lenses.

4.3 Results

Fig 4.1 shows the model image, residual image and source reconstruction for each of the seven lenses modelled in this work. The lens model parameters from our fitting procedure are given in Table 4.4.

Different interferometer configurations probe distinct scales and reach varying surface brightness limits. Fig 4.3 shows how, for each system, the inferred magnification is sensitive to this effect. By working down a list of a subset of source-plane pixels ordered by their flux density (and chosen to have a surface brightness above a certain threshold), the lensed image and average magnification of these pixels was computed. The source flux density fraction tells us how much of the total flux density is con-

Table 4.4: Lens model parameters. The columns are the Einstein radius, $\theta_{\rm E}$, coordinates $(x_{\rm c}, y_{\rm c})$ of the centroid of the mass profile with respect to the phase-tracking centre of the observations ($x_{\rm c}$ and $y_{\rm c}$ correspond to west and north respectively), the lens profile orientation measured as the angle counterclockwise from east to the semi major axis, ϕ, the lens profile axis ratio, q, the density slope of the power-law, α, the magnitude of the external shear field, γ, and the external shear field direction, ϕ_γ, measured counterclockwise from east.

ID	$\theta_{\rm E}$ (arcsec)	$(x_{\rm c}, y_{\rm c})$ (arcsec)	ϕ (deg)	q	α	γ	ϕ_γ (deg)
G09v1.97 (lens 1)	0.43 ± 0.01	(-1.87 ± 0.01, -0.78 ± 0.01)	7 ± 1	0.51 ± 0.01	2.18 ± 0.02	-	-
G09v1.97 (lens 2)	0.54 ± 0.01	(-1.51 ± 0.01, -0.08 ± 0.01)	9 ± 2	0.90 ± 0.01	2.03 ± 0.08	-	-
HELMS18 (lens 1)	3.80 ± 0.02	(-4.67 ± 0.01, 4.02 ± 0.01)	156 ± 2	0.37 ± 0.05	2.05 ± 0.03	0.17 ± 0.02	102 ± 3
HELMS18 (lens 2)	1.46 ± 0.02	(-0.63 ± 0.01, 2.18 ± 0.01)	109 ± 5	0.66 ± 0.06	1.93 ± 0.02	-	-
HELMS5	0.75 ± 0.01	(-0.78 ± 0.01, -1.88 ± 0.01)	97 ± 1	0.27 ± 0.01	2.37 ± 0.01	0.17 ± 0.01	103 ±1
HELMS7 (lens 1)	0.54 ± 0.01	(1.91 ± 0.02, 1.09 ± 0.02)	23 ± 1	0.79 ± 0.02	2.08 ± 0.06	0.13 ± 0.01	165 ± 2
HELMS7 (lens 2)	0.40 ± 0.01	(0.88 ± 0.01, 1.24 ± 0.01)	84 ± 2	0.27 ± 0.02	1.84 ± 0.03	-	-
HELMS2	1.01 ± 0.01	(-0.06 ± 0.01, -1.52 ± 0.02)	118 ± 2	0.44 ± 0.02	1.89 ± 0.01	0.18 ± 0.01	112 ± 1
HELMS15	1.05 ± 0.01	(2.43 ± 0.01, 0.18 ± 0.01)	150 ± 1	0.56 ± 0.01	1.99 ± 0.03	0.12 ± 0.01	48 ± 1
HELMS40	0.21 ± 0.01	(1.81 ± 0.01, 1.32 ± 0.01)	151 ± 4	0.49 ± 0.05	1.64 ± 0.04	0.19 ± 0.02	170 ± 5

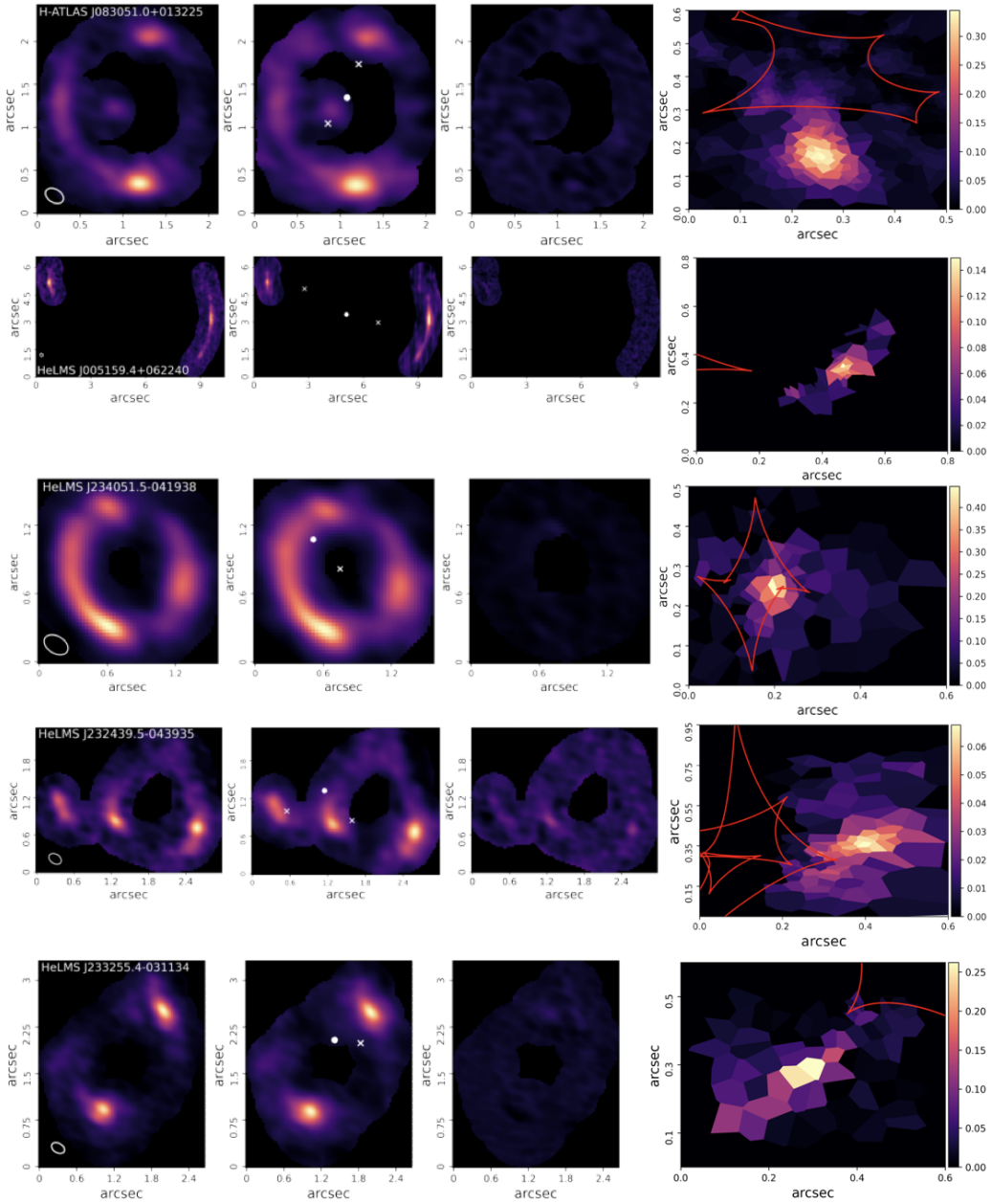


Figure 4.1: Lens reconstructions for each system. The left column shows the cleaned ALMA image. The middle left column shows the lensed image of our reconstructed source, i.e. the model image with the lens centroid(s) and centre of the source plane indicated by white crosses and white circles respectively. The middle right column shows the residual image for our model, i.e. the image of the observed visibilities minus the image of the model visibilities. The right column shows the reconstructed source. The colour bar indicates the 873μm surface brightness in Jy/arcsec^2 for each of the panels related to a particular source and all residuals are $< 3\sigma$. The caustics are shown in red. North is up and east is left. The beam is shown as a white ellipse in the bottom left corner of the cleaned ALMA images.

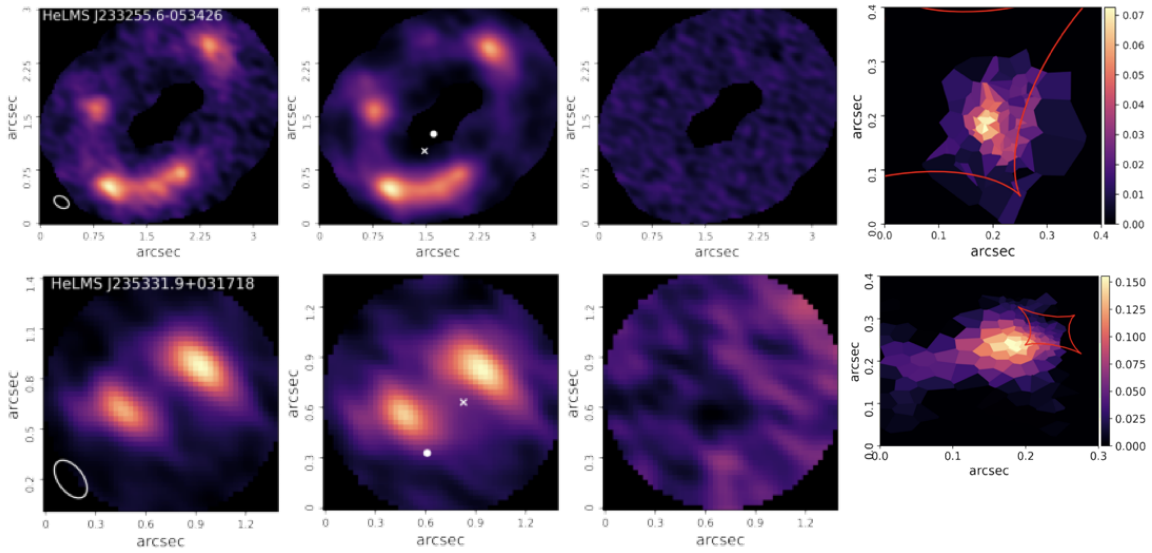


Figure 4.2: Lens reconstructions for each system. The left column shows the cleaned ALMA image. The middle left column shows the lensed image of our reconstructed source, i.e. the model image with the lens centroid(s) and centre of the source plane indicated by white crosses and white circles respectively. The middle right column shows the residual image for our model, i.e. the image of the observed visibilities minus the image of the model visibilities. The right column shows the reconstructed source. The colour bar indicates the 873μm surface brightness in Jy/arcsec^2 for each of the panels related to a particular source and all residuals are $< 3\sigma$. The caustics are shown in red. North is up and east is left. The beam is shown as a white ellipse in the bottom left corner of the cleaned ALMA images.

tained within the subset of source-plane pixels as we move down the list. Likewise, the image flux density fraction tells us how much of the total image flux density is contributed by the lensed image of our subset of source-plane pixels. This process was repeated 100 times using a randomised source plane pixelisation for each, to produce an averaged magnification profile.

4.3.1 Intrinsic source properties

For each lens system, we have determined the intrinsic properties of the background source. To achieve this, we have demagnified the sub-mm photometry (see Table 4.3) by the total source magnification factors given in Table 4.5, taking into account the uncertainties on our magnification values. With the source redshifts given in Table 4.2, we fitted the rest-frame photometry with two SEDs. Firstly, a single temperature, optically thick SED was fitted, of the form:

$$S_\nu \propto [1 - \exp\left(-(\nu/\nu_0)^\beta\right)]B(\nu, T_\mathrm{d}), \tag{4.2}$$

where S_ν is the flux density at frequency ν, ν_0 is the frequency at which the optical depth is equal to unity, β is the dust emissivity index, T_d is the dust temperature and $B(\nu, T_\mathrm{d})$ denotes the Planck function. Secondly, a dual temperature, optically thin SED was fitted, of the form:

$$S_\nu = \nu^\beta[N_\mathrm{c}B(\nu, T_\mathrm{c}) + N_\mathrm{w}B(\nu, T_\mathrm{w})], \tag{4.3}$$

where N is the weighting of the cold and warm components (denoted with subscripts) and T is the dust temperature of the two components (denoted by the subscripts). This allows for the computation of an upper and lower bound in the range of possible

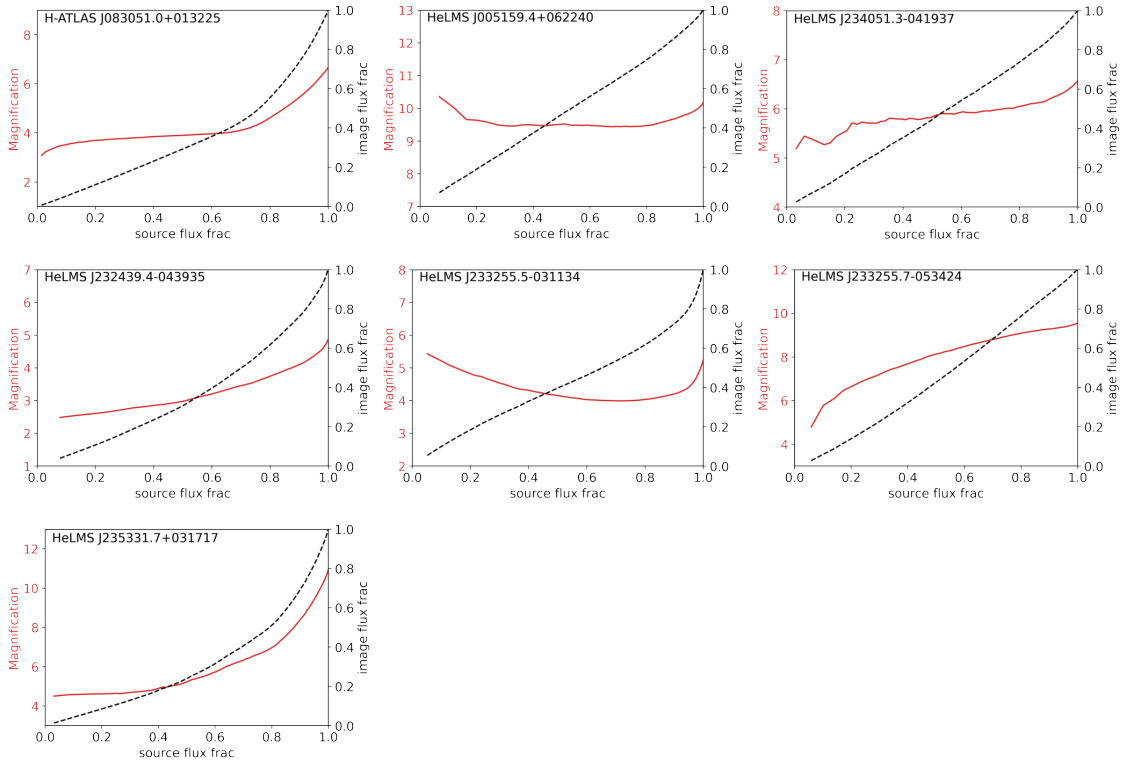


Figure 4.3: The magnification profile for each of the reconstructed background sources. Each plot shows how the total magnification (solid red line) and image flux density fraction (dashed black line) varies as a function of the source flux density fraction above a surface brightness threshold (more details given in the main text). The magnification profiles represent the average of 100 realisations of the source plane pixelisation for the best-fit lens model. The plot demonstrates how the inferred magnification can vary with different interferometer configurations.

Table 4.5: Intrinsic source properties. The columns are total magnification, $\mu_{\rm tot}$, dust mass computed from the single temperature optically thick SED, $M_{\rm d}^{\rm thick}$, dust mass assuming a dual temperature optically thin SED, $M_{\rm d}^{\rm thin}$, temperature of the optically thick SED, $T^{\rm thick}$, temperatures of the optically thin SED, $T^{\rm thin}$, the optical depth at 100 μm for the optically thick SED, τ_{100}, the demagnified luminosity (computed as the integral of the optically thin SED from 3 to 1100 μm), $L_{\rm IR}$, the H_2 gas mass computed using the scaling relation of Hughes et al. (2017), $M_{\rm gas}$, and the SFR scaled from $L_{\rm IR}$ using the procedure given by Kennicutt & Evans (2012) with a Kroupa IMF. Dust masses are expressed as $\log(M_{\rm d}/{\rm M}_\odot)$, luminosities as $\log(L_{\rm IR}/{\rm L}_\odot)$, and gas masses as $\log(M_{\rm gas}/{\rm M}_\odot)$. The source HeLMS J235331.9+031718 appears twice, displaying the intrinsic source properties for the upper and lower value of source redshift being considered (indicated in the ID column).

ID	$\mu_{\rm tot}$	$M_{\rm d}^{\rm thick}$	$M_{\rm d}^{\rm thin}$	$T^{\rm thick}$ (K)	$T^{\rm thin}$ (K)	τ_{100}	$L_{\rm IR}$	$M_{\rm gas}$	SFR (M$_\odot$/yr)
G09v1.97	6.7 ± 0.5	8.5	9.2	70	32 / 70	3.1	13.4 ± 0.1	11.4 ± 0.1	3700 ± 500
HELMS18	10.2 ± 0.5	8.2	8.9	62	28 / 59	6.9	12.6 ± 0.1	11.1 ± 0.1	590 ± 50
HELMS5	6.7 ± 0.5	8.3	9.1	62	26 / 53	7.8	13.1 ± 0.1	11.4 ± 0.1	1900 ± 300
HELMS7	4.9 ± 0.5	8.5	9.1	63	30 / 55	4.2	13.0 ± 0.1	11.3 ± 0.1	1500 ± 300
HELMS2	5.5 ± 0.5	8.5	9.1	60	28 / 51	4.9	13.1 ± 0.1	11.5 ± 0.1	1900 ± 300
HELMS15	9.2 ± 0.5	8.5	9.0	48	28 / 63	4.9	12.5 ± 0.1	11.2 ± 0.1	500 ± 50
HELMS40 (z=2.0)	8.7 ± 0.5	8.7	9.0	37	25 / 31	3.3	12.1 ± 0.1	11.1 ± 0.1	180 ± 50
HELMS40 (z=3.7)	8.7 ± 0.5	8.2	8.5	59	20 / 40	1.4	12.7 ± 0.1	10.8 ± 0.1	800 ± 50

dust masses, which were determined using the method described in Dunne et al. (2011). In this work, we have used the 873 μm ALMA flux density and computed the dust mass absorption coefficient by extrapolating the 850 μm value of $\kappa_{850} = 0.077\ \mathrm{m^2 kg^{-1}}$ (James et al., 2002 – see Dunne et al., 2000 for more details). The uncertainties in the dust mass absorption coefficient are known to be large, but a relative comparison between dust masses computed with this value is still valuable. The same value of κ_{873} was used for all of our galaxies, which assumes that the physical properties of the dust, such as grain size and density, are constant for our sample of galaxies. The scaling relations from Hughes et al. (2017) were used to calculate the H_2 gas mass.

During the fitting of the optically thin SED, the temperature and normalisation of both dust components were allowed to vary. In the case of the optically thick SED, the temperature, normalisation, and opacity at 100 μm, τ_{100}, were varied during the fit. Throughout, the emissivity index, β, was fixed to 2.0 (see Smith et al., 2013). The fitted SEDs and the demagnified source photometry can be seen in Fig 4.4. The best fit parameters for these SEDs can be found in Table 4.5, along with the demagnified luminosity of the sources, computed by integrating the optically thin SED from 3-1100 μm. Finally, the SFRs of the sources are given, computed using the conversion from luminosity

$$\log(\mathrm{SFR}) = \log(\mathrm{L_{IR}}) - 43.41 \tag{4.4}$$

given by Kennicutt & Evans (2012), who employ the Initial Mass Function (IMF) of Kroupa (2001). This conversion assumes that all the infrared emission originates from star forming regions, which may result in biased SFR values if an active galactic nucleus (AGN) is significantly contributing to the luminosity of the galaxy.

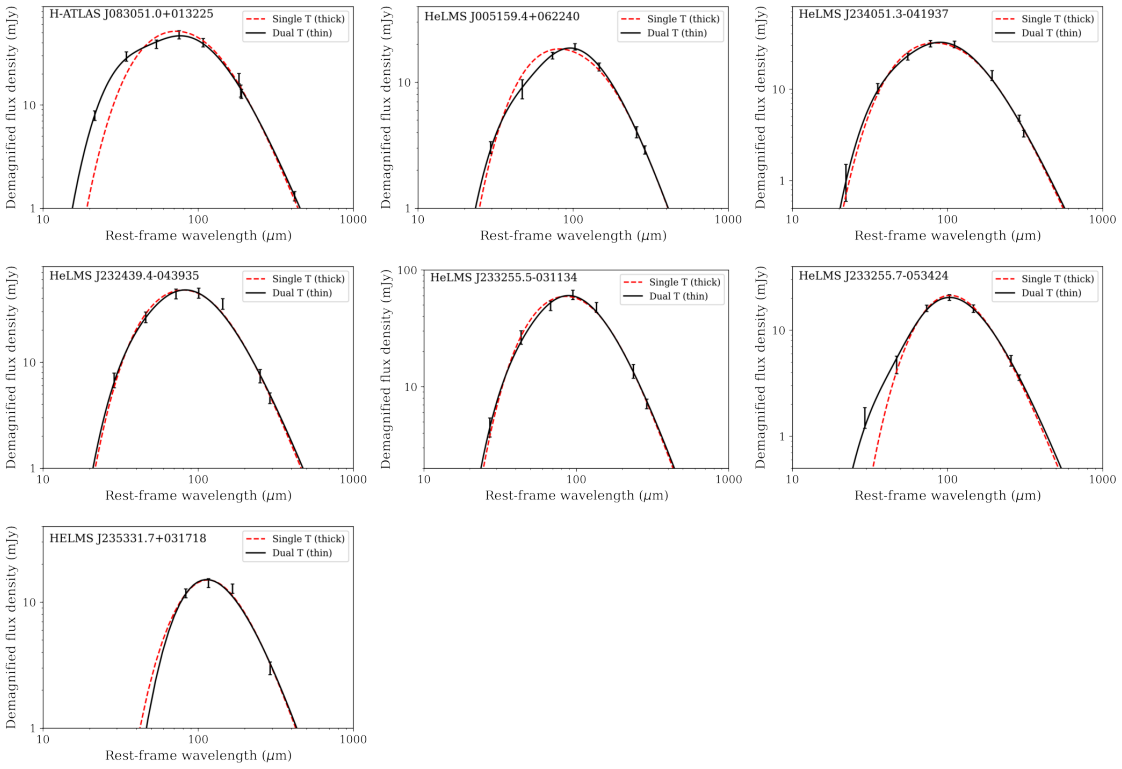


Figure 4.4: SEDs of the lensed background sources. The two-temperature optically thin fit (solid black line) and the single-temperature optically thick fit (dashed red line) are shown in each subplot. The measured photometry and uncertainties, shown as the error bars in the plots, have been demagnified by the appropriate lensing magnifications μ, given in Table 4.5. The SED for HeLMS J235331.7+031718 is shown at a redshift of $z = 2$.

4.3.2 Object notes

H-ATLAS J083051.0+013225 - This lens system has an almost complete ∼1.5 arcsec Einstein ring with three major image components, along with a fainter central image. The central image is due to the rare line of sight configuration of the dual deflectors. Keck Adaptive Optics K_s band imaging of this lensing system (see Calanog et al., 2014) shows two galaxies, but the lensing features are ambiguous due to the low signal-to-noise ratio. Upon superimposing the lensed emission from the background source as detected in ALMA data, it is clear that both of these galaxies are interior to the Einstein ring (see Yang et al., 2019). Long slit spectroscopy, presented in Bussmann et al. (2013), provides evidence that these two galaxies are at different redshifts (0.626 and 1.002 respectively). Bussmann et al. (2013) present the modelling of SMA data for this system, fitting two SIE models for the foreground deflectors and a Sérsic model for the background source. They infer a magnification factor $\mu = 6.9 \pm 0.6$, which is in agreement with the magnification we find with our best fit model of $\mu = 6.7 \pm 0.5$. Yang et al. (2019) present ALMA band 4 data, along with the band 7 data presented in this thesis, and find a significantly higher magnification factor of $\mu = 10.5^{+0.5}_{-0.6}$, using a double SIE lens model and a dual disk model for the background source.

The SFR of this lensed source is extremely high at $3700 \pm 500 M_\odot yr^{-1}$, in reasonable agreement with the value reported by Zhang et al. (2018), who used the magnification factor from Bussmann et al. (2013) to determine the SFR. This compares to the lower values of SFR inferred by Yang et al. (2019) who find $600 \pm 300\ M_\odot yr^{-1}$ and $900 \pm 400\ M_\odot yr^{-1}$ for each of the components in their source model. Our value of SFR is reduced to $2400\ M_\odot yr^{-1}$ when using the magnification factor of 10.5 found by Yang et al. (2019). A possible explanation for the large discrepancy in SFR is differential magnification of the source. Yang et al. (2019) find evidence of differential magnification for the compact and extended components of their source model, which

they have taken into account in their SED modelling. In addition to this, whilst the lens model that we find is similar to that of Yang et al. (2019), we have allowed the power-law slope to vary rather than using SIE profiles. The higher values (>2) of slope that we find will tend to decrease the overall magnification of the source plane.

Our reconstructed source for this system shows significant disturbance, with a main component oriented north-south and a fainter component oriented east-west.

HeLMS J005159.4+062240 - This system is a group-scale lens with a doubly-imaged background source. Optical imaging from the Sloan Digital Sky Survey Data Release 14 (Abolfathi et al., 2018) reveals two lensing galaxies within the lensed arcs. Tergolina (2020) investigated these objects, finding that one was a red and passive early-type galaxy (ETG) ($z = 0.60246 \pm 0.00004$), whilst the other is the host of a quasar ($z = 0.59945 \pm 0.00009$; Okido et al., 2020). We were able to find a lens model constructed from two elliptical power-law density profiles and an external shear field that reconstructs the background source as a single component. Our reconstructed source appears elongated with a disturbed morphology.

With a magnification factor of $\mu = 10.2 \pm 0.5$, this system has the highest total magnification in our sample. Our measurement of the intrinsic source luminosity, $\log(L_{\rm IR}/{\rm L}_\odot) = 12.6 \pm 0.1$, is in reasonable agreement with the far-infrared (far-IR) luminosity given in Nayyeri et al. (2016), once the lens magnification has been taken into account. It is unlikely that the quasar, hosted in one of the lens galaxies, is contributing to the infrared luminosity, as it is not detected at the lens centroid in our ALMA data. The peak of the SED is bounded by the SPIRE photometry, and the PACS flux densities indicate the presence of a warm dust component. As such, the two temperature SED provides a better fit to the data. With an SFR of $\sim 590 \pm 50\,{\rm M}_\odot{\rm yr}^{-1}$ and a dust mass range of $10^{8.2} - 10^{8.9}{\rm M}_\odot$, this source's SFR to dust mass ratio is typical

of high-redshift SMGs and low redshift ULIRGs according to the empirical relations in Rowlands et al. (2014).

HeLMS J234051.5-041938 - This system displays a nearly-complete ∼0.75 arcsec Einstein ring with three image components. A singular power-law with external shear provides a good fit to the data, with the power-law index preferring relatively high values of around 2.3. The source has evidence of a disturbed morphology, displaying extended faint emission to the west and east.

The peak of the source SED is constrained by our ALMA and SPIRE photometry. The PACS flux densities do not indicate the presence of a significant warm dust component, with both the single and double temperature SEDs providing excellent fits to the data.

This source's intrinsic luminosity of $\log(L_{\rm IR}/{\rm L}_{\odot}) = 13.1 \pm 0.1$ agrees well with the luminosity calculated by Nayyeri et al. (2016). An SFR of $\sim 1900 \pm 300\,{\rm M}_{\odot}{\rm yr}^{-1}$ and a dust mass range of $10^{8.3} - 10^{9.1}\,{\rm M}_{\odot}$ places this source above the typical dust mass to SFR ratio for high-redshift SMGs and local ULIRGs.

HeLMS J232439.5-043935 - This system exhibits an unusual image configuration with three distinct image components, the two westernmost of which show faint extended features, whilst the eastern component is more compact. Two power-law density profiles with an external shear field provide an excellent fit to the data, reconstructing a source with faint extended emission to the north.

The peak in flux density for this source occurs close to the 250 μm SPIRE measurement, with the constraints coming from the PACS measurements. The SED does not show clear evidence of a significant second temperature component, with both SED models providing good fits to the measured flux densities.

With an intrinsic luminosity of $\log(L_{\rm IR}/{\rm L}_{\odot}) = 13.0 \pm 0.1$, our estimate is consistent

within our uncertainties of the value given by Nayyeri et al. (2016). An SFR of $\sim 1500 \pm 300\,\mathrm{M_{\odot}yr^{-1}}$ and a dust mass range of $10^{8.5} - 10^{9.1}\mathrm{M_{\odot}}$ means that this source has a higher than typical ratio of SFR to dust mass for high-redshift SMGs and local ULIRGs.

HeLMS J233255.4-031134 - This is a doubly-imaged source with some faint extended emission emanating from the southern image; it is well fit by a single power-law density profile and an external shear field. The reconstructed source is relatively compact and featureless.

The peak of the SED is constrained by the SPIRE photometry, with the PACS flux densities not showing any major warm dust component. The SED is well described by an optically thick model.

The intrinsic source luminosity we have obtained of $\log(L_{\mathrm{IR}}/\mathrm{L_{\odot}}) = 13.1 \pm 0.1$ is in agreement with Nayyeri et al. (2016). The SFR of $\sim 2000 \pm 300\,\mathrm{M_{\odot}yr^{-1}}$ for this source agrees with the value quoted by Zhang et al. (2018) within our stated uncertainties. A dust mass range of $10^{8.5} - 10^{9.1}\mathrm{M_{\odot}}$ means that this galaxy also lies above the mean SFR to dust mass ratio, as given by Rowlands et al. (2014).

HeLMS J233255.6-053426 - This quadruple image system is well described by a power-law density profile embedded in an external shear field. The source exhibits a relatively featureless, compact morphology. The peak of the source SED is well constrained by the ALMA and SPIRE photometry. The PACS photometry indicates the presence of a warmer dust component, as shown by the relatively high 100 μm PACS flux density measurement, and the significantly better fit of the two temperature SED. Our measurement of the intrinsic source luminosity, $\log(L_{\mathrm{IR}}/\mathrm{L_{\odot}}) = 12.5 \pm 0.1$, is in agreement with the far-IR luminosity given in Nayyeri et al. (2016) using our magnification factor of 9.2 to demagnify the quoted value. With an SFR of $\sim 500 \pm$

50 $M_{\odot}yr^{-1}$ and a dust mass range of $10^{8.5} - 10^{9.0}$ $M_{\odot}$, this source's SFR to dust mass ratio is consistent with typical high-redshift SMGs and low redshift ULIRGs as determined by the empirical relations given in Rowlands et al. (2014).

HeLMS J235331.9+031718 - This double image system has an extremely small image separation, with an Einstein radius of ∼0.1 arcsec. This system is described by a single power-law density profile and an external shear field. The inferred slope of the power-law density profile is relatively low at $\alpha = 1.64 \pm 0.04$, contributing to the high magnification of this source, which also lies near a lensing caustic cusp. The source itself appears to be mostly compact, with an extended feature to the southeast, which is readily visible in the observed lensed image.

There is no redshift measurement for this source, and so we have opted to use the range of redshifts ($z \sim 2 - 3.7$) present in our sample to calculate redshift dependent quantities. There are also no PACS flux density measurements for this source, and so we rely on the SPIRE and ALMA measurements to constrain the SED. The peak of the SED appears to lie within the SPIRE wavelengths, and fitting the optically thick SED for the range of redshifts considered gives a temperature range of $37 - 59$ K. Without the PACS measurements and with the extra free parameters of the optically thin SED, it is not possible to meaningfully infer the presence of a warmer dust component.

4.4 Discussion

Combining our sample with that of Dye et al. (2018), who carried out a similar analysis on a set of 6 strongly lensed sub-mm galaxies observed as part of the same ALMA programme, we can start to make more significant comparisons between our results and those found by other surveys. The HERschel ULIRG Survey (HERUS;

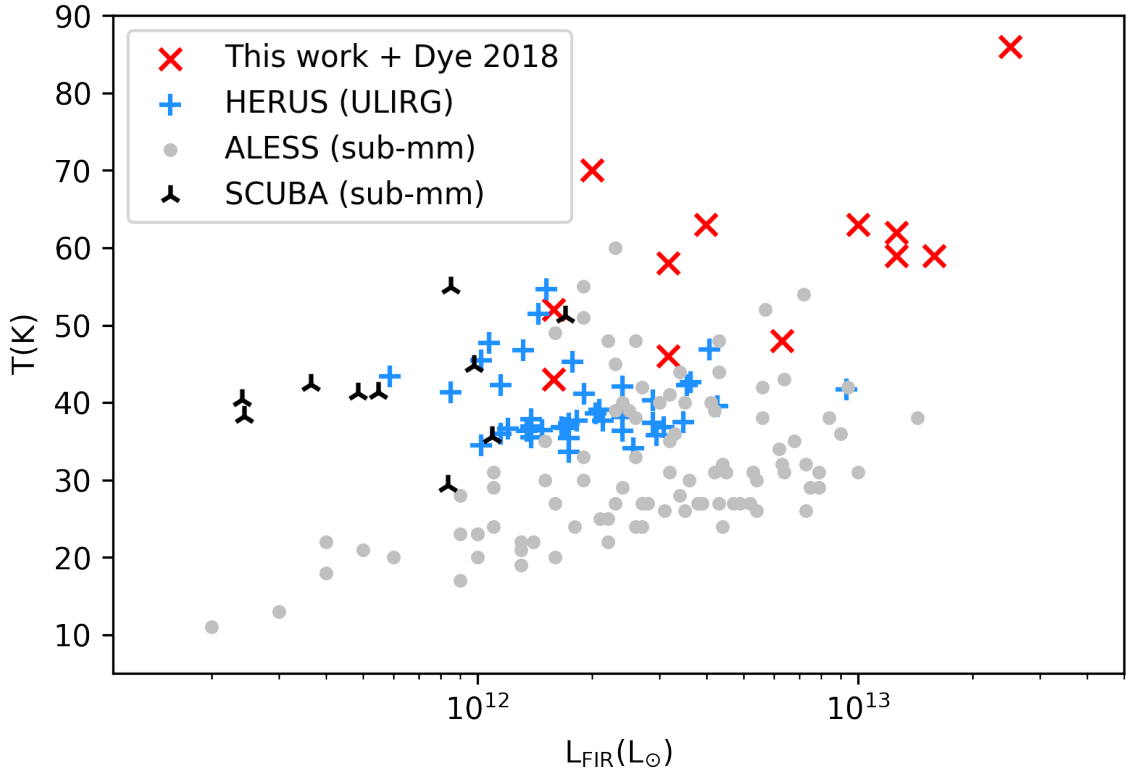


Figure 4.5: Temperature distribution as a function of far-IR luminosity. The red crosses show the luminosities taken from this work (excluding HeLMS J235331.9+031718), and Dye et al. (2018). The spectroscopic redshift range of this combined dataset is approximately $z \sim 1-4$. The blue pluses show the data taken from the HERUS Survey (Clements et al., 2017), with a spectroscopic redshift range of approximately $z \sim 0.02-0.26$. The grey dots come from the ALESS survey (Swinbank et al., 2013), with a photometric redshift range of $z \sim 0.5-6.5$. The black markers show data taken from the SCUBA-2 Lensing Cluster Survey (Hsu et al., 2017), with a redshift range of $z \sim 0.7-4$ (mixed spectroscopic and photometric redshifts).

Clements et al., 2017) is a sample of 43 local ULRIGS, selected at 60 μm by the Infrared Astronomical Satellite. The ALMA LESS survey is a set of cycle 0 and cycle 1 ALMA observations of the sub-mm sources detected in the LABOCA ECDFS Sub-mm Survey (LESS). The 10 sources we present in Figure 4.5 come from the SCUBA-2 Cluster lensing survey and are those sources that have corresponding detections by Herschel, and thus IR luminosities have been derived. Comparing the relationship between the far-IR luminosity and the effective dust temperature (see Fig. 4.5) from the optically thick model between our sample, the ULIRG population from HERUS and the sub-mm sources in ALESS, we can see that our sources tend to possess both more extreme luminosities and higher dust temperatures. The median luminosity of the ALESS sample is $L_{IR} = (3.0 \pm 0.3) \times 10^{12}\,L_{\odot}$, compared with the median luminosity of the HERUS sample of $L_{IR} = (1.7 \pm 0.3) \times 10^{12}\,L_{\odot}$, and finally, that of our sample: $L_{IR} = (5.1 \pm 2.5) \times 10^{12}\,L_{\odot}$. Given our small sample size and large scatter on this measurement, our luminosities are overall consistent with both the HERUS and ALESS samples. The median effective temperature of the ALESS sample is $T = 31 \pm 1\,\mathrm{K}$, which is consistent at the 2σ level with that of the HERUS sample, for which $T = 38 \pm 3\,\mathrm{K}$. The median dust temperature of our sample is significantly higher at $T = 59 \pm 3\,\mathrm{K}$. The high dust temperatures can be interpreted as a product of the extreme SFRs present in our sample, though it is important to bear in mind that the significance of these results is likely explained by the combination of selection effects due to our observing of the brightest galaxies detected by Herschel and the sample of ALESS galaxies being selected at $870\mu m$. Comparison with the sources of the SCUBA-2 Cluster Lensing survey, whose sources have experienced lower magnification factors on the whole than those in the galaxy-scale lenses of our sample, seems to support the notion that our Herschel selected sample is a biased one, preferentially selecting the most intrinsically bright sources.

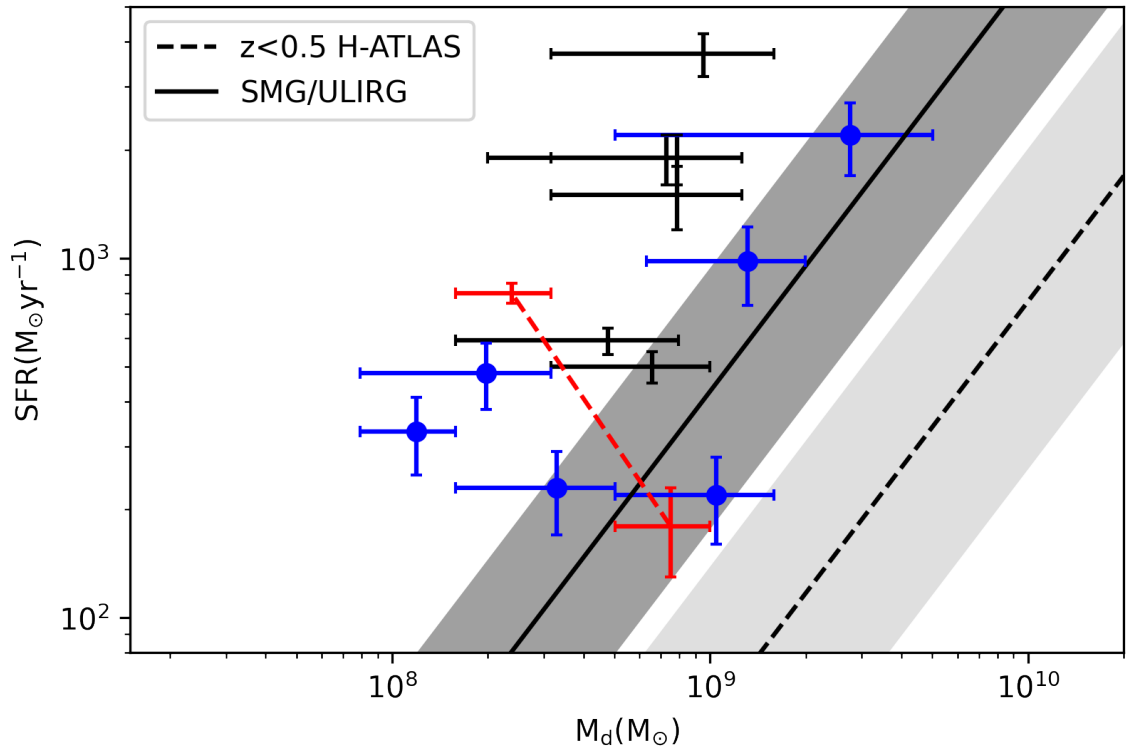


Figure 4.6: SFR, computed with the method of Kennicutt & Evans (2012), plotted against dust masses for our seven lensed sources (black pluses), with the six sources from Dye et al. (2018) shown as blue pluses with a central circle. The horizontal error bars represent the range of dust masses for each source encompassed by $M_{\rm d}^{\rm thick}$ and $M_{\rm d}^{\rm thin}$, whilst the vertical error bars represent the uncertainty in SFR, plotted at the midpoint of the dust mass range. Shown in red are the values for the two extremes of redshift for HeLMS J235331.9+031718. Also shown are the empirical relations between $M_{\rm d}$ and SFR determined by Rowlands et al. (2014) for high-redshift SMGs and low redshift ULIRGs (solid line with 1σ spread indicated in dark grey) and the population of galaxies detected in H-ATLAS with $z < 0.5$ (dashed line with 1σ spread indicated in light grey).

Fig 4.6 shows the degree to which our sources exceed the ratio of the SFR (determined using the scaling relation of the far-IR luminosity given by Kennicutt & Evans, 2012) to the dust mass. Given that our SFRs are derived from scaled far-IR luminosity, we can conclude that our sources have higher than expected luminosities for the amount of gas available for star formation. An obvious interpretation of this fact would be that a large fraction of the luminosity is due to an AGN, but without additional imaging we cannot confirm this. It is also important to note that this excess could be at least in part explained by selection bias, as we have chosen the brightest Herschel sources to follow up on.

Converting the rest frame 850 μm flux density of our sources to H_2 gas mass (see Table 4.5) using the scaling relation given by Hughes et al. (2017), we find that our sources all lie on or above the mean relationship between SFR and H_2 gas mass, as determined by Scoville et al. (2016). One interpretation is that our sources possess a higher star formation efficiency (SFE), provided that dust is an accurate tracer for molecular gas. Fitting a line parallel to the SMG/ULIRG relation shown in Fig. 4.6, and treating the range of dust masses as the 1σ error, we find an increase in SFE by a factor of six when compared with the value implied by the SMG/ULIRG relation from Rowlands et al. (2014). This increases to a factor of 50 when compared to the $z < 0.5$ H-ATLAS galaxies.

An unknown fraction of our sources could have their SFR overestimated due to significant contamination in the infrared by strongly obscured AGN. In order to quantify this effect, additional observational evidence would need to be considered. For example, the stellar mass could be estimated from broadband SED fitting. However, this procedure would either require careful lens light subtraction or additional follow-up observations in the X-ray hard band to reveal the obscured nucleus.

Out of our combined sample of 13 reconstructed sources, nine show signs of possess-

ing a disturbed morphology, with at least two being potential mergers. Morphological studies of sub-mm galaxies have found that it is common for them to exhibit peculiar and complex morphologies; this is suggestive of a high frequency of mergers in the population (Chapman et al., 2003; Ricciardelli et al., 2010). It has been found that samples of ULIRGs in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) (Grogin et al., 2011) fields are more likely to have undergone interactions and mergers than a field galaxy sample Kartaltepe et al. (2012). However, it has also been shown that the morphologies of a sample of spectroscopically confirmed SMGs do not differ significantly from a sample of star forming galaxies in the field Swinbank et al. (2010). Galaxy morphology is strongly correlated with star formation history (Larson et al., 1980; Strateva et al., 2001; Lee et al., 2013). At $z \sim 3$, massive galaxies are mostly star forming disks, with the SFR peaking at $1 < z < 2$ (Madau & Dickinson, 2014), and then dropping as the fraction of massive quiescent galaxies grows rapidly. This period of galaxy evolution is extremely dramatic, and many different mechanisms have been proposed to explain the build up of ETGs we see today. Galaxy mergers are one such mechanism, as they are effective at disturbing the morphology and building a central bulge in a galaxy (Hopkins et al., 2006a; Snyder et al., 2011). They have also been shown to trigger starbursts and AGNs, which can lead to strong supernovae and/or AGN winds contributing to the quenching of a galaxy (Bekki et al., 2005).

4.5 Conclusions

We have modelled seven ALMA observations of strongly lensed sub-millimetre galaxies. Four of these systems are galaxy-galaxy scale strong lenses, which are well described by a single power-law mass profile; the remaining three are group-scale lenses,

which have been successfully fitted with two power-law mass profiles. Where we found it improved the fit, an external shear term was also included in the lens model. In this work, we have opted to model the visibility data directly, rather than to work with CLEANed image plane data. Whilst the uv-plane method is more computationally expensive, it does not suffer from the image pixel covariances introduced due to incomplete sampling of the uv-plane. However, Dye et al. (2018) showed that with sufficient sampling of the uv-space, both the image-plane and direct visibility modelling approaches produce very similar results. We have fitted a smooth power-law density profile (and in some cases two profiles) to each of the lensing systems, and found that most of the lenses are close to isothermal. This result is expected in massive ETGs due to the combination of an inner Sérsic profile representing the baryonic component and an outer Navarro-Frenk-White (Navarro et al., 1996) profile representing the dark matter component (Lapi et al., 2012). In some instances, there are significant deviations from an isothermal power-law slope, which may be due to degeneracies between parameters in our model or reflect the true nature of the lens, but we leave a more thorough explanation of this to further work.

By obtaining the total magnification factors from our models, we have demagnified the available sub-millimetre source photometry. Fitting rest frame SEDs to these data allowed us to determine the dust temperatures, dust masses, intrinsic luminosities, and SFRs of our lensed sources. In order to estimate a range of possible dust masses for these sources, we fitted the photometry with both single temperature, optically thick SEDs, as well as dual temperature, optically thin SEDs. Using the midpoint of this range to calculate the SFR to dust mass ratio, we find that all seven of our sources lie above the mean ratio for the SMG/ULIRG population as described in Rowlands et al. (2014).

Our combined sample of strongly lensed sub-mm galaxies contains a majority of

sources that display a disturbed morphology. Nine of the 13 galaxies in our sample are visually classified as being disturbed, with at least two of them having evidence of being mergers (H-ATLAS J142935.3-002836, and H-ATLAS J083051.0+01322). Other observations of sub-mm galaxies have found similarly high fractions of disturbed morphologies, such as those made by Chapman et al. (2004), who used high-resolution optical and radio imaging of 12 sub-mm galaxies in order to study their spatially extended star formation activity. It has been suggested that high density molecular gas is more commonly found in galaxy mergers than quiescent systems, and that this can be used to predict the star formation mode of a galaxy (Papadopoulos & Geach, 2012). We are not able to conclusively say what fraction of our sample's disturbed morphologies are a result of mergers, but the source with the most extreme ratio of SFR and gas mass (H-ATLAS J083051.0+01322) does display a significantly disturbed morphology, and is identified as being a merger by Yang et al. (2019).

Chapter 5

Lens modelling biases in CLEANed images

As has been discussed in Section 1.2.1, SMGs are thought to be the progenitors of quiescent galaxies seen at low redshift, and as such are an important area of study for our understanding of galaxy evolution. The extreme SFRs of these sources must be fuelled by large gas reservoirs, and studies have found strong evidence for this (Huynh et al., 2017; Dye et al., 2018; Kaasinen et al., 2019). A difficulty of studying these high redshift sources is the necessary angular resolution needed to probe their ISM at sub-kpc scales, or the scale of giant molecular clouds ($\sim$ 100 pc). Gravitational lensing is uniquely placed to aid in the study of SMGs due to the boosting of the background source's apparent flux and magnification of their extent on the sky. The combination of this, with the high-resolution configurations of ALMA, has huge potential for investigating the physical properties of the ISM and furthering our understanding of the processes at play in these high redshift galaxies (Dye et al., 2015; Rybak et al., 2015b; Dye et al., 2022b).

Understanding the properties of lensed SMGs is contingent upon our ability to

produce accurate lens models, and thus faithfully reconstruct the properties of the background source. Likewise, if we wish to study the properties of the lensing galaxy, for example to study the evolution of the slope of the density profile (see Section 1.2.2), we must be able to accurately model the lens mass profile. However, when dealing with ALMA (or any other interferometer) observations, there are a number of considerations one must take into account before embarking upon the analysis. Chief amongst these are whether to model the visibility data directly, or to employ a deconvolution algorithm such as `clean` (Högbom, 1974) to form an image that can subsequently be modelled. Modelling the visibility data has the benefit of working with the purest form of the data, but comes at the cost of time and computational resources. For this reason, it is often attractive to work with a derived image-plane representation of the data, which is significantly easier to model, but introduces artefacts due to the discrete sampling of the uv-plane in the form of correlated noise. This is further complicated by the many choices afforded in the production of these images, such as how to weight the visibilities (see Section 1.6.2). There has been some work showing that the differences between the inferred models when working with the image-plane versus the uv-plane data are negligible, in cases of good uv-plane sampling Dye et al. (2018). However, there are still many questions to be asked and explored on this topic, and that is what we set out to do here. Some of the questions that we seek to answer are: Does uv-plane modelling provide a more accurate lens model than its image plane counterparts? Does the amount of time-binning effect the resulting lens model? Does our choice of visibility weighting scheme lead to the inference of a different lens model? Do these different weighting schemes introduce different biases?

The layout of this chapter is as follows: Section 5.1 describes in detail the processes used to produce all the simulated data products used in this paper. Section 5.2 outlines

the lens modelling methodology we have used. The results of our lens modelling are presented in Section 5.3, and their interpretation are described in Section 5.4. Finally, Section 5.5 presents a brief summary and our main conclusions. Throughout this paper, we assume a flat ΛCDM cosmology using the 2015 Planck results (Planck Collaboration et al., 2016), with Hubble constant $H_0 = 67.7\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ and matter density parameter $\Omega_m = 0.307$.

5.1 Simulated Data

This work relies on a variety of simulated data products to compare modelling results between them. Interferometer datasets, and images derived from interferometer measurements, are used. In this section, we describe the process used to create the simulated data products used in this paper.

The starting point for all of our simulated data was three sets of eight ray-traced lensing simulations, produced using `PyAutoLens` (Nightingale & Hayes, 2020). These three sets of simulated images consist of one SIS mass model and two SIE mass models, the parameters of which can be found in Table 5.1. In each case, a spherical Gaussian source was used to produce eight lensed images, with the source being stepped away from the optic axis by 0.1 arcsec each time. In the case of the SIE lensed images, the two sets represent the source being stepped through a fold and cusp configuration (see Figure 5.1). These simulated images are then used as the sky brightness distribution for our simulated observations.

We used the Common Astronomy Software Application V6.3.0.48 (`CASA`) and the `simobserve` task to produce mock ALMA measurement sets. We used the ALMA cycle7.3 array configuration and three different observing times (60s, 600s, 6000s) were employed (see Table 5.2 for a summary of the setup). The coverage of

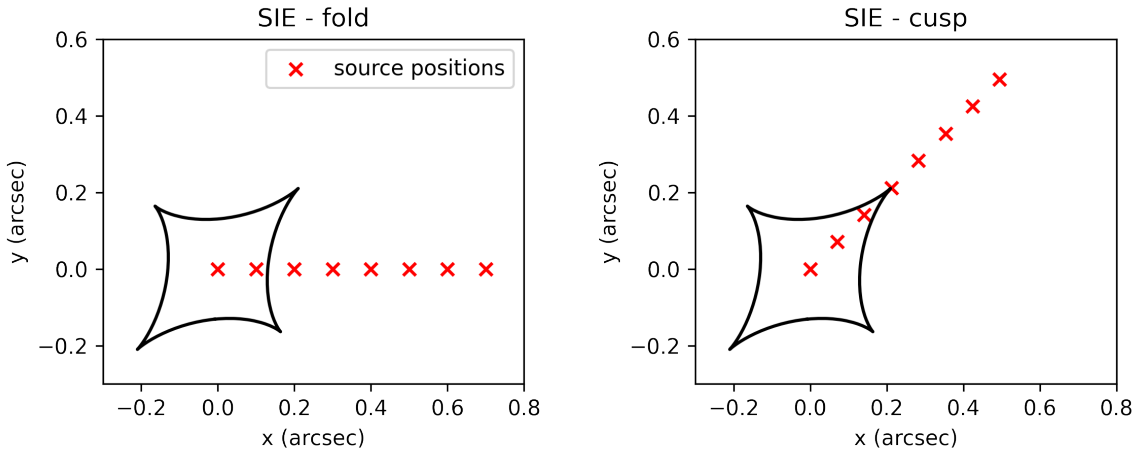


Figure 5.1: The source positions relative to the caustics of the two SIE datasets used in this work. A lensed image was created for each of the source positions. The left-handed panel shows the source stepping away from the optic axis in the direction of the caustic fold, whilst the right-handed panel shows the source moving towards the cusp of the caustic.

Table 5.1: A summary of the two lens models used to create the simulated images in this work. The columns are the Einstein radius θ_E, the lens centroid (x_c, y_c), corresponding to west and north respectively, the orientation of the lens ϕ, measured counter-clockwise from east to the lens semi-major axis, and the lens profile axis ratio.

Lens model	θ_E (arcsec)	(x_c, y_c) (arcsec)	ϕ (deg)	q
SIS	1	(0, 0)	N/A	1
SIE	1	(0, 0)	45	0.67

the uv-plane for each of these observing setups can be seen in Figure 5.3. We assume a Spectral Energy Distribution (SED) for our lensed sources that is constant with S_ν. The spectral setup in our simulated ALMA observations was identical for each system. The band 7 continuum observations, comprised of four spectral windows, each with a width of 2000 MHz and centred on the frequencies 337.5 GHz, 339.5 GHz, 347.5 GHz and 349.5 GHz. The central frequency of 343.404 GHz corresponds to a wavelength of 873 μm. Each spectral window consists of 128 frequency channels, resulting in a spectral resolution of 15.6 MHz. The noise model for these simulated ALMA observations is generated by the `simobserve` task, using pre-defined values that correspond to good observing conditions (i.e. precipitable water vapour of 0.5 mm and an ambient temperature of 269 K). The three on-source integration times of 60s, 600s, and 6000s lead to a total number of 21672, 216720 and 2167200 visibilities respectively. In order to make the challenge of modelling these large numbers of visibilities more tractable, we time bin the data, using three different sizes of bins, 10s, 20s and 30s. In addition to reducing the computational cost of the modelling process, this allows us to make an initial exploration of the question of how much time binning is appropriate for problems of this type.

As well as modelling the visibility data directly, we perform image-plane modelling for each system by producing images via the `tclean` algorithm. When creating images with `tclean`, one of the choices that must be made is how to weight the visibility data. A common choice is natural weighting, where visibilities are weighted according to the inverse of the noise variance of that visibility, determined during calibration (see Section 1.6.2), which provides good signal-to-noise and sensitivity, but relatively poor resolution. Another option is that of uniform weighting, where the weights are adjusted based on the density of points in the uv-plane, giving greater weighting to sparse regions. This achieves better resolution, but at the cost of poor signal-to-noise.

A third option, Briggs weighting, allows for a smoothly varying compromise between the two aforementioned methods by modifying the value of the 'robustness' parameter R. This parameter varies between the values of -2, which is approximately equivalent to a uniform weighting and $+2$, which closely resembles the natural weighting scheme (Briggs, 1995). For a more detailed discussion of synthesis imaging, see Section 1.6.2. From each of our simulated measurement sets, we produce two images; one with natural weighting and another with Briggs weighting with the robustness parameter set to zero, which provides a good tradeoff between resolution and sensitivity. We use an image pixel scale of 0.1 arcsec pixel^{-1}, providing good sampling of the minor axis of the primary beam. A threshold value of twice the background noise was used as the stopping criteria for the `tclean` algorithm in all cases.

5.2 Methodology

All the modelling in this work was carried out using the semi linear inversion method (Warren & Dye, 2003) and the extensions to this method that adapt to the source-plane magnification or source-plane brightness distribution (Nightingale et al., 2018) as it's implemented with `PyAutoLens` (Nightingale & Hayes, 2020). The standard image-plane method was employed for modelling the data produced via the `tclean` algorithm, whilst modelling of the visibility data was carried out using a modified version of this technique as implemented within PyAutoLens.

5.2.1 Image-plane modelling

In order to model our image-plane data using the semi-linear inversion method, we needed to acquire an estimate for the noise in our images. This was achieved by

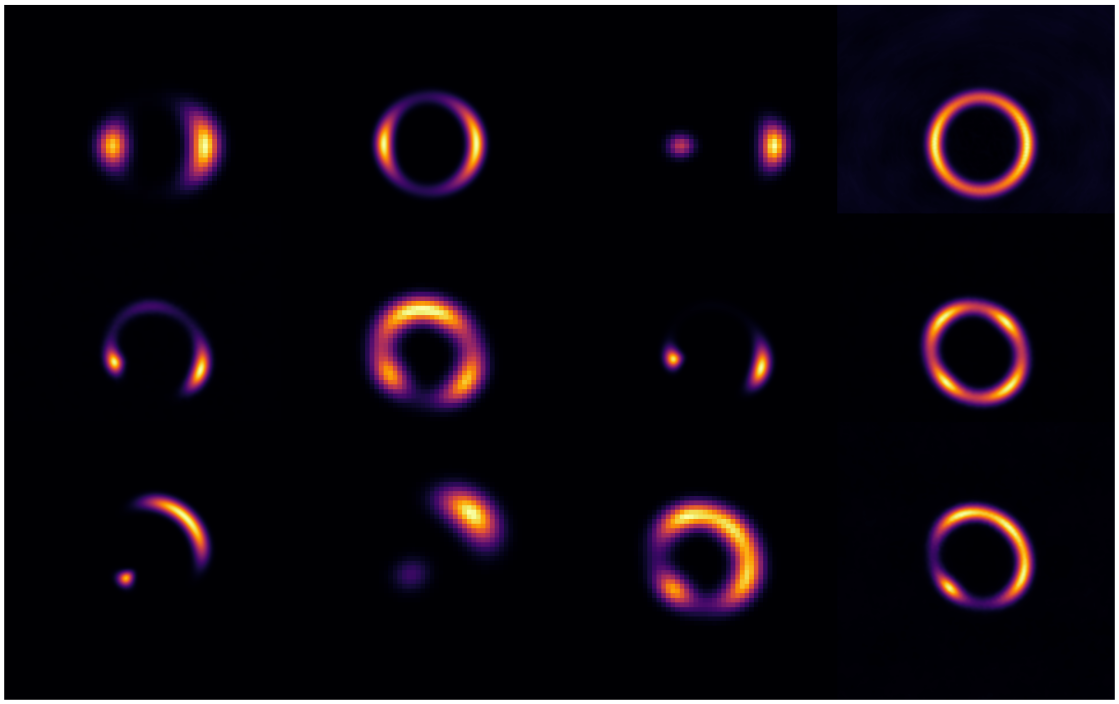


Figure 5.2: A selection of the images produced via the `tclean` algorithm, with a variety of on-source integration times, ALMA configurations and weighting schemes. The first row contains sources lensed by the SIS model. The second row contains sources lensed by the SIE model, with the source positions in the fold configuration. The third row contains sources lensed by the SIE model, with the source positions in the cusp configuration.

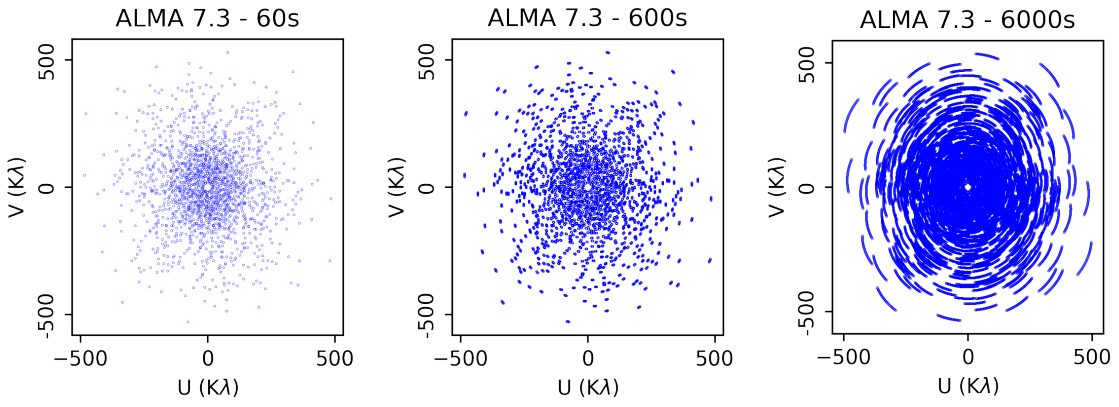


Figure 5.3: The uv-coverage of the ALMA cycle7.3 configuration as a function of the different observation times used in the creation of the simulated datasets in this work.

Table 5.2: A summary of the targeted position of each source, the number of antennae used in each observation, the continuum sensitivity, the shortest & longest baseline in the ALMA configuration, and the angular resolution of the setup.

ALMA config	Position (RA DEC)	$N_{\rm ant}$	Sensitivity (μJy/beam)	Shortest baseline (m)	Longest baseline (m)	Angular resolution (arcsec)
Cycle 7.3	19:00:00 -40:00:00	43	(220, 70, 22)	15	500	0.41

measuring the standard deviation of a 25×25 pixel region of each image, far away from any lensed source emission. Due to the incomplete sampling of the uv-plane, and the covariant noise that this introduces into the images, we opted to use a 1σ value of twice the measured standard deviation in our images.

The first stage of our modelling process was to reconstruct the source on a Voronoi grid, adapted to the magnification of the source plane. We assumed that all our observations were properly centred on the lens, and used a Gaussian prior on the (x, y) coordinated of the lens centroid with $\mu = 0$ and $\sigma = 0.2''$. A uniform prior of width $0.5''$, centred on the true value, was used for the Einstein radius. Uniform priors covering the whole range of parameter space were used for the ellipticity components. We adopted a gradient regularisation scheme, as we have done throughout this thesis, with a constant weight for the source plane. This magnification based fit was then used to initialise a new fit, using a Voronoi grid, this time adapted to the reconstructed source light (Nightingale et al., 2018). The nested sampling algorithm `dynesty` (Speagle, 2020) was used to maximise the Bayesian evidence (Suyu et al., 2006) and find the optimal lens model parameters. For a more detailed description of the semi-linear inversion process works, see Section 1.7.

5.2.2 uv-plane modelling

The uv-plane modelling carried out in this work follows the same setup as the described above, but using the modified version of the semi-linear inversion method that was developed by Powell et al. (2020) and has been implemented within `PyAutoLens`. See Section 1.7.3 for more details on how this process works.

5.2.3 Lens Model

To fit our data, we have once again used the elliptical power-law density profile. All of our simulated lenses have isothermal profiles, but we are interested in our ability to recover the slope of the density profile, and so we allow for this extra degree of freedom in our modelling. The surface mass density, κ, of this profile is given by:

$$\kappa(R) = \frac{3-\alpha}{1+q}\left(\frac{\theta_{\rm E}}{R}\right)^{\alpha-1}, \tag{5.1}$$

where $\theta_{\rm E}$ is the Einstein radius in arc seconds, the power-law α index and R the elliptical radius defined by $R = \sqrt{x^2 + y^2/q^2}$, where q is the axis ratio (Suyu, 2012). The orientation of the lens ϕ is measured as the angle counterclockwise from east to the semi-major axis of the elliptical lens profile. The centre of the lens profile is given by the coordinates in the image-plane $(x_{\rm c}, y_{\rm c})$. This results in six lens model parameters.

5.3 Results

5.3.1 Time-binning visibilities

The first question we wish to answer is, "to what extent are our uv-plane lens modelling results impacted by our choice of time binning in the visibility data?". Time averaging of visibilities is a means to reduce the overall data volume and thus reduce the computational cost of performing lens modelling in the uv-plane. Since the process of visibility modelling is more computationally expensive, compared to its image plane counterpart, it is useful to know if our attempts to make the task more tractable have negative impacts on our results. Here, we focus on the effects of time-averaging

on our shortest observation (60s on-source integration time), since a relatively small change in the size of the time-binning will produce the largest effect on datasets with low numbers of visibilities. i.e a 30 second time-binning value is a much more extreme value to use for the 60s integration time observations compared with a 6000s integration time.

As described in Section 5.1, for each of the three lensing scenarios considered in this work (and each of the 8 source positions), we produced three visibility datasets with different time-binning values of 10s, 20s, and 30s. We used `PyAutoLens` to model each of these datasets, and the results are presented in Figures 5.4, 5.5, and 5.6.

Inspecting the results displayed in Figures 5.4, 5.5, and 5.6, it is clear to see that for all three of the simulated scenarios, the value of the time-binning plays little to no role in the overall performance of the uv-plane modelling. Except for a handful of outliers, the three time-bin values are all consistent with one another and there is no striking evidence that moving to smaller values of time-averaging is worth the increased computational cost. In particular, we see no improvement in the bias to lower axis ratios at any source positions in the SIS model when moving to smaller values of time-binning. Likewise, the tendency to overestimate the value of the slope of the power law mass model at small source offsets in the SIS model is not clearly improved by employing less time-averaging. These results therefore justify our use of 30s time averaging for all of our datasets, as the reduction in computational cost outweighs any potential benefit in modelling accuracy. As such, all the uv-plane modelling results in the rest of this work have been time-averaged with a time bin of 30s, unless explicitly stated otherwise.

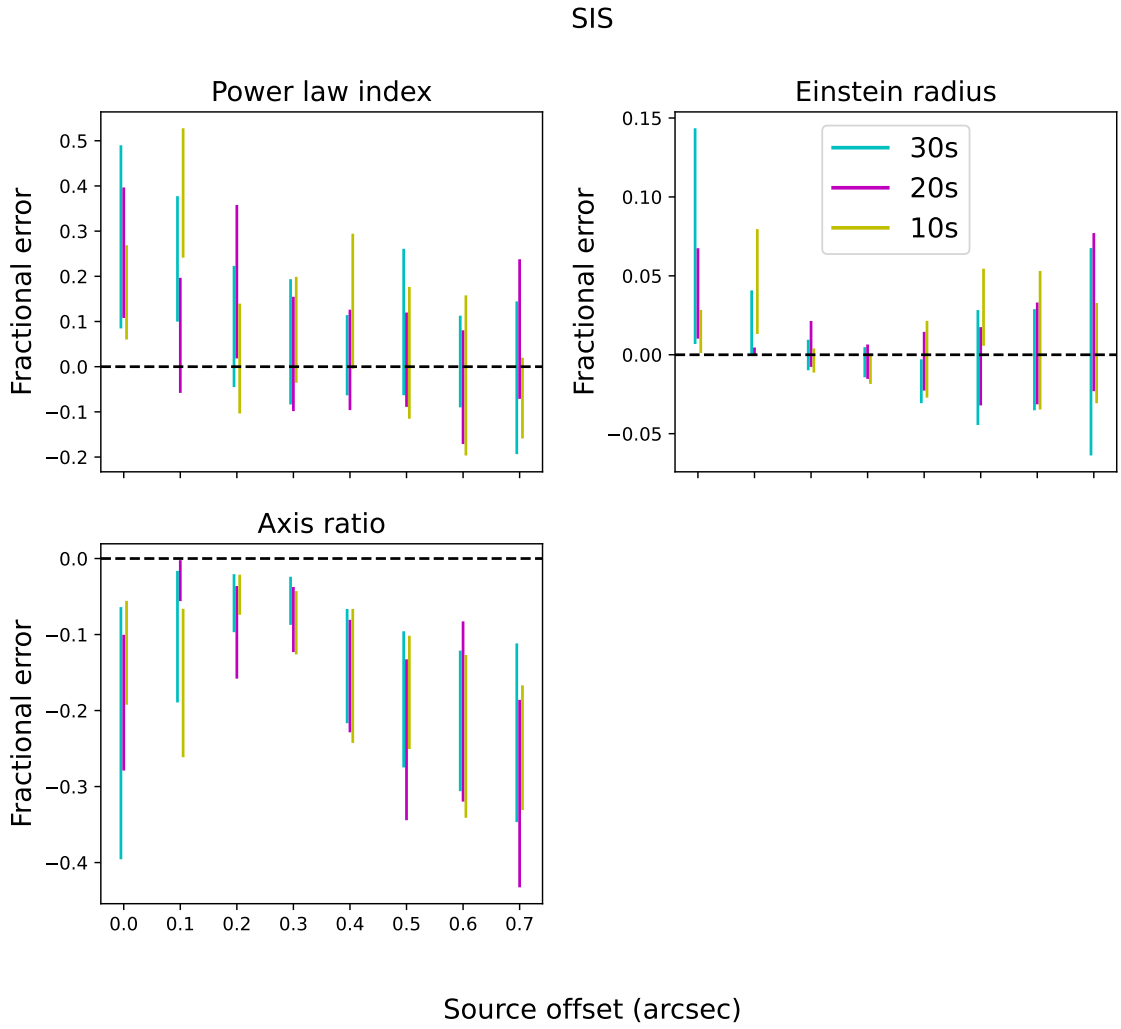


Figure 5.4: A comparison of uv-plane lens modelling results for three different time-binning values of visibilities. In each case, we show the results from time-binning at 10, 20 and 30 seconds (shown in yellow, magenta and cyan respectively). The three panels show the 1-σ spread for the lens parameters of the SIS model, as a function of source offset from the lens centroid.

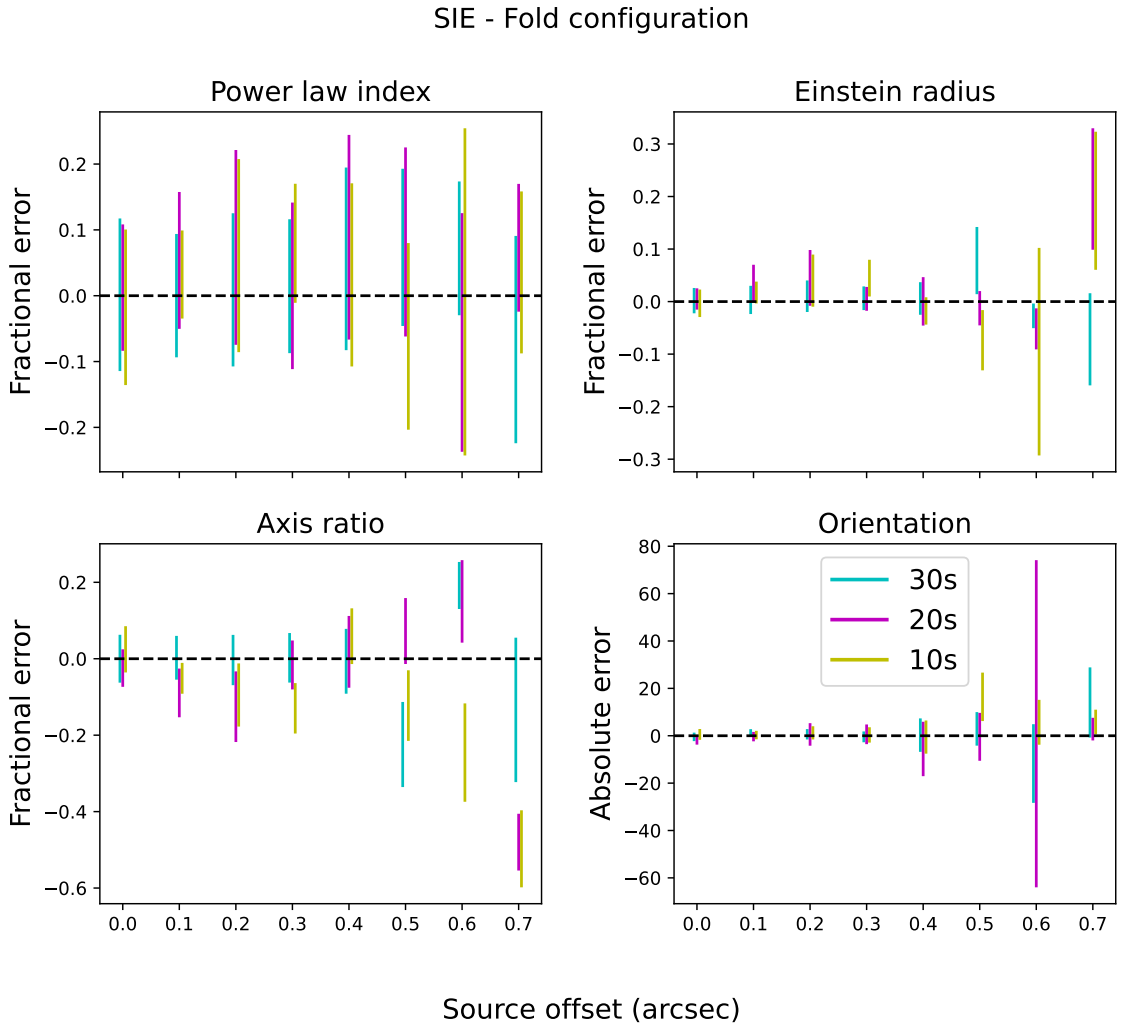


Figure 5.5: A comparison of uv-plane lens modelling results for three different time-binning values of visibilities. In each case, we show the results from time-binning at 10, 20 and 30 seconds (shown in yellow, magenta and cyan respectively). The four panels show the 1-σ spread for the lens parameters of the SIE model (source moving through the caustic fold), as a function of source offset from the lens centroid.

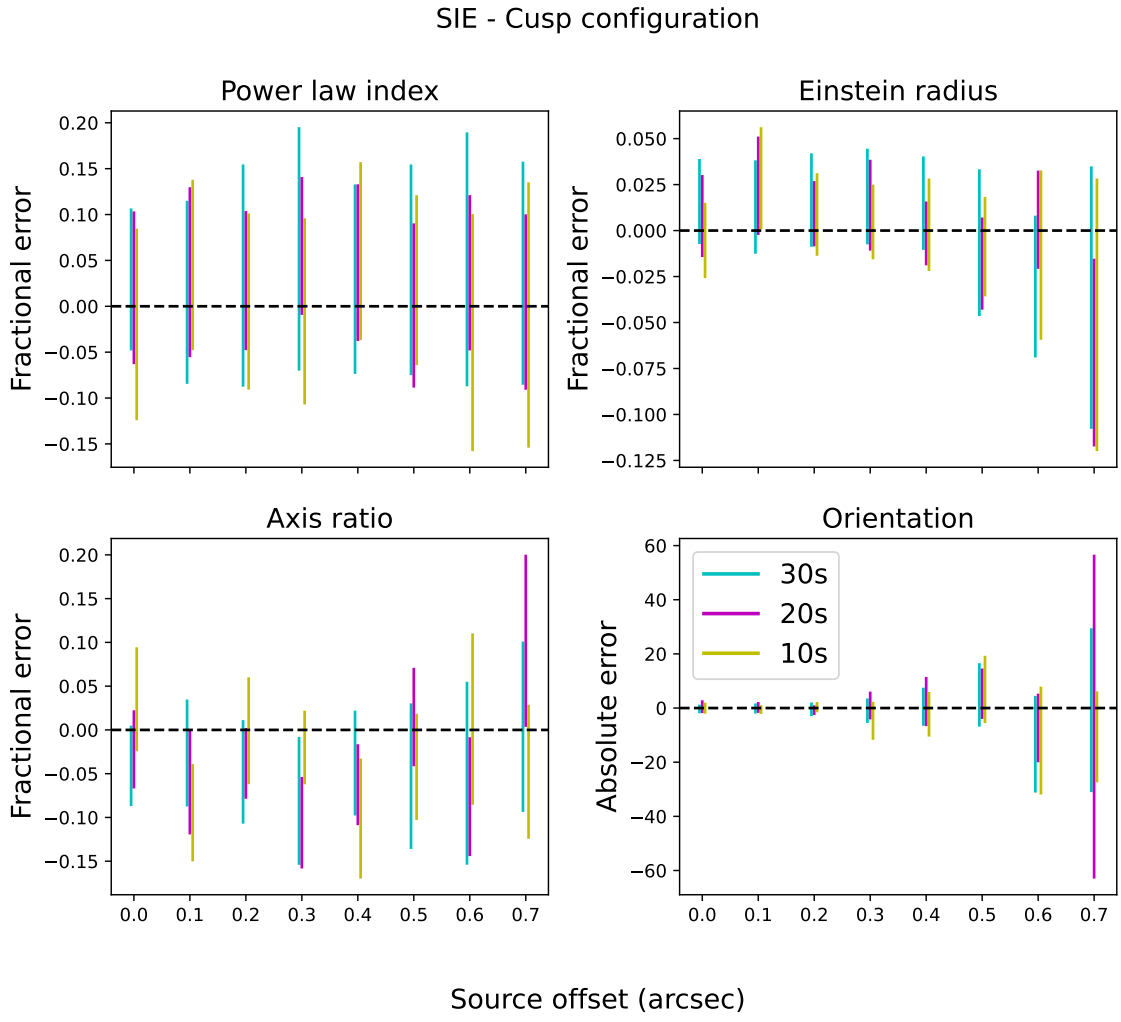


Figure 5.6: A comparison of uv-plane lens modelling results for three different time-binning values of visibilities. In each case, we show the results from time-binning at 10, 20 and 30 seconds (shown in yellow, magenta and cyan respectively). The four panels show the 1-σ spread for the lens parameters of the SIE model (source moving through the caustic cusp), as a function of source offset from the lens centroid.

5.3.2 Modelling results

In this section, we present the results of our lens modelling for each of the different data products considered in this work. In some instances, it is clear that our modelling has not converged upon the global solution, but rather upon an unphysical, demagnified solution (see Maresca et al., 2020, for more discussion). Where this is the case (confirmed by visual inspection of the source reconstruction), we have opted to mask the data point, as we are interested in biases on lens model parameters that correspond to the global solution.

Figure 5.7 displays the ensemble of results from modelling the SIS lens with the source at varying offsets from the optical axis. Each quadrant of the plot shows the fractional error on a lens parameter as a function of source offset for both natural and Brigg's weighting, as well as for the direct visibility modelling approach. Each subpanel contains the results for the three on-source integration times as indicated by the legend. The modelling of the naturally weighted images performs well for determining the power law index, but significant biases in the inferred value of Einstein radius and axis ratio begin to appear at approximately $0.4''$ of source offset. The results for the Briggs weighted images are largely similar, but the extent to which the inferred values for θ_E and q are biased at large source offsets is less than that of the naturally weighted images. The direct modelling of visibility data appears to show a small bias in the value of α at small source offsets. The inferred values of Einstein radius for this approach do not show any evidence of being biased at large values of source offset. There is, however, a bias at all source offsets in the axis ratio value inferred by the direct visibility modelling approach. This bias remains approximately constant over the range of offsets considered here.

Figure 5.8 follows the same layout as Figure 5.7, with the addition of the bottom

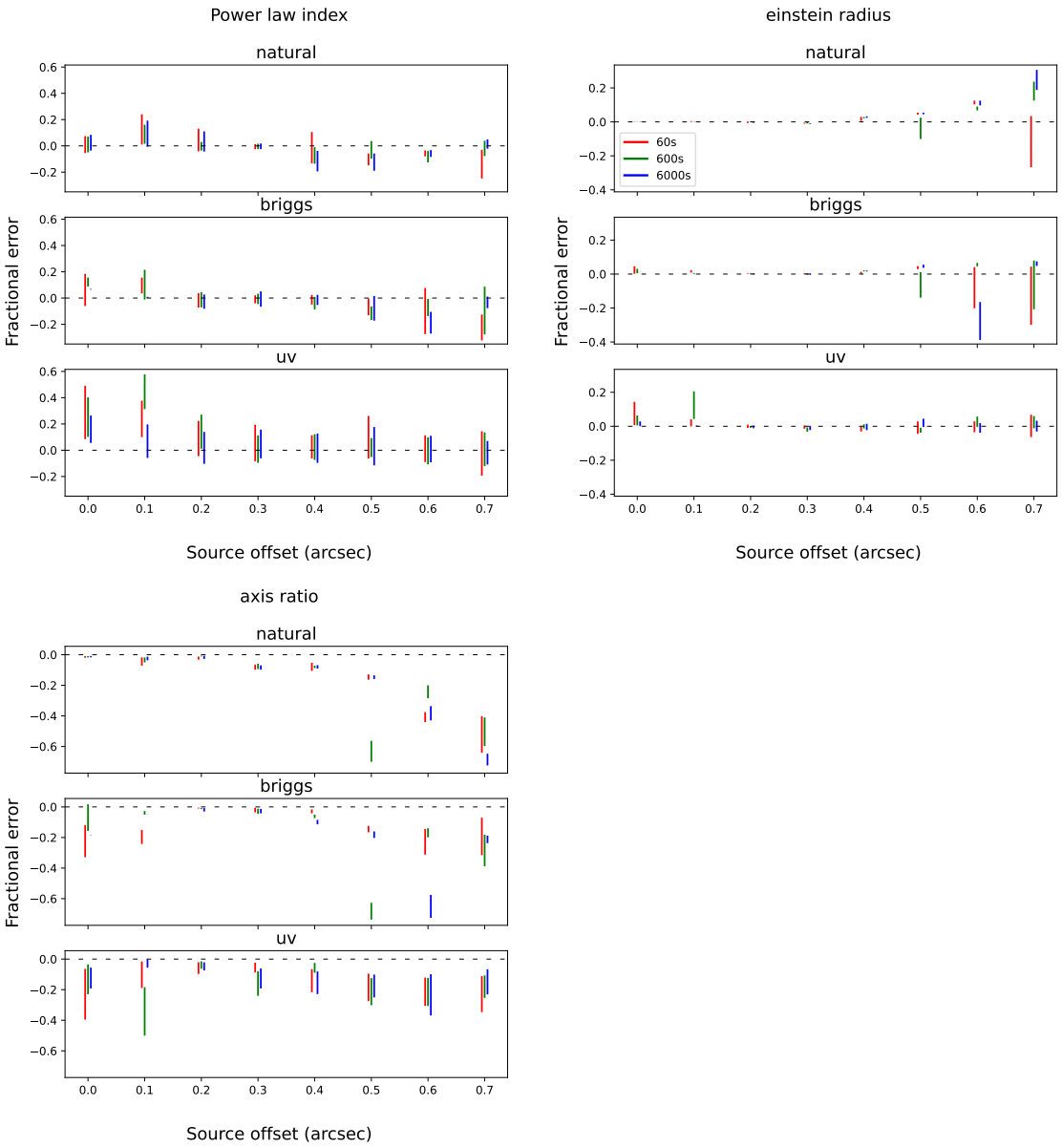


Figure 5.7: Fractional error in SIS lens model parameters as a function of the source offset from the optical axis. The error bars represent the 1-σ range on the inferred parameter values. The legend indicates the on-source integration time of each measurement. The three sub-panels for each parameter represent modelling carried out on images created with natural weighting, Briggs weighting, and direct visibility modelling respectively (from top to bottom).

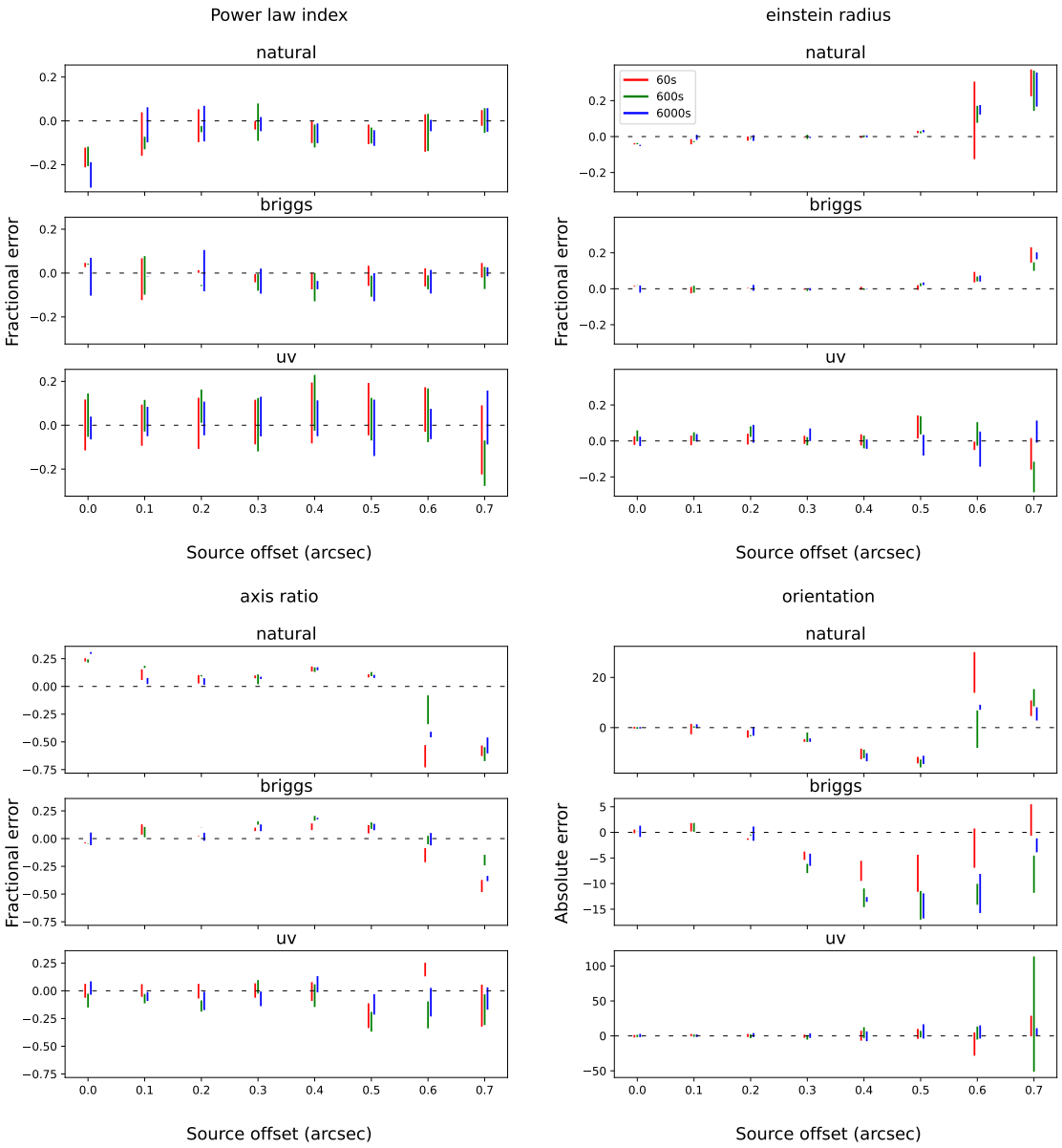


Figure 5.8: Fractional and absolute error in SIE lens model parameters (fold configuration) as a function of the source offset from the optical axis. The error bars represent the 1-σ range on the inferred parameter values. The legend indicates the on-source integration time of each measurement. The three sub-panels for each parameter represent modelling carried out on images created with natural weighting, Briggs weighting, and direct visibility modelling respectively (from top to bottom).

right panel displaying the results for the orientation of the mass profile. Comparing the results of the two image-plane approaches with that of the uv-plane results, we see that all three methods perform comparably well at recovering the true value of the power law index over the full range of source offsets considered. The naturally weighted image results, for all three on-source integration times, show some evidence of a bias towards lower values of α when the source is on the optical axis. However, both image-plane datasets show a tendency to under-estimate the power law index, whilst still being consistent with the truth. For the Einstein radius, we see again that the naturally weighted image result biases towards lower values when the source is on the optical axis; an issue that does not arise when working with either the Briggs weighted imaging or the visibility data. At large source offsets, both the image-plane approaches start to bias towards higher values of θ_E, but once again, the magnitude of this bias is greater for the naturally weighted images. The uv-plane results do not suffer from this bias, maintaining accurate recovery of the parameter value at all source offsets. Biases in the power law index and Einstein radius also translate into biases in the axis ratio, due to the well known degeneracy between this triplet of parameters. The more extreme biases in θ_E of the naturally weighted images result in larger biases in q compared with those we see for the Briggs weighted images. There is some significant scatter in the uv-plane modelling results, but no clear systematic bias. The uv-plane modelling provides a consistent and accurate measurement of the lens orientation, whilst both image plane datasets appear to suffer from the same bias. The bias in lens orientation correlates with the bias in axis ratio; where the value of q is over-estimated, the bias in lens orientation increases.

Figure 5.9 follows the same layout as Figure 5.8. For the inference of the power law index, the visibility modelling approach significantly outperforms the image plane modelling, across the full range of source offsets. Similarly to the fold configuration,

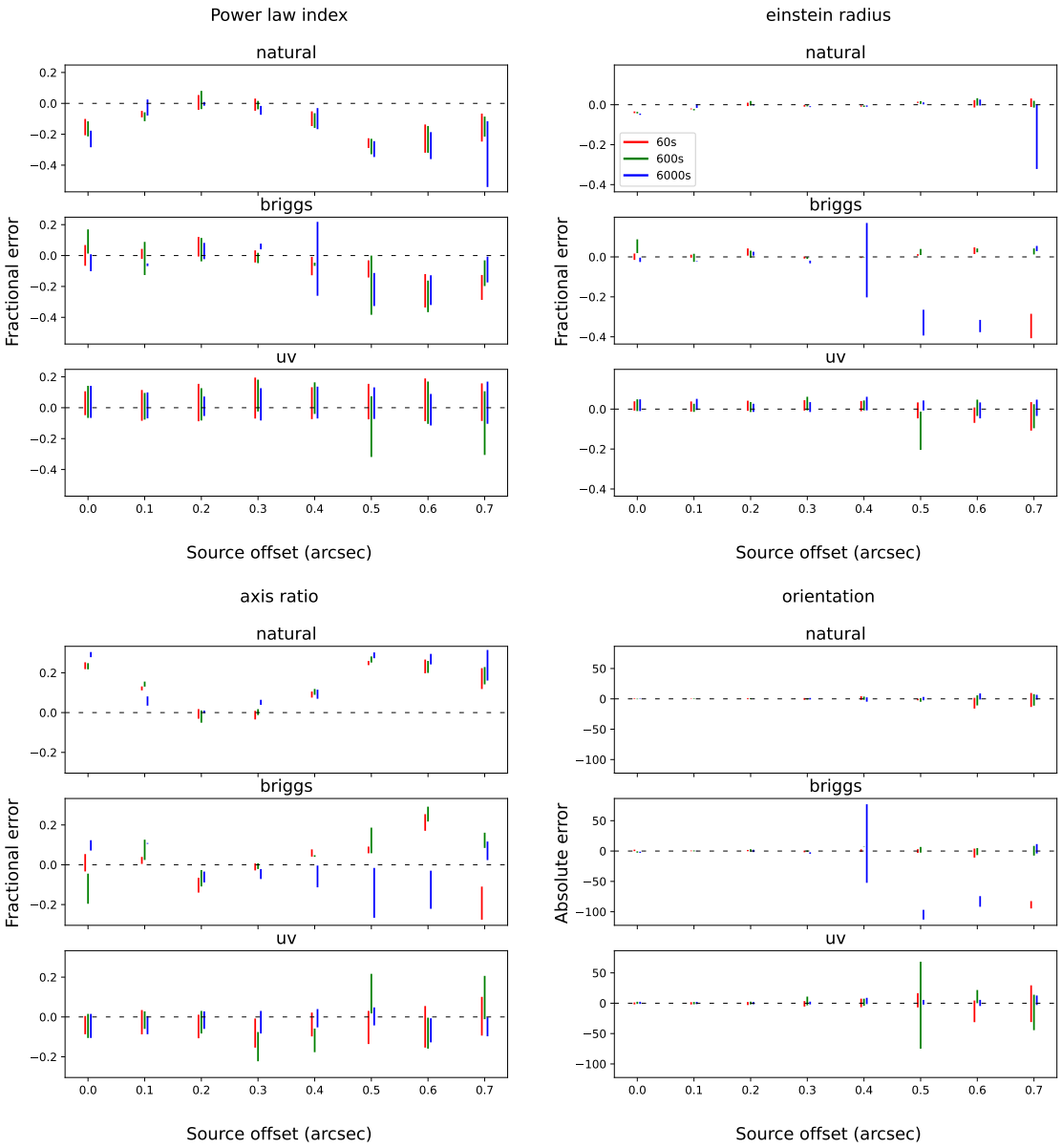


Figure 5.9: Fractional and absolute error in SIE lens model parameters (cusp configuration) as a function of the source offset from the optical axis. The error bars represent the 1-σ range on the inferred parameter values. The legend indicates the on-source integration time of each measurement. The three sub-panels for each parameter represent modelling carried out on images created with natural weighting, Briggs weighting, and direct visibility modelling respectively (from top to bottom).

the results of modelling the naturally weighted images show a tendency to under-estimate α when the source is on the optical axis. In fact, with the source on the optical axis, these images are identical except for different realisations of random noise. This behaviour is once again not seen in either the Briggs weighted images or the visibility modelling. At larger source offsets, both image plane methods tend to under-estimate α. As expected, the $0.0''$ offset source for the naturally weighted images shows a bias towards lower values in θ_E. In general, the modelling of all three data sources provides a reliable estimate of the Einstein radius. The visibility modelling approach provides the least biased estimate of the lens axis ratio for all source offsets. Both the naturally and Briggs weighted datasets exhibit significant biases at large source offsets. The lens orientation is recovered without any signs of bias in any of the data sources.

5.4 Discussion

The relatively good performance of the modelling of the Briggs weighted data, compared to that of the naturally weighted data, suggests that there is important information on small scales that is being lost by using natural weighting. At least, for this simple lensing scenario, with a parametric source, any benefits from the increased sensitivity of natural weighting appear to be outweighed by the greater resolution achieved by Briggs weighting. In order to explore this effect, we generate a new set of lensed images using `PyAutoLens`. We consider two scenarios: one where the source is on the optic axis; and another where the source is located at $(0.7'', 0.0)$ with respect to the centre of the lens. For each scenario, we produce ray-traced images of the source, lensed by a power law profile with $\theta_E = 1''$, axis ratio varying as $q \sim [0.4, 0.9]$, and power law index varying as $\alpha \sim [1.8, 2.2]$. Following the same procedure as described

in Section 5.1, we produce two images using the `tclean` algorithm, one with natural weighting and the other with Briggs weighting and a robustness parameter of 0.

Figures 5.10 and 5.11 show the variance of the Fourier transformed true source surface brightness distributions as a function of axis ratio, and power law index respectively, with the source on the optical axis. This is computed by stacking a set of i images in the z-plane, and computing the variance along this axis for each (k_1, k_2) position in the Fourier transformed true source surface brightness distributions. There is a wealth of information contained at high k-numbers, indicated by the intensity at large radial distances from the centre of the Fourier transformed images. It can be seen from Figure 5.10 that the variance of Fourier components, as a function of axis ratio, at large k increases with increasing power law index. Figure 5.11 shows that the variance of Fourier components, as a function of power law index, at high k numbers has a weak radial dependence on the axis ratio, but particularly for very elliptical lenses, there is a strong directional component that is most prominent at low k numbers. In both instances, there is significant variance at high k numbers, indicating that variations in the lens model parameters, q and α, lead to changes at high spatial frequencies, that if captured, could be useful in constraining the lens model.

Figures 5.12 and 5.13 respectively, show how this variance as a function of axis ratio and power law index changes when we displace the source from the optic axis to (0.7", 0.0). Generally speaking, the variance in both cases becomes less circularly symmetric and more pronounced at high k-numbers. This strong azimuthal dependence on the variance, however, does not change the conclusion that there exists a large amount of information at high k-values that could in principle be used to constrain our lens model.

The significance of this variance at high k-numbers can be seen by looking at the relative sensitivity to these scales in the naturally and Briggs weighted images. Due

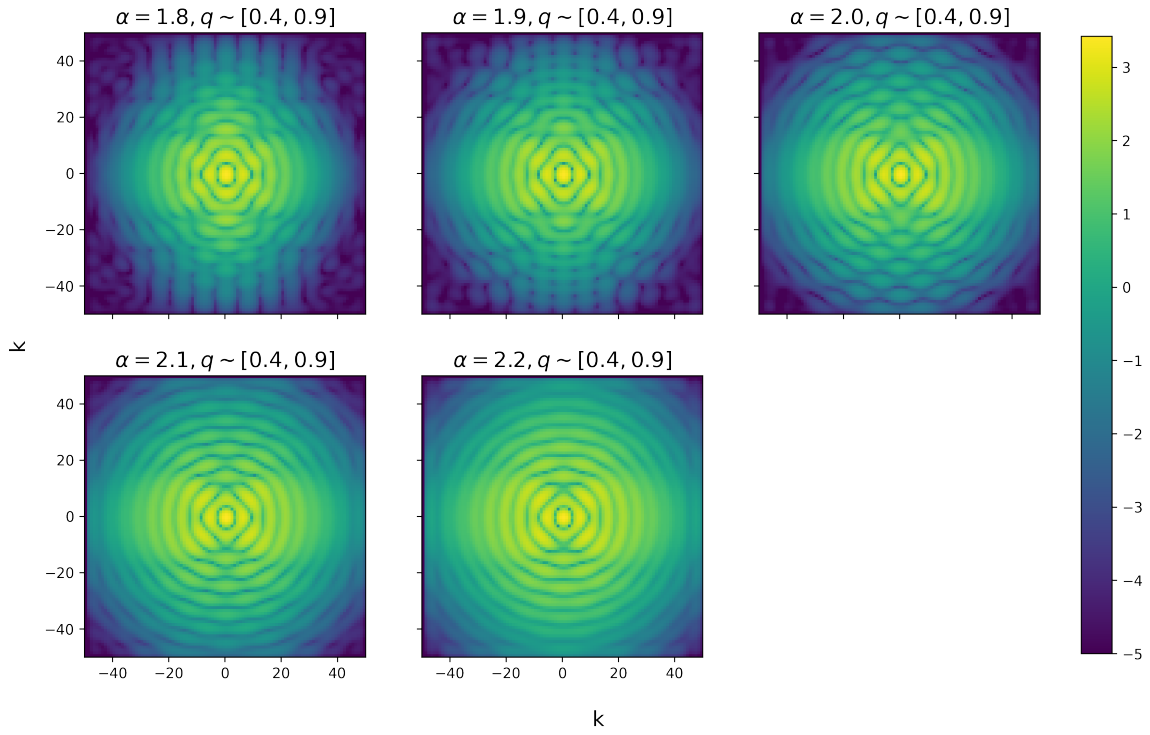


Figure 5.10: Variance of Fourier transformed true source surface brightness as a function of axis ratio, for a range of power law index values. The source is on the optic axis and the images have been created at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

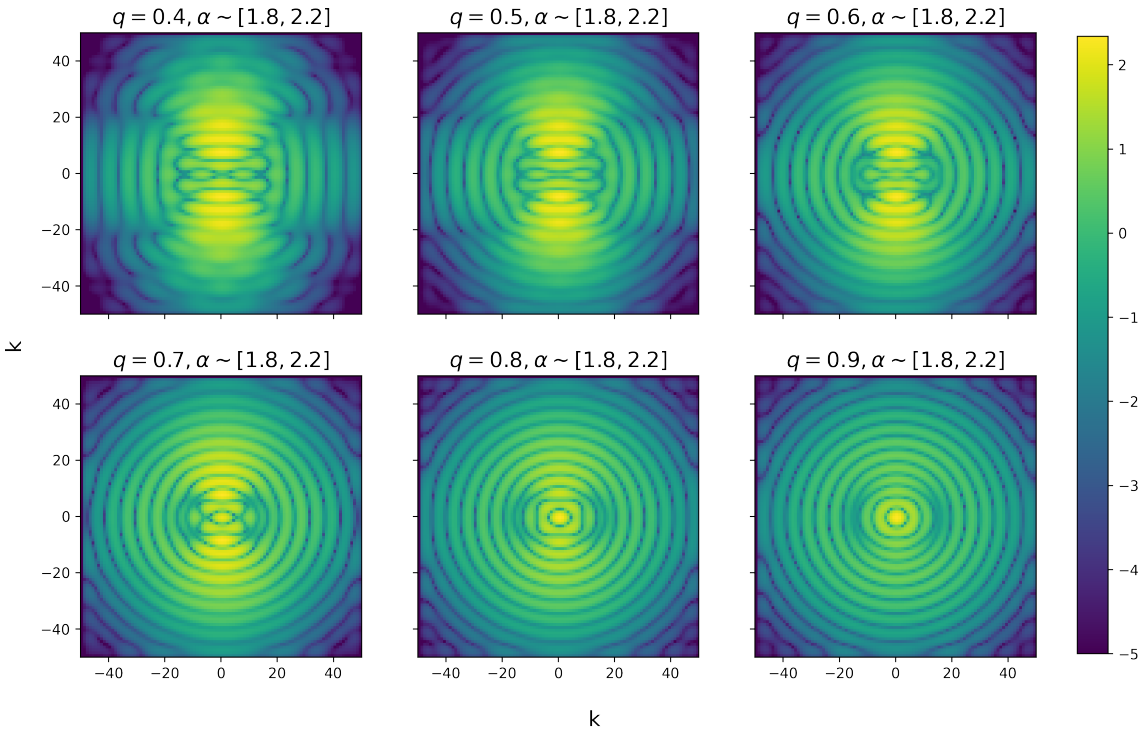


Figure 5.11: Variance of Fourier transformed true source surface brightness as a function of power law index, for a range of axis ratio values. The source is on the optic axis and the images have been created at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

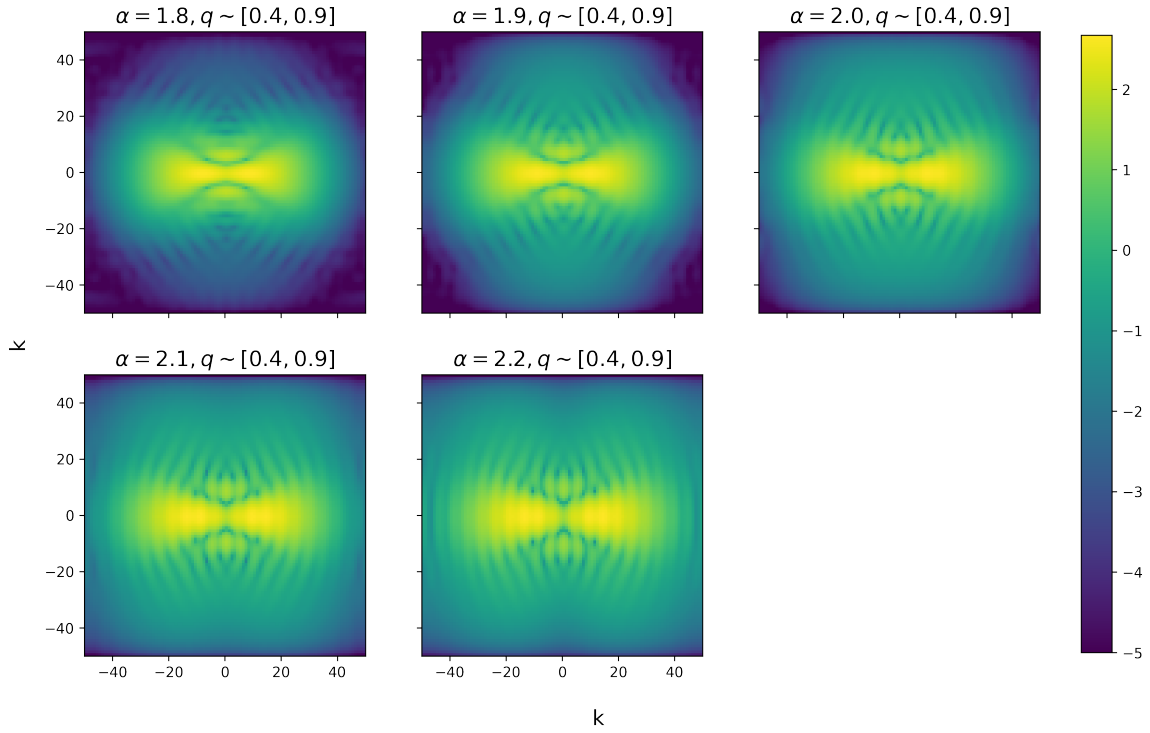


Figure 5.12: Variance of Fourier transformed true source surface brightness as a function of axis ratio, for a range of power law index values. The source is placed at (0.7", 0.0") relative to the centre of the lens and the images have been created at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

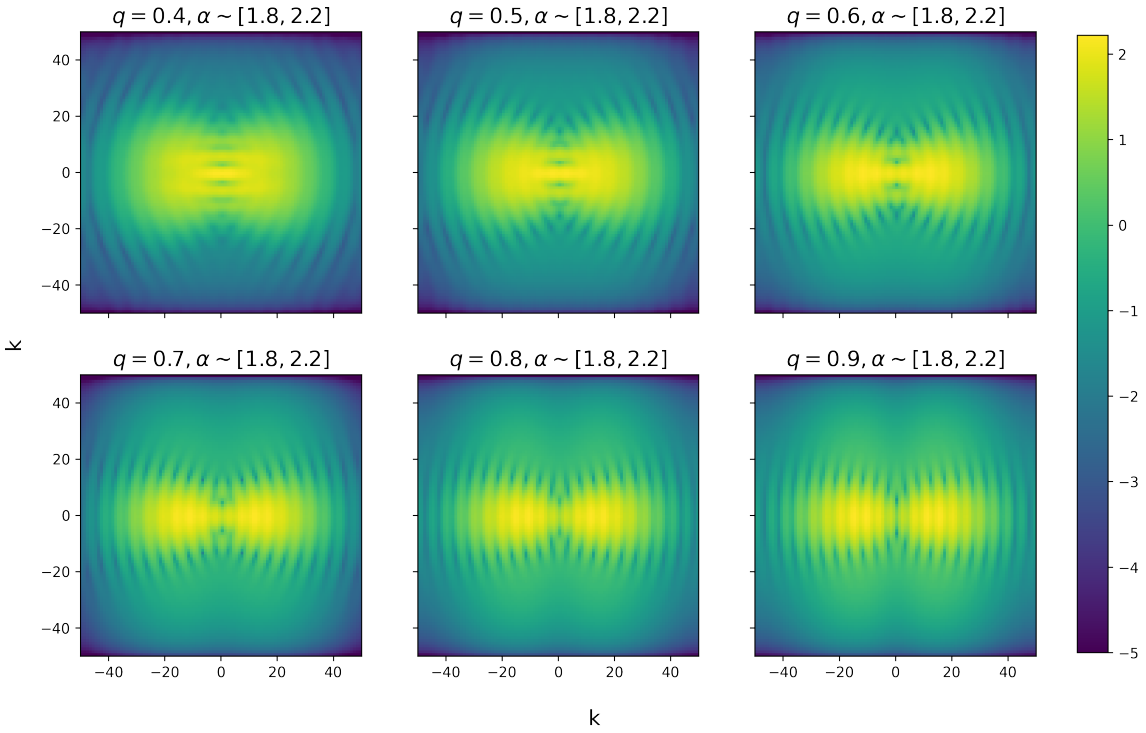


Figure 5.13: Variance of Fourier transformed true source surface brightness as a function of power law index, for a range of axis ratio values. The source is placed at (0.7", 0.0") relative to the centre of the lens and the images have been created at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

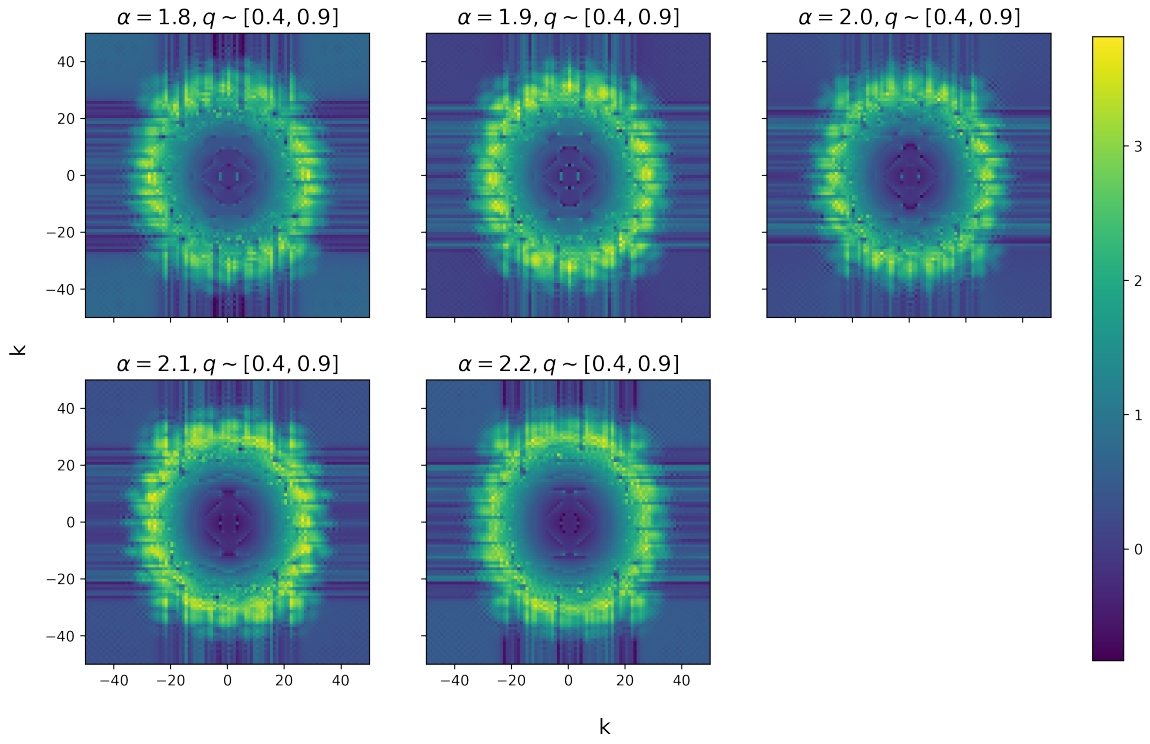


Figure 5.14: Ratios of the variance of Fourier transformed Briggs and natural weighted CLEAN images as a function of axis ratio, for a range of power law indices. The source is on the optic axis and the images have been reconstructed at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

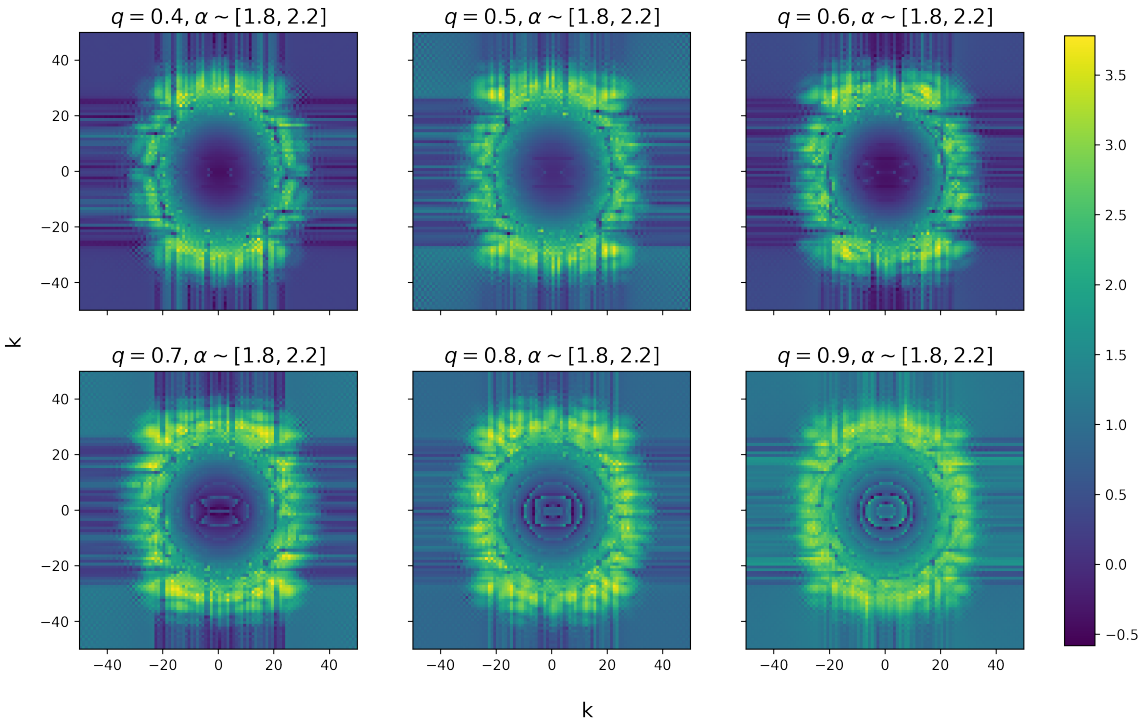


Figure 5.15: Ratios of the variance of Fourier transformed Briggs to natural weighted CLEAN images as a function of power law index, for a range of axis ratios. The source is on the optic axis and the images have been reconstructed at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

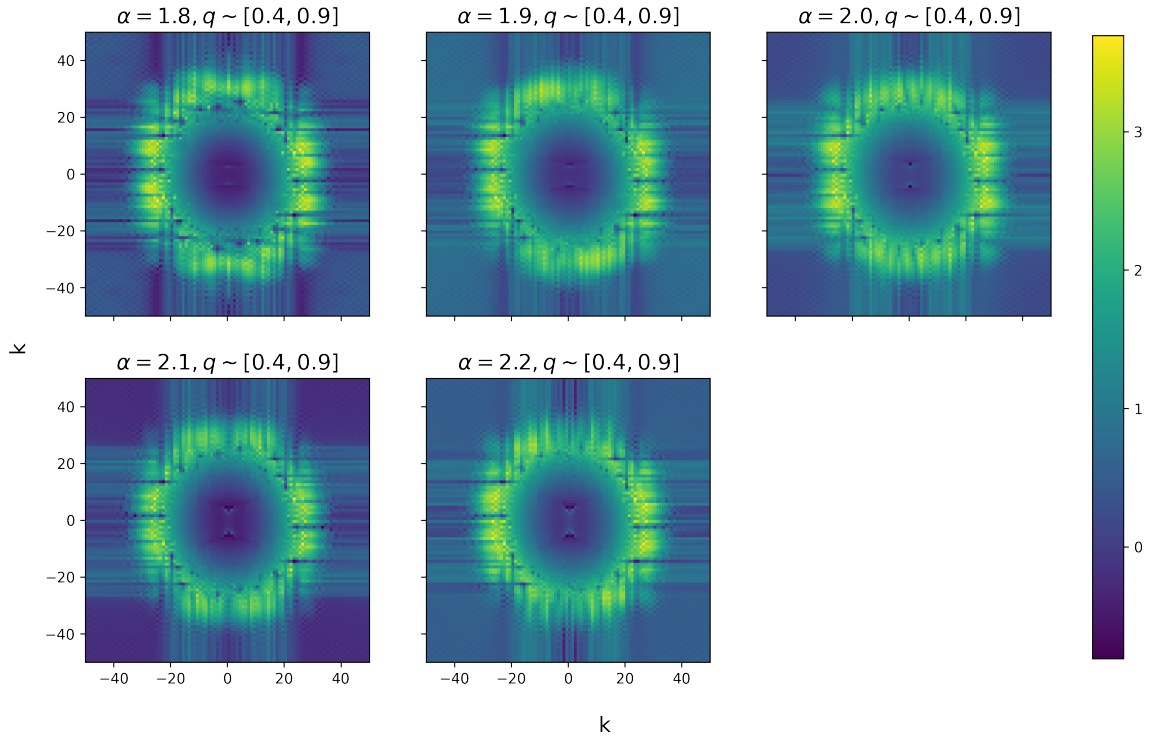


Figure 5.16: Ratios of the variance of Fourier transformed Briggs and natural weighted CLEAN images as a function of axis ratio, for a range of power law indices. The source is placed at $(0.7'', 0.0)$ with respect to the centre of the lens, and the images have been reconstructed at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

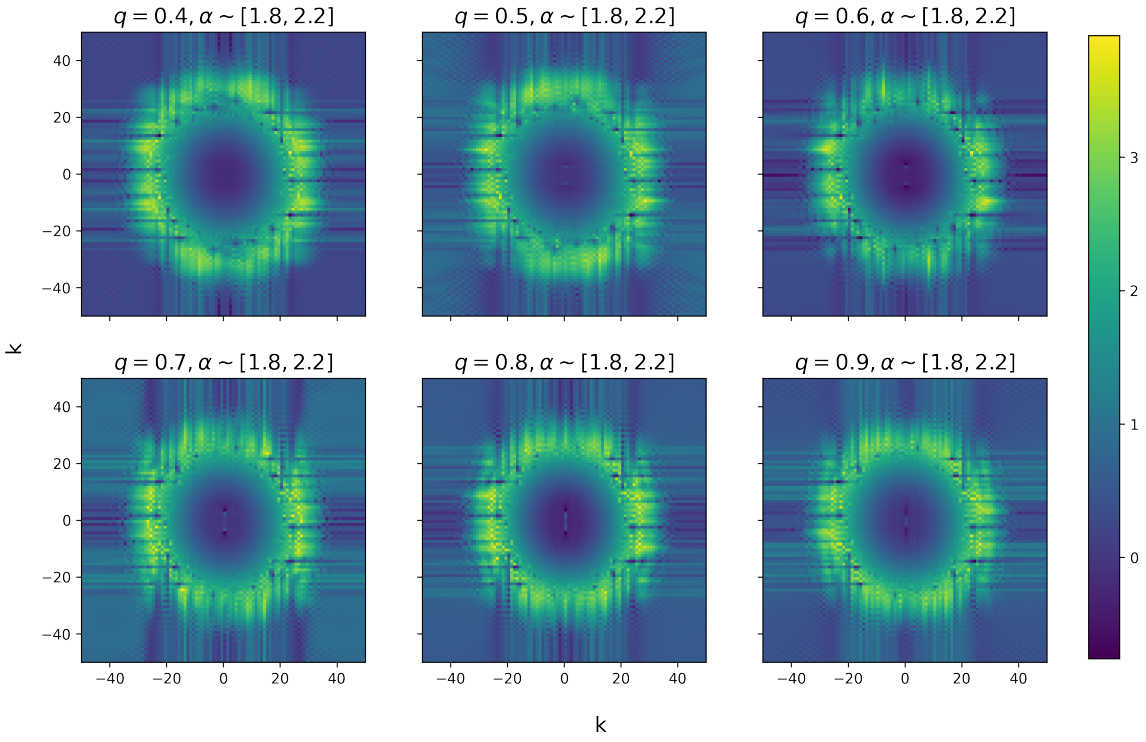


Figure 5.17: Ratios of the variance of Fourier transformed Briggs and natural weighted CLEAN images as a function of power law index, for a range of axis ratios. The source is placed at $(0.7'', 0.0)$ with respect to the centre of the lens, and the images have been reconstructed at a resolution of 0.1 arcsec pixel^{-1} to match that of the CLEANed images for the ALMA cycle 7.3 configuration.

to the nature of the weighting schemes, it is expected that the naturally weighted images will contain less information on the smallest scales compared to the Briggs weighted images, as discussed in section 1.6.2. Figures 5.14 and 5.15 show the ratio of the variance of the Fourier transformed Briggs and naturally weighted CLEAN images, as a function of axis ratio and power law index, respectively, for the case where the source is on the optic axis. It is plain to see from the annular structure in these figures that the Briggs weighted CLEAN images are capturing information at higher k-numbers than the naturally weighted images, in regions where the true surface brightness distributions show significant variation, as shown in Figures 5.10 and 5.11. This increased angular resolution in the Briggs weighted images likely plays a key role in the ability to accurately recover the lens model parameters, more so than the better point source sensitivity of the naturally weighted images. This effect becomes even more important when considering cases where the source is significantly displaced from the optic axis, such as in Figures 5.12 and 5.13. Due to the strong azimuthal dependence of the variance in these images, a greater proportion of the signal is contained at high k-values. As shown by Figures 5.16 and 5.17, it is in these high k-regions that the Briggs weighted images are able to capture more of the signal. In addition to this, there is also less information within the Einstein ring to constrain the lens model at large source offsets, which can partly explain the difficulty in recovering the true lens model parameters.

The difference in lens modelling accuracy between the Briggs and naturally weighted image datasets can, at least in part, be explained by the different angular resolutions achieved by the two weighting schemes. In order to quantify this effect, we analyse the azimuthally averaged two-dimensional Fourier transforms of our true source distributions and the associated images produced by `tclean` from the simulated ALMA observations. Hereafter, we refer to this azimuthally averaged

two-dimensional Fourier transform as a power spectrum, but acknowledge that technically speaking, it is not.

Inspecting the power spectra of the simulated lensed images reveals a bimodal split in the data as a function of source position offset. As can be seen in Figure 5.18, the power spectra fall into two camps. For small source offsets ($x \leq 0.3$) the power spectra for each simulated image is virtually identical, whilst for source offsets greater than $x \geq 0.4$ we begin to see an ever-increasing power at large k numbers. This divide occurs at approximately the point where any source emission connecting the multiple images of the background source is too faint to be of consequence, and thus the image becomes dominated by smaller structure. From this, it is natural to suspect that the lower angular resolution achieved by naturally weighted images is to blame for the relatively poorer performance in recovery of lens model parameters at larger source offsets.

In Figure 5.19 we compare the power spectra of the underlying simulated image with that of the images produced with the `tclean` algorithm using both Briggs and natural weighting, for the two extremes of source offset in the SIE cusp configuration. With the source on the optical axis, the naturally weighted image is underestimating the power for $10 \leq k \leq 40$, whereas the Briggs weighted image is overestimating the power in the same region. As we move the source to its extreme offset, the power spectrum of the underlying simulation has a much shallower decline at high k-numbers. This serves to dramatically increase the amount by which the naturally weighted image is underestimating the power at these high frequencies. In contrast, the Briggs weighted image is capable of reproducing this higher frequency structure out to larger values of k, but ultimately does underestimate the power at the highest frequencies.

We define the quantity of excess power ϵ, as the difference between the total power

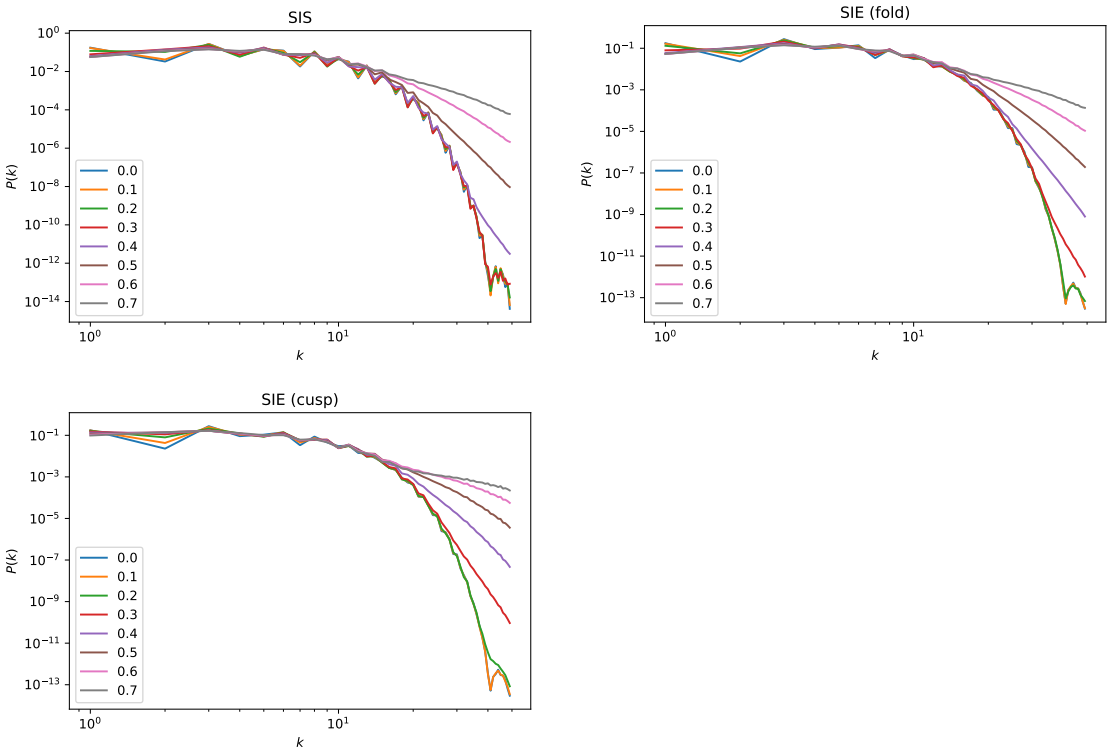


Figure 5.18: Power spectra of the underlying simulated lensed images and a selection of CLEAN images. The top left panel shows the power spectra for each of the source positions considered with the SIS lens model. The top right panel shows the power spectrum for each of the source positions moving through the fold configuration of the SIE lens model. The bottom left panel shows the power spectrum for each of the source positions moving through the cusp configuration of the SIE lens model. The legends in each plot indicate the distance between the source and lens centroid in arcseconds.

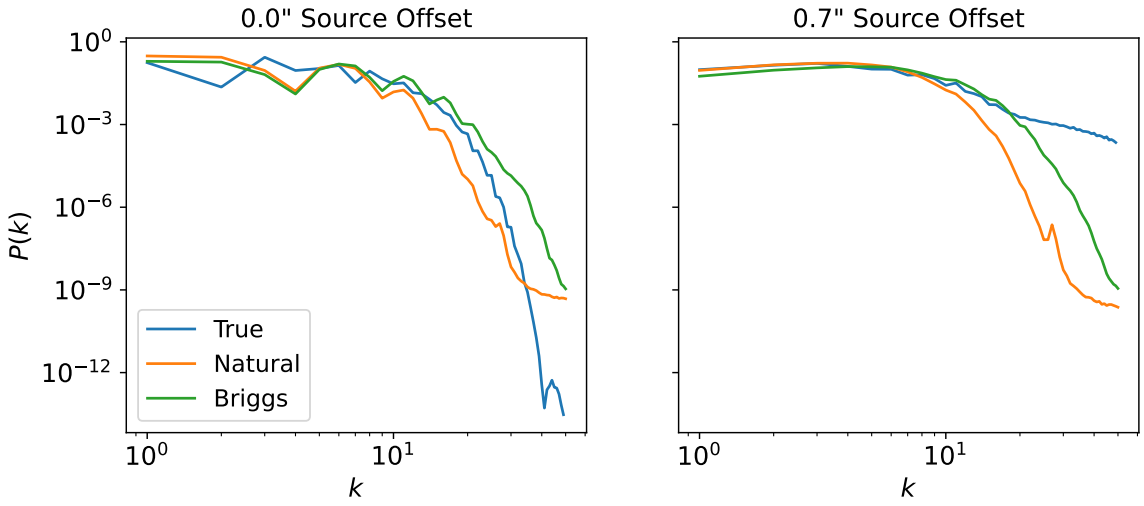


Figure 5.19: A comparison between the power spectra of the simulated image, and the images created with Briggs and natural weighting for two extremes of source offset position in the SIE cusp configuration.

of the underlying simulation and that of the image produced with `tclean` in question, above a threshold value of $k = 10$,

$$\epsilon = \int_{10}^{50} P_T(k)dk - \int_{10}^{50} P_C(k)dk, \tag{5.2}$$

where P_T and P_C are the power spectra of the simulated image and `tclean` image respectively. The threshold value of $k = 10$ was chosen as a reasonable approximation to where the power spectra in all cases begin to diverge significantly from the true underlying power spectra. We then use this value of ϵ to explore the biases in our lens modelling results as both a function of source offset and excess power.

The fractional error on the key SIS lens model parameters as a function of excess power, for the 6000s integration time ALMA observations are shown in Figure 5.20. There are several trends in this data that have explanatory capabilities for the behaviour we are seeing in the modelling results. Notably, in the case of the naturally weighted images, the negative excess power is neatly correlated with the source offset, with more negative values of ϵ corresponding to larger values of source offset. Particularly in the cases of Einstein radius and axis ratio, these more negative values of excess power result in larger absolute values of fractional error. This is suggestive that the naturally weighted images are more susceptible to the degeneracy between θ_E and q. The relationship between fractional error in α and ϵ is strikingly different for the naturally weighted and Briggs images, displaying non-linear behaviour in the case of the former. In general, the range of values of excess power occupied by the naturally and Briggs weighted images are comparable, but the correlation with fractional errors on the lens parameters is much weaker for the Briggs weighted images.

The fractional error of the SIE lens model parameters (fold configuration) as a function of excess power, for the 6000s integration time ALMA observations, are

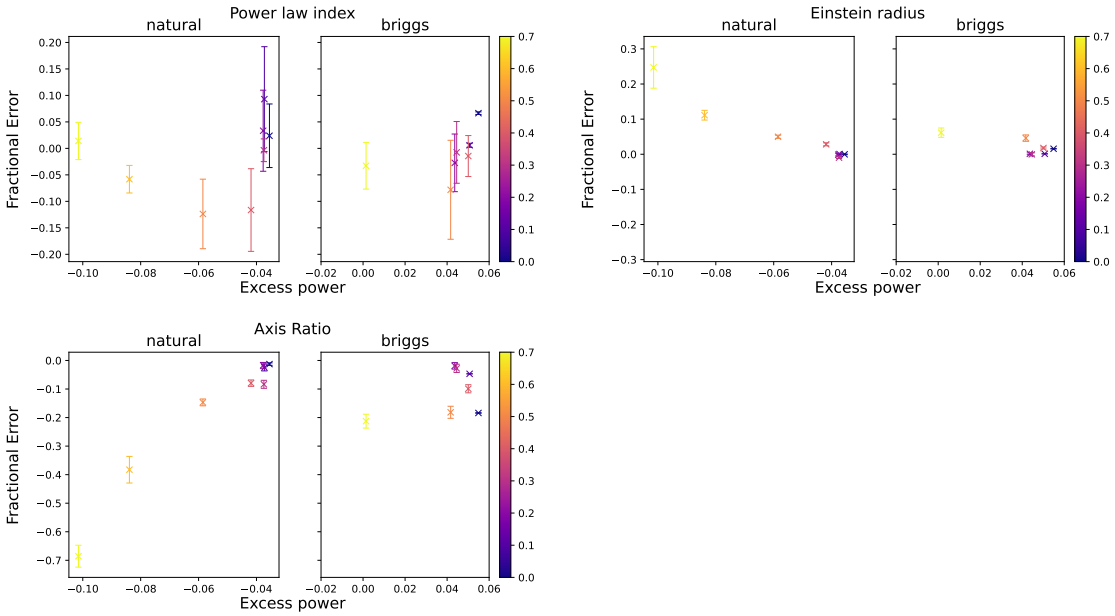


Figure 5.20: Fractional error in SIS lens model parameters as a function of the excess power, as defined in Equation 5.2. Plotted side by side are the same quantities for the naturally and Briggs weighted images created from the 6000s integration time ALMA observations. The colour bar indicates the source offset in arcseconds.

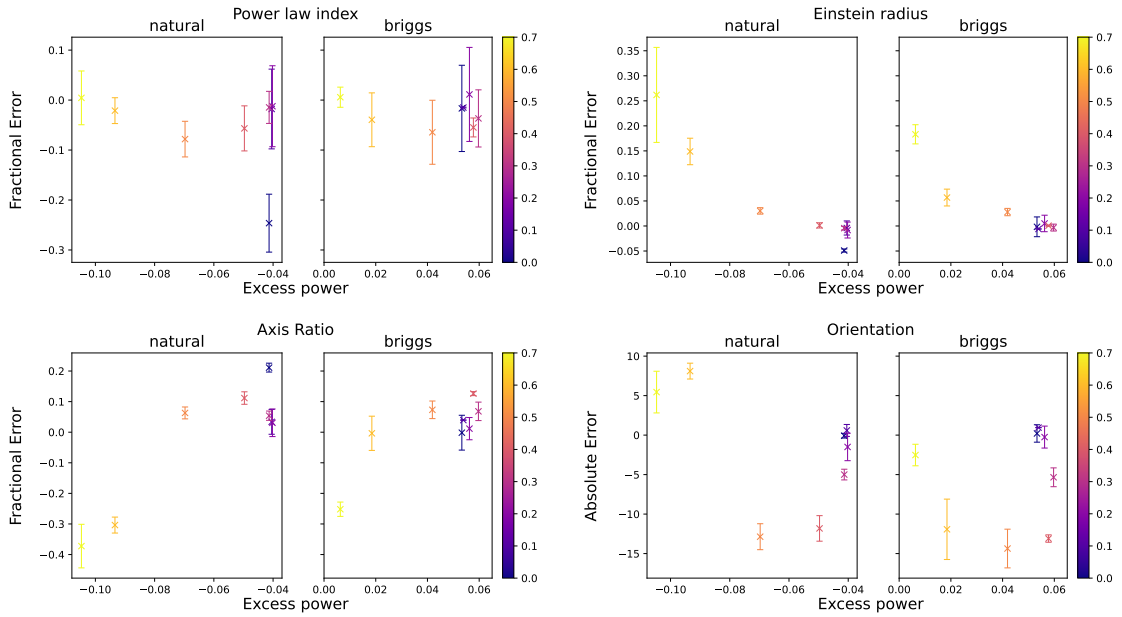


Figure 5.21: Fractional error in SIE lens model (fold configuration) parameters as a function of the excess power, as defined in Equation 5.2. Plotted side by side are the same quantities for the naturally and Briggs weighted images created from the 6000s integration time ALMA observations. The colour bar indicates the source offset in arcseconds.

shown in Figure 5.21. Here, there is no correlation between the excess power and the fractional error in α, with both the naturally and Briggs weighted images performing comparably. For the Einstein radius, in both cases, the excess power is strongly correlated with the bias in the parameter value. The correlation between ϵ and the bias in q is far stronger for the naturally weighted images, compared to the Briggs weighted images.

The fractional error of the SIE lens model parameters (cusp configuration) as a function of excess power, for the 6000s integration time ALMA observations, are shown in Figure 5.22. There is a clear correlation between ϵ and the bias in α for the naturally weighted images when compared with the relationship for the Briggs weighted images. As the negative excess power becomes greater, so does the bias in α. There is no correlation between the excess power and fractional error in Einstein radius for either datasets. Extreme values of negative excess power are associated with a larger bias in q for the naturally weighted images. There is no correlation between ϵ and the fractional error in ϕ.

5.5 Conclusions

In this chapter, we have introduced a set of realistic simulated ALMA observations, produced using the capabilities of `CASA`. From these simulated measurement sets, we have used the `tclean` algorithm to produce images, both with natural weighting and Brigg's weighting. With these data products, we have been able to explore a number of different questions related to lens modelling methodology: How does time-binning the visibility data effect our lens model accuracy? Does the weighting scheme chosen for use in `tclean` introduce biases into our lens model? and, is the extra computational cost of performing lens modelling in the uv-plane worth it?

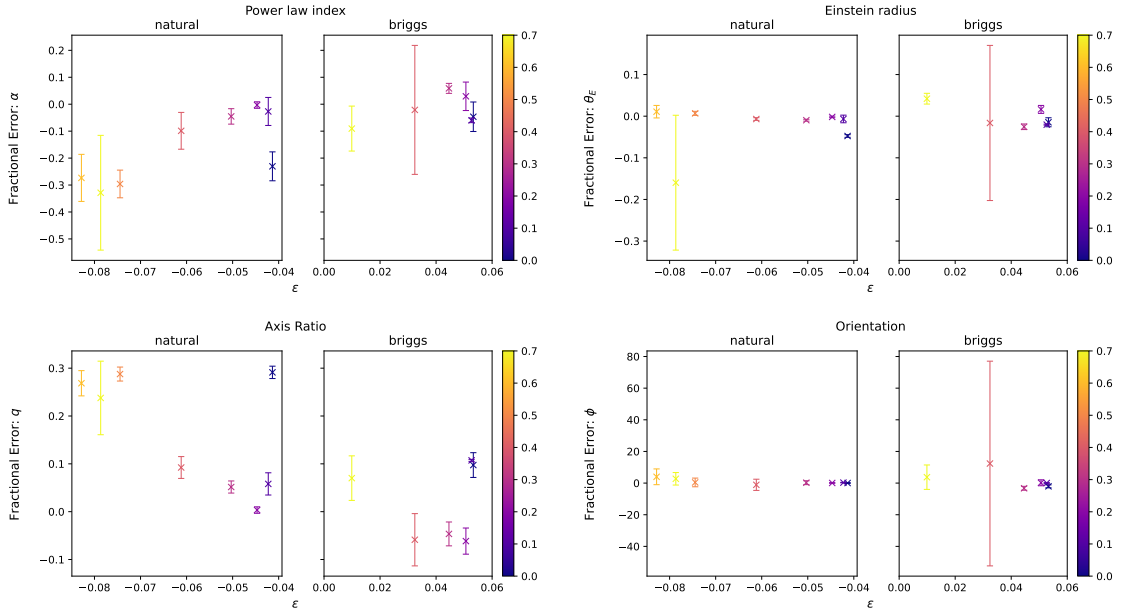


Figure 5.22: Fractional error in SIE lens model (cusp configuration) parameters as a function of the excess power, as defined in Equation 5.2. Plotted side by side are the same quantities for the naturally and Briggs weighted images created from the 6000s integration time ALMA observations. The colour bar indicates the source offset in arcseconds.

We have shown that, at least for the simple parametric source and low angular resolution ALMA setup used here, relatively extreme values of time-binning can be used without negatively impacting our ability to recover the lens model parameters. We used three values of time binning, 10s, 20s, and 30s, and found that the results were all consistent with one another, and so the computational benefits of larger time-bins clearly outweighed any potential cost in modelling accuracy. We have not explored what effect, if any, this might have on our ability to recover intrinsic source properties. These conclusions would likely differ had we considered more complex background sources, or used longer baseline ALMA setups. We leave further exploration of these questions to future work.

Our lens modelling results indicate that the higher resolution images, produced by using Brigg's weighting with `tclean`, lead to less biased results. At large source offsets from the lens centroid, both image weight schemes led to an over-estimation of the Einstein radius, but this effect was more pronounced for the naturally weighted images. Similarly, in the case of the SIS and SIE (fold configuration), we observed a tendency to under-estimate the axis ratio at large source offsets. This behaviour was stronger in the case of the naturally weighted images. We found, in particular, that when the source lay on the optic axis of our SIE lens, our modelling of naturally weighted images led to a 20 per cent underestimation of the power law index. This behaviour was not seen in our Brigg's weighted images.

Generally speaking, the uv-plane modelling outperformed both of our image plane tests in almost every instance. In the case of the SIS, the uv-plane approach exhibited a slight bias towards higher values of α at small source offsets, but was generally consistent. The inferred value of Einstein radius was accurate at all source offsets, and whilst the axis-ratio values were under-predicted, they do not display the same bias as a function of source offset; instead, the level of bias remains constant. In both

the SIE fold and SIE cusp configurations, the uv-plane approach performs remarkably well, showing no clear signs of bias in the recovery of any lens model parameters.

These results indicate that performing lens modelling in the uv-plane is the most robust way to recover the true lens model parameters when working with relatively compact ALMA configuration datasets. Whilst the computational cost of doing so is high, it is justified by its accuracy.

There is a massive amount of scope for exploring the ideas in this work further. We have only considered one compact ALMA configuration (with three different on-source integration times). Understanding how these effects vary as a function of angular resolution is an important question for informing our choices of methodology, as the computational cost of performing visibility modelling only increases with higher resolution data. We used the same source in all of our tests (albeit at different positions), and so could not answer any questions pertaining to whether the effects we see are more pronounced for more compact or extended sources. Additionally, the source we have used was a simple parametric one, and so understanding how these lens modelling biases are effected by more complex source morphologies is an interesting question. Furthermore, we have only considered the impact on our ability to recover the lens model parameters, and have not explored the consequences for determining the intrinsic source properties.

Chapter 6

Conclusions and Future Work

In this final section, we summarise the main findings of this thesis and discuss the prospects for future work. In particular, we will discuss the possibilities of extending the work of Chapter 2 and Chapter 3 to increase the level of automation in our lens modelling pipelines, and the need for further investigation into performing lens modelling on CLEANed images.

6.1 Summary of main results

Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting

We have developed a novel automated technique for the efficient modelling of strongly lensed galaxies. Firstly, we have trained an approximate Bayesian CNN to predict the mass model parameters and their associated uncertainties. These values are then used to initialise lens model priors for `PyAutoLens`, which we use to re-fit the data and produce a reconstructed background source. We explored

two techniques for initialising `PyAutoLens`; one that centres a fixed width uniform prior on the CNN's prediction, and another that uses the CNN's uncertainty on a parameter to initialise a Gaussian prior with a standard deviation equal to that of the CNN's prediction (centred on the CNN predicted value). We find that whilst the CNN is vastly more time-efficient, the combination of the CNN with `PyAutoLens` produces results with greater accuracy than either method alone. Additionally, the incorporation of the CNN's predictions into `PyAutoLens` speeds up the run-times for lens modelling. This work represents a promising step towards a fully automated lens modelling pipeline, which is becoming increasingly important as we move towards the era of LSST and Euclid.

Auto-identification of unphysical source reconstructions

We have shown that the convergence upon unphysical source reconstructions, which can often de-rail the lens modelling process, can be effectively avoided through the use of a CNN. Such unphysical source reconstructions represent valid solutions to the lens equation, and are, as such, difficult to discriminate against when optimising a lens model. The existence of these solutions often leads to the need for human intervention in the lens modelling process, to carefully tune the priors of the lens model, or to manually check the converged models for sensible source reconstruction. Obviously, this is not a process that lends itself to the modelling of the many thousands of strong lenses we expect to discover with upcoming surveys, such as LSST and Euclid. Whilst the techniques of combining `PyAutoLens` with a CNN's predictions can greatly reduce the occurrence of these issues, as discussed in Chapter 2, they can still arise nonetheless. To tackle this problem, we trained a CNN to identify these unphysical source reconstructions, and developed a simple prescription to re-initialise the lens modelling process with new priors in order to converge upon

the true solution. This methodology allows for less human intervention in the lens modelling process, and is thus important for the development of a truly automated lens modelling pipeline.

Modelling high-resolution ALMA observations of strongly lensed dusty star forming galaxies detected by Herschel

We presented modelling of ∼0.1 arcsec resolution ALMA imaging of seven strong gravitationally lensed galaxies detected by the Herschel Space Observatory. Four of these systems were galaxy-galaxy strong lenses, with the remaining three being group-scale lenses. Through careful modelling of visibilities, we inferred the mass profiles of the lensing galaxies and by determining the magnification factors, we investigated the intrinsic properties and morphologies of the lensed sub-millimetre sources. We found that these sub-millimetre sources all have ratios of star formation rate to dust mass that are consistent with, or in excess of, the mean ratio for high-redshift sub-millimetre galaxies and low redshift ultra-luminous infrared galaxies. Reconstructions of the background sources revealed that the majority of our sample display disturbed morphologies. The majority of our lens models have mass density slopes close to isothermal, but some systems show significant differences.

Lens modelling with interferometric datasets

We have shown that the methodological choices that one can make when working with interferometer datasets has a significant impact on the inferred lens model. By producing a series of simulated ALMA observations, and images with different weighting schemes via the `tclean` algorithm, we explored several questions that lens modellers are faced with when it comes to choosing their methodologies. We looked at the impact that time binning of the visibility data had on our recovery of

the true lens model parameters, and found that it did not play a significant role, though we did not consider the impact that this may have on the recovery of the intrinsic source properties. We compared the performance of image-plane modelling applied to two different sets of images, created with either natural weighting or Brigg's weighting (R=0), and found that in general, the increased resolution of the Brigg's weighted images led to more accurate lens models. Last, but certainly not least, we compared the performance of the image-plane modelling to that of the uv-plane method, and found that the uv-plane method performs significantly better, and is less prone to several of the biases we observed in our image-plane results.

6.2 Future improvements

Error handling with CNNs

We argued in Chapter 2 that one of the drawbacks to using CNNs was that their performance was so tightly coupled with our ability to create high quality training data sets. Whilst we showed that the CNN still performed reasonably well when applied to the EAGLE lenses in this work, despite having not been trained on such lenses, it did under-predict its uncertainty values. This had a knock-on effect when used in combination with `PyAutoLens` that the prior distributions used in the parametric modelling were both too narrow and often not properly centred. This likely led to a dramatic reduction in the performance of this combined method. Addressing these issues in the training data would allow us to properly explore the potential of this method when applied to extremely realistic lensing simulations.

Furthermore, whilst state-of-the-art lens finding algorithms are very good, they are not perfect (Lanusse et al., 2017b; Cheng et al., 2020; Petrillo et al., 2017).

Categorising non-lensed objects and data artefacts as lensed galaxies can lead to issues in the creation of a fully automated lens modelling pipeline. We did not explore what the output of the CNN would be if given an object such as a ring galaxy to predict upon. An extension to this work would be to explore how we can mitigate the issue of the CNN failing silently and simply outputting a lens model to be passed to the parametric fitting algorithm.

Auto-identification of Unphysical Source Reconstructions

In Chapter 3 we showed that a CNN could be employed to detect when our sampling algorithms had converged upon solutions corresponding to over/under-magnified source reconstructions. We also developed a simple prescription for re-initialising our lens model priors to avoid these solutions and ultimately converge upon the true solution. Our implementation of this technique was relatively crude, and waited until the stopping criteria of the modelling process had been met before providing a classification of the result. To improve upon this, we could use the CNN to reject/accept samples during the modelling process, by checking the 'intermediate' source reconstructions, to effectively cut down on the amount of time spent exploring these regions of parameter space. Additionally, information could be gathered about these rejected samples, and the relationships between unphysical and physical source reconstructions could be used to increase the density of sampling in the likely region of parameter space.

Lens modelling with Interferometer datasets

There is a huge scope for exploring the work of Chapter 5 further. The lensing scenario we used as the basis of our work was incredibly simple, with a parametric source and either an SIS or SIE mass model acting as the lens. Simply increasing

the complexity of the background source could lead to interesting discovery, since so much of the behaviour we observed seemed to be linked to the resolution achieved in the observations. Also linked to the source, we did not carry out any investigation into how the different methodologies impacted our ability to infer the intrinsic properties of the source. The parameter space for exploring the many ALMA configurations and options when producing images with `tclean` is vast, but some obvious examples of what to do next are: What happens when we use a more extended ALMA configuration? If Brigg's weighting outperforms natural weighting, does that imply the existence of an optimal robustness parameter? How does time-binning effect our results when a more complex source is used?

6.3 Outlook

In the near future, surveys such as LSST and Euclid are set to come online, revolutionising the field of strong gravitational lensing. These two projects are expected to discover many thousands of galaxy-galaxy strong lensing systems (Collett, 2015b), providing statistical power that has been lacking up to this point.

The multi-band imaging of LSST will allow for the determination of photometric redshift estimates, for both the lens and source galaxies. This will allow for measurements of the mass of the lens galaxy enclosed within the Einstein radius, as well as providing accurate aperture mass-to-light ratios for each lens galaxy. Studies of the mean density profile and evolution of early-type galaxies will greatly benefit from the large sample sizes of LSST lenses. Currently, surveys such as SLACS have used relatively small numbers of low redshift, luminous galaxies to place constraints on the mean logarithmic slope of the density profiles of early type galaxies. With the LSST galaxy-galaxy strong lensing sample, this ensemble size can be increased from tens to

many thousands, and dramatically increase both the redshift range and mass range of galaxies included (Marshall et al., 2022). Synergies with other upcoming survey telescopes, such as the Nancy Grace Roman Space Telescope (Spergel, D., 2015), will allow for tighter constraints on the lens profiles of LSST detected strong lenses (Eifler et al., 2019).

Our current sample of strong lenses capable of being used for measuring time-delays, and thus the Hubble-Lemaître constant, is relatively small. This sample contains many lenses that are difficult to model, such as having multiple lens galaxies, or being contained within a cluster potential. LSST will provide a large sample of time-delay lenses, and thus allow for the selection of a subsample of lenses that do not suffer from these issues. It is expected that LSST will observe 2600 lensed quasars and 330 lensed supernovae, with 90 of these being Type Ia (Marshall et al., 2022). After selections for environment and complexity have been made, it is reasonable to expect a sample of approximately 100 high quality time delay lenses, allowing for an order of magnitude increase in measurement precision over our current sample. It is known that characterisation of the external convergence is the dominant source of uncertainty in determining time-delay measurements, and so significant effort will need to be put into managing any systematics introduced by the adopted treatments (LSST Dark Energy Science Collaboration, 2012)

Astronomy is undergoing a revolution in the scale of the data products being produced, from projects as diverse as Gaia (Gaia Collaboration et al., 2016), which will provide accurate positions and radial-velocities for more than a billion stars in our galaxy, to the Square Kilometer Array (SKA; Dewdney et al., 2009), which will be capable of mapping the primordial distribution of gas in the universe. In light of the unprecedented challenges that this volume of data will present, astronomers are developing many new tools to rise to the occasion, with machine learning algorithms

being very prevalent among them. Machine learning algorithms are typically divided into two categories; supervised and unsupervised. The former requires labelled data for the algorithm to learn a mapping from the input to the output, whilst the latter is able to learn relationships that exist in the data, without the need for labelling. These ML algorithms can be used to great effect to efficiently deal with huge amounts of data and perform traditionally CPU intensive tasks, such as lens modelling, measuring galaxy morphology and predicting photometric redshifts. One particular challenge faced by these algorithms is their specificity to the data they were trained upon. Typically, when training a convolutional neural network, for example, one would use many images with the characteristics of the survey we wish to apply our network to. Such networks fail to generalise to other surveys or imaging characteristics, which is of particular importance when we wish to use them to search for rare astrophysical objects. In the case of finding strong gravitational lenses, we have so few observations that we must rely on simulated data to train the algorithms, but this can pose a challenge when applying the tool to real data. One such method for dealing with this issue, is transfer learning, where knowledge gained in the solution of one problem can be leveraged by the same network to solve another (Zhuang et al., 2019). The application of such techniques will be vital to properly exploiting the wealth of information contained within upcoming datasets, such as achieving a high purity and completeness in the search for strong gravitational lenses in the LSST sample (LSST Dark Energy Science Collaboration, 2012).

Bibliography

ALMA Partnership et al., 2015, ApJ, 808, L4

Abolfathi B., et al., 2018, ApJS, 235, 42

Alaghband-Zadeh S., et al., 2012, MNRAS, 424, 2232

Amvrosiadis A., et al., 2018, Monthly Notices of the Royal Astronomical Society, 475, 4939

An F. X., et al., 2019, ApJ, 886, 48

Asboth V., et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 1989

Assef R. J., et al., 2011, ApJ, 728, 56

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511

Autonomio 2019, autonomio/talos, `http://www.github.com/autonomio/talos`

Bakx T. J. L. C., et al., 2017, Monthly Notices of the Royal Astronomical Society, 473, 1751

Baldry I. K., et al., 2012, MNRAS, 421, 621

Barkana R., 1998, The Astrophysical Journal, 502, 531

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125

Beckwith S. V., et al., 2006a, The Astronomical Journal, 132, 1729

Beckwith S. V. W., et al., 2006b, AJ, 132, 1729

Bekki K., Couch W. J., Shioya Y., Vazdekis A., 2005, Monthly Notices of the Royal Astronomical Society, 359, 949

Birrer S., Amara A., 2018, Physics of the Dark Universe, 22, 189

Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102

Birrer S., et al., 2020, A&A, 643, A165

Blain A. W., 1996, MNRAS, 283, 1340

Blain A. W., Smail I., Ivison R., Kneib J.-P., Frayer D. T., 2002, Physics Reports, 369, 111

Blandford R., Narayan R., 1986, ApJ, 310, 568

Bolton A. S., Burles S., Koopmans L. V., Treu T., Moustakas L. A., 2006, The Astrophysical Journal, 638, 703

Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Gavazzi R., Moustakas L. A., Wayth R., Schlegel D. J., 2008, ApJ, 682, 964

Bolton A. S., et al., 2012, ApJ, 757, 82

Bournaud F., et al., 2014, ApJ, 780, 57

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Brammer G. B., et al., 2009, The Astrophysical Journal, 706, L173

Brewer B. J., et al., 2012, MNRAS, 422, 3574

Briggs D. S., 1995, PhD thesis, New Mexico Institute of Mining and Technology, United States

Brownstein J. R., et al., 2011, The Astrophysical Journal, 744, 41

Bussmann R. S., et al., 2012, ApJ, 756, 134

Bussmann R. S., et al., 2013, The Astrophysical Journal, 779, 25

Bussmann R. S., et al., 2015, ApJ, 812, 43

Cañameras R., et al., 2017a, A&A, 600, L3

Cañameras R., et al., 2017b, A&A, 604, A117

Cabanac R. A., et al., 2007, Astronomy & Astrophysics, 461, 813

Calanog J. A., et al., 2014, The Astrophysical Journal, 797, 138

Cappellari M., et al., 2013, MNRAS, 432, 1862

Carlstrom J. E., et al., 2011, PASP, 123, 568

Carollo C. M., et al., 2013, ApJ, 773, 112

Casey C. M., Narayanan D., Cooray A., 2014, Physics Reports, 541, 45

Cava A., Schaerer D., Richard J., Pérez-González P. G., Dessauges-Zavadsky M., Mayer L., Tamburello V., 2018, Nature Astronomy, 2, 76

Chapman S. C., Windhorst R., Odewahn S., Yan H., Conselice C., 2003, ApJ, 599, 92

Chapman S. C., Smail I., Windhorst R., Muxlow T., Ivison R. J., 2004, The Astrophysical Journal, 611, 732

Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, ApJ, 622, 772

Cheng T.-Y., Li N., Conselice C. J., Aragón-Salamanca A., Dye S., Metcalf R. B., 2020, Monthly Notices of the Royal Astronomical Society, 494, 3750

Chomiuk L., Povich M. S., 2011, AJ, 142, 197

Clements D. L., et al., 2017, Monthly Notices of the Royal Astronomical Society, 475, 2097

Collett T. E., 2015a, The Astrophysical Journal, 811, 20

Collett T. E., 2015b, The Astrophysical Journal, 811, 20

Couchman H. M. P., Rees M. J., 1986, MNRAS, 221, 53

Courbin F., Saha P., Schechter P., 2008, Quasar Lensing. Springer-Verlag, pp 1–54, doi:10.1007/3-540-45857-3_1

Crain R. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1937

Cropper M., et al., 2016, in MacEwen H. A., Fazio G. G., Lystrup M., Batalha N., Siegler N., Tong E. C., eds, Vol. 9904, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave. SPIE, pp 269 – 284, doi:10.1117/12.2234739, https://doi.org/10.1117/12.2234739

Davé R., Finlator K., Oppenheimer B. D., Fardal M., Katz N., Kereš D., Weinberg D. H., 2010, MNRAS, 404, 1355

Davies A., Serjeant S., Bromley J. M., 2019, Monthly Notices of the Royal Astronomical Society, 487, 5263

Dekel A., Birnboim Y., 2006, Monthly Notices of the Royal Astronomical Society, 368, 2

Dekel A., Silk J., 1986, ApJ, 303, 39

Dekel A., et al., 2009a, Nature, 457, 451

Dekel A., Sari R., Ceverino D., 2009b, The Astrophysical Journal, 703, 785

Dessauges-Zavadsky M., et al., 2019, Nature Astronomy, 3, 1115

Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proceedings, 97, 1482

Dozat T., 2015, In Proc. ICLR Workshop

Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297

Dudzevičiūtė U., et al., 2020, Monthly Notices of the Royal Astronomical Society, 494, 3828

Dunne L., Eales S., Edmunds M., Ivison R., Alexander P., Clements D. L., 2000, Monthly Notices of the Royal Astronomical Society, 315, 115

Dunne L., et al., 2011, Monthly Notices of the Royal Astronomical Society, 417, 1510

Dye S., et al., 2015, Monthly Notices of the Royal Astronomical Society, 452, 2258

Dye S., et al., 2018, Monthly Notices of the Royal Astronomical Society, 476, 4383

Dye S., et al., 2022a, Monthly Notices of the Royal Astronomical Society

Dye S., et al., 2022b, MNRAS, 510, 3734

Eales S., et al., 2010, PASP, 122, 499

Eales S. A., et al., 2018, MNRAS, 481, 1183

Eifler T., et al., 2019, BAAS, 51, 418

Falco E. E., Gorenstein M. V., Shapiro I. I., 1985, ApJ, 289, L1

Feroz F., Hobson M. P., Bridges M., 2009, Monthly Notices of the Royal Astronomical Society, 398, 1601

Fixsen D. J., 2009, ApJ, 707, 916

Fixsen D. J., Dwek E., Mather J. C., Bennett C. L., Shafer R. A., 1998, ApJ, 508, 123

Fukushima K., 1980, Biological Cybernetics, 36, 193

Gaia Collaboration et al., 2016, A&A, 595, A1

Gal Y., Ghahramani Z., 2016a, in 4th International Conference on Learning Representations (ICLR) workshop track.

Gal Y., Ghahramani Z., 2016b, in Proceedings of the 33rd International Conference on Machine Learning (ICML-16). pp 1050–1059

Gavazzi R., Treu T., Marshall P. J., Brault F., Ruff A., 2012, ApJ, 761, 170

Geach J. E., et al., 2013, MNRAS, 432, 53

Geach J. E., et al., 2017, MNRAS, 465, 1789

Glorot X., Bengio Y., 2010, in Teh Y. W., Titterington M., eds, Proceedings of Machine Learning Research Vol. 9, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. PMLR, Chia Laguna Resort, Sardinia, Italy, pp 249–256, https://proceedings.mlr.press/v9/glorot10a.html

Goodfellow I., Bengio Y., Courville A., 2016, Deep Learning. MIT Press

Griffin M. J., et al., 2010, A&A, 518, L3

Grogin N. A., et al., 2011, ApJS, 197, 35

Guo Q., White S. D. M., 2008, Monthly Notices of the Royal Astronomical Society, 384, 2

Guth A. H., 1981, Phys. Rev. D, 23, 347

Harrington K. C., et al., 2021, ApJ, 908, 95

Hartigan J. A., Wong M. A., 1979, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, 100

Hauser M. G., Dwek E., 2001, Annual Review of Astronomy and Astrophysics, 39, 249

He Z., et al., 2020, Monthly Notices of the Royal Astronomical Society, 497, 556

Heitmann K., et al., 2019, The Astrophysical Journal Supplement Series, 245, 16

Hezaveh Y. D., Levasseur L. P., Marshall P. J., 2017a, Nature, 548, 555

Hezaveh Y. D., Perreault Levasseur L., Marshall P. J., 2017b, Nature, 548, 555

Hill R., Masui K. W., Scott D., 2018, Applied Spectroscopy, 72, 663

Högbom J. A., 1974, A&AS, 15, 417

Holland W. S., et al., 1999, Monthly Notices of the Royal Astronomical Society, 303, 659

Holwerda B. W., et al., 2015, MNRAS, 449, 4277

Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142

Hopkins P. F., Hernquist L., Cox T. J., Matteo T. D., Robertson B., Springel V., 2006a, The Astrophysical Journal Supplement Series, 163, 1

Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Springel V., 2006b, The Astrophysical Journal Supplement Series, 163, 50

Hsu L.-Y., Desai V., Murphy E. J., Cowie L. L., Heywood I., Momjian E., Barger A. J., Smail I., 2017, The Astrophysical Journal, 840, 29

Hu W., White M., 1997, in Microwave Background Anisotropies. pp 333–338 (arXiv:astro-ph/9606140)

Huang X., et al., 2021, The Astrophysical Journal, 909, 27

Hubble E. P., 1927, The Observatory, 50, 276

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Huertas-Company M., et al., 2013, MNRAS, 428, 1715

Hughes D. H., et al., 1998, Nature, 394, 241

Hughes T. M., et al., 2017, Monthly Notices of the Royal Astronomical Society: Letters, 468, L103

Huynh M. T., et al., 2017, MNRAS, 467, 1222

Ilbert O., et al., 2013, A&A, 556, A55

Inoue K. T., Minezaki T., Matsushita S., Chiba M., 2016, Monthly Notices of the Royal Astronomical Society, 457, 2936

Ivezic Z., et al., 2008, Serbian Astronomical Journal, 176, 1

Ivison R. J., et al., 2013, ApJ, 772, 137

Jacobs C., Glazebrook K., Collett T., More A., McCarthy C., 2017, Monthly Notices of the Royal Astronomical Society, 471, 167

Jacobs C., et al., 2019, The astrophysical journal supplement series, 243, 17

James A., Dunne L., Eales S., Edmunds M. G., 2002, Monthly Notices of the Royal Astronomical Society, 335, 753

Kaasinen M., et al., 2019, ApJ, 880, 15

Kartaltepe J. S., et al., 2012, ApJ, 757, 23

Keeton C. R., 2001, arXiv preprint astro-ph/0102341

Kendall A., Gal Y., 2017, in Advances in neural information processing systems. pp 5574–5584

Kennicutt Robert C. J., 1998, ARA&A, 36, 189

Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Kereš D., Katz N., Davé R., Fardal M., Weinberg D. H., 2009, MNRAS, 396, 2332

Knabel S., et al., 2020, The Astronomical Journal, 160, 223

Koopmans L. V. E., Treu T., 2003, The Astrophysical Journal, 583, 606

Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006a, ApJ, 649, 599

Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006b, ApJ, 649, 599

Koopmans L. V. E., et al., 2009, ApJ, 703, L51

Kormann R., Schneider P., Bartelmann M., 1994, Astronomy and Astrophysics, 284, 285

Korytov D., et al., 2019, The Astrophysical Journal Supplement Series, 245, 26

Kravtsov A. V., Borgani S., 2012, Annual Review of Astronomy and Astrophysics, 50, 353

Kroupa P., 2001, MNRAS, 322, 231

LSST Dark Energy Science Collaboration 2012, Large Synoptic Survey Telescope: Dark Energy Science Collaboration, doi:10.48550/ARXIV.1211.0310, `https://arxiv.org/abs/1211.0310`

Lagattuta D. J., et al., 2010, ApJ, 716, 1579

Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Póczos B., 2017a, Monthly Notices of the Royal Astronomical Society, 473, 3895

Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Póczos B., 2017b, Monthly Notices of the Royal Astronomical Society, 473, 3895

Lapi A., et al., 2011, ApJ, 742, 24

Lapi A., Negrello M., González-Nuevo J., Cai Z. Y., De Zotti G., Danese L., 2012, ApJ, 755, 46

Lapi A., et al., 2018, ApJ, 857, 22

Larson R. B., 1974, Monthly Notices of the Royal Astronomical Society, 169, 229

Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692

Laureijs R., et al., 2011a, arXiv preprint arXiv:1110.3193

Laureijs R., et al., 2011b, arXiv

Lee B., et al., 2013, ApJ, 774, 47

Levasseur L. P., Hezaveh Y. D., Wechsler R. H., 2017, The Astrophysical Journal Letters, 850, L7

Li C., Lee C., 1993, Pattern Recognition, 26, 617

Li C., White S. D. M., 2009, MNRAS, 398, 2177

Li N., Gladders M. D., Rangel E. M., Florian M. K., Bleem L. E., Heitmann K., Habib S., Fasel P., 2016a, The Astrophysical Journal, 828, 54

Li N., Gladders M. D., Rangel E. M., Florian M. K., Bleem L. E., Heitmann K., Habib S., Fasel P., 2016b, The Astrophysical Journal, 828, 54

Linde A. D., 1982, Physics Letters B, 108, 389

Loeb A., Barkana R., 2001, ARA&A, 39, 19

Mackenzie T., et al., 2011, Monthly Notices of the Royal Astronomical Society, 415, 1950

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Maresca J., Dye S., Li N., 2020, arXiv preprint arXiv:2012.04665

Marinoni C., Hudson M. J., Giuricin G., 2002, The Astrophysical Journal, 569, 91

Marshall P., et al., 2022, The LSST Science Book, `https://www.lsst.org/sites/default/files/docs/sciencebook/SB_12.pdf`

Martig M., Bournaud F., Teyssier R., Dekel A., 2009, The Astrophysical Journal, 707, 250

Matteo T. D., Springel V., Hernquist L., 2005, Nature, 433, 604

McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, Astronomical Society of the Pacific Conference Series Vol. 376, Astronomical Data Analysis Software and Systems XVI. p. 127

Metcalf R. B., et al., 2019, Astronomy & Astrophysics, 625, A119

More A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 455, 1191–

Morningstar W. R., Hezaveh Y. D., Levasseur L. P., Blandford R. D., Marshall P. J., Putzky P., Wechsler R. H., 2018, arXiv preprint arXiv:1808.00011

Morningstar W. R., et al., 2019, ApJ, 883, 14

Mukhanov V. F., Chibisov G. V., 1981, ZhETF Pisma Redaktsiiu, 33, 549

Mukherjee S., et al., 2018, Monthly Notices of the Royal Astronomical Society, 479, 4108

NASA & ESA 2013, Hubble galaxy classification, https://sci.esa.int/web/hubble/-/52791-the-hubble-tuning-fork-classification-of-galaxies

NASA & WMAP Science Team 2006, Timeline of the Universe, https://map.gsfc.nasa.gov/media/060915/index.html

Nair V., Hinton G. E., 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML'10. Omnipress, Madison, WI, USA, p. 807–814

Narayanan D., Hayward C. C., Cox T. J., Hernquist L., Jonsson P., Younger J. D., Groves B., 2010a, MNRAS, 401, 1613

Narayanan D., et al., 2010b, MNRAS, 407, 1701

Narayanan D., et al., 2015, Nature, 525, 496

Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563

Nayyeri H., et al., 2016, The Astrophysical Journal, 823, 17

Negrello M., Perrotta F., González-Nuevo J., Silva L., de Zotti G., Granato G. L., Baccigalupi C., Danese L., 2007, MNRAS, 377, 1557

Negrello M., et al., 2010, Science, 330, 800

Negrello M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 465, 3558

Newman A. B., Ellis R. S., Bundy K., Treu T., 2012, ApJ, 746, 162

Newton E. R., Marshall P. J., Treu T., Auger M. W., Gavazzi R., Bolton A. S., Koopmans L. V. E., Moustakas L. A., 2011, ApJ, 734, 104

Niemi S.-M., 2015, Euclid Visible InStrument (VIS) Python Package (VIS-PP) Documentation, http://www.mssl.ucl.ac.uk/~smn2/

Nightingale J. W., Dye S., 2015, Monthly Notices of the Royal Astronomical Society, 452, 2940

Nightingale J., Hayes R., 2020, PyAutoLens: Open-source Strong Gravitational Lensing, https://github.com/Jammy2211/PyAutoLens

Nightingale J. W., Dye S., Massey R. J., 2018, Monthly Notices of the Royal Astronomical Society, 478, 4738

Nightingale J. W., Massey R. J., Harvey D. R., Cooper A. P., Etherington A., Tam S. I., Hayes R. G., 2019, Monthly Notices of the Royal Astronomical Society, 489, 2049

Nightingale J. W., et al., 2021, Journal of Open Source Software, 6, 2825

Okido D. H., Furlanetto C., Trevisan M., Tergolina M., 2020, Proceedings of the International Astronomical Union, 15, 188–189

Oliver S. J., et al., 2012, Monthly Notices of the Royal Astronomical Society, 424, 1614

Oogi T., Habe A., 2012, Monthly Notices of the Royal Astronomical Society, 428, 641

Oser L., Naab T., Ostriker J. P., Johansson P. H., 2012, ApJ, 744, 63

Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, Monthly Notices of the Royal Astronomical Society, 378, 1550

Papadopoulos P. P., Geach J. E., 2012, The Astrophysical Journal, 757, 157

Park J. W., Wagner-Carena S., Birrer S., Marshall P. J., Lin J. Y.-Y., Roodman A., Collaboration L. D. E. S., et al., 2021, The Astrophysical Journal, 910, 39

Pearson E. A., et al., 2013, MNRAS, 435, 2753

Pearson J., Li N., Dye S., 2019, Monthly Notices of the Royal Astronomical Society, 488, 991

Pearson J., Maresca J., Li N., Dye S., 2021, Monthly Notices of the Royal Astronomical Society, 505, 4362

Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perrotta F., Magliocchetti M., Baccigalupi C., Bartelmann M., De Zotti G., Granato G. L., Silva L., Danese L., 2003, MNRAS, 338, 623

Petrillo C., et al., 2017, Monthly Notices of the Royal Astronomical Society, 472, 1129

Planck Collaboration et al., 2016, A&A, 594, A13

Planck Collaboration et al., 2020a, A&A, 641, A6

Planck Collaboration et al., 2020b, A&A, 641, A6

Poci A., Cappellari M., McDermid R. M., 2017, Monthly Notices of the Royal Astronomical Society, 467, 1397

Poglitsch A., et al., 2010, A&A, 518, L2

Powell D., Vegetti S., McKean J. P., Spingola C., Rizzo F., Stacey H. R., 2020, Monthly Notices of the Royal Astronomical Society, 501, 515

Press W. H., Schechter P., 1974, ApJ, 187, 425

Puglisi A., et al., 2019, ApJ, 877, L23

Racca G. D., et al., 2016, in Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave. p. 99040O

Radeka V., Frank J., Geary J., Gilmore D., Kotov I., O'Connor P., Takacs P., Tyson J., 2009, Journal of Instrumentation, 4, P03002

Refsdal S., 1964, Monthly Notices of the Royal Astronomical Society, 128, 307

Remus R.-S., Burkert A., Dolag K., Johansson P. H., Naab T., Oser L., Thomas J., 2013, ApJ, 766, 71

Ricciardelli E., Trujillo I., Buitrago F., Conselice C. J., 2010, MNRAS, 406, 230

Richards G. T., et al., 2006, AJ, 131, 2766

Riechers D. A., Cooray A., Pérez-Fournon I., Neri R., 2021, The Astrophysical Journal, 913, 141

Rodney S. A., et al., 2016a, ApJ, 820, 50

Rodney S. A., et al., 2016b, ApJ, 820, 50

Rowan-Robinson M., Wang L., 2010, Monthly Notices of the Royal Astronomical Society, 406, 720

Rowlands K., et al., 2014, Monthly Notices of the Royal Astronomical Society, 441, 1017

Ruff A. J., Gavazzi R., Marshall P. J., Treu T., Auger M. W., Brault F., 2011, ApJ, 727, 96

Rybak M., McKean J. P., Vegetti S., Andreani P., White S. D. M., 2015a, MNRAS, 451, L40

Rybak M., Vegetti S., McKean J. P., Andreani P., White S. D. M., 2015b, Monthly Notices of the Royal Astronomical Society: Letters, 453, L26

Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749

Schaye J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521

Schneider P., Sluse D., 2014, A&A, 564, A103

Schneider, Peter Sluse, Dominique 2013, A&A, 559, A37

Schneider P., Kochanek C., Wambsganss J., 2005, Gravitational Lensing: Strong, Weak and Micro. Springer

Scoville N., et al., 2016, ApJ, 820, 83

Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14

Shankar F., et al., 2017, The Astrophysical Journal, 840, 34

Shapley A. E., 2011, ARA&A, 49, 525

Simpson J. M., et al., 2014, ApJ, 788, 125

Smith D. J. B., et al., 2013, Monthly Notices of the Royal Astronomical Society, 436, 2435

Snyder G. F., Cox T. J., Hayward C. C., Hernquist L., Jonsson P., 2011, ApJ, 741, 77

Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, Monthly Notices of the Royal Astronomical Society, 423, 1992

Sonnenfeld A., Treu T., Gavazzi R., Suyu S. H., Marshall P. J., Auger M. W., Nipoti C., 2013, The Astrophysical Journal, 777, 98

Sonnenfeld A., Treu T., Marshall P. J., Suyu S. H., Gavazzi R., Auger M. W., Nipoti C., 2015, The Astrophysical Journal, 800, 94

Sonnenfeld A., et al., 2020, Astronomy & Astrophysics, 642, A148

Speagle J. S., 2020, Monthly Notices of the Royal Astronomical Society, 493, 3132

Spergel, D. 2015

Spilker J. S., et al., 2016, ApJ, 826, 112

Springer ed. 2006, Gravitational Lensing: Strong, Weak and Micro (arXiv:astro-ph/0407232)

Strateva I., et al., 2001, AJ, 122, 1861

Sun F., et al., 2021, The Astrophysical Journal, 908, 192

Sunyaev R. A., Zeldovich Y. B., 1970, Ap&SS, 7, 3

Suyu S. H., 2012, Monthly Notices of the Royal Astronomical Society, 426, 868

Suyu S. H., Marshall P., Hobson M., Blandford R., 2006, Monthly Notices of the Royal Astronomical Society, 371, 983

Swinbank A. M., et al., 2010, MNRAS, 405, 234

Swinbank A. M., et al., 2013, Monthly Notices of the Royal Astronomical Society, 438, 1267

Tadaki K.-i., et al., 2020, ApJ, 901, 74

Tagore A. S., Jackson N., 2016, Monthly Notices of the Royal Astronomical Society, 457, 3066

Tergolina M., 2020, Master's thesis, Federal University of Rio Grande do Sul

Tessore N., Metcalf R. B., 2015, Astronomy & Astrophysics, 580, A79

Tielens A. G. G. M., Allamandola L. J., 1987, Composition, Structure, and Chemistry of Interstellar Dust. Vol. 134, Springer, doi:10.1007/978-94-009-3861-8_16

Toft S., et al., 2014, ApJ, 782, 68

Tohill C., Ferreira L., Conselice C. J., Bamford S. P., Ferrari F., 2021, The Astrophysical Journal, 916, 4

Trayford J. W., Theuns T., Bower R. G., Crain R. A., Lagos C. d. P., Schaller M., Schaye J., 2016, Monthly Notices of the Royal Astronomical Society, 460, 3925

Treu T., Koopmans L. V. E., 2004, ApJ, 611, 739

Treu T., Dutton A. A., Auger M. W., Marshall P. J., Bolton A. S., Brewer B. J., Koo D. C., Koopmans L. V., 2011, Monthly Notices of the Royal Astronomical Society, 417, 1601

Vanderriest C., Schneider J., Herpe G., Chevreton M., Moles M., Wlerick G., 1989, A&A, 215, 1

Vegetti S., Koopmans L. V. E., 2009, Monthly Notices of the Royal Astronomical Society, 392, 945

Vieira J. D., et al., 2013, Nature, 495, 344

Wang S. X., et al., 2013, ApJ, 778, 179

Warren S. J., Dye S., 2003, ApJ, 590, 673

Whitney A., Ferreira L., Conselice C. J., Duncan K., 2021, The Astrophysical Journal, 919, 139

Wong K. C., et al., 2020, Mon. Not. Roy. Astron. Soc., 498, 1420

Wootten A., Thompson A. R., 2009, Proceedings of the IEEE, 97, 1463

Yang C., et al., 2019, A&A, 624, A138

Zhang Z.-Y., et al., 2018, Monthly Notices of the Royal Astronomical Society, 481, 59

Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q., 2019, A Comprehensive Survey on Transfer Learning, doi:10.48550/ARXIV.1911.02685, https://arxiv.org/abs/1911.02685