Morphological and kinematic indicators of structural transformation in galaxies

by

Connor Bottrell
B.Sc., University of Victoria, 2014
M.Sc., University of Victoria, 2016

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Physics and Astronomy

Morphological and kinematic indicators of structural transformation in galaxies

by

Connor Bottrell
B.Sc., University of Victoria, 2014
M.Sc., University of Victoria, 2016

Supervisory Committee

Dr. Luc Simard, Supervisor
(National Research Council of Canada)

Dr. Sara Ellison, Supervisor
(Department of Physics & Astronomy, University of Victoria)

Dr. Katherine S. Elvira, Outside Member
(Department of Chemistry, University of Victoria)

ABSTRACT

The observed properties of galaxies are intricately connected to their respective evolutionary histories. Establishing these connections – tying the morphologies, dynamics, and other properties of galaxies to the dominant events and processes from which they originate – is the central challenge in creating a self-consistent framework for how galaxies form and evolve. Overcoming this challenge requires that two criteria be satisfied: (1) accurate characterization of the *physical states* of galaxies; and (2) creation of models that connect the *observed features* of galaxies to their evolutionary histories. This thesis chiefly concerns the identification and characterization of morphological and kinematic indicators for structural transformation in galaxies and their connections to galaxy mergers – including merger status (merger or non-merger) and merger stage.

Accurate measurement of the morphological structures of galaxies is a cornerstone for making connections to their evolutionary pathways. However, without significant overlap between the observational footprints of deep and shallow galaxy imaging surveys, the extent to which structural measurements for large galaxy samples are robust to image quality (e.g. depth, spatial resolution) cannot be established. Deep images from the Sloan Digital Sky Survey (SDSS) Stripe 82 co-adds provide a unique solution to this problem – offering $1.6-1.8$ magnitudes improvement in depth with respect to SDSS Legacy images. Having similar spatial resolution to Legacy, the co-adds make it possible to examine the sensitivity of parametric morphologies to depth alone. Using the GIM2D surface-brightness decomposition software, I provide public morphology catalogs for 16,908 galaxies in the Stripe 82 *ugriz* co-adds. The methods and selection are completely consistent with those of previous analyses in the shallow images. Measurements in the deep and shallow images are rigorously compared. No systematics in total magnitudes and sizes are found except for faint galaxies in the u-band and the brightest galaxies in each band. However, characterization of bulge-to-total fractions is significantly improved in the deep images. Furthermore, statistics used to determine whether single-Sérsic or two-component (e.g. bulge+disc) models are required become more bimodal in the deep images. Lastly, I show that morphological asymmetries (commonly linked to mergers) are enhanced in the deep images and that the enhancement is positively correlated with the asymmetries measured in Legacy images.

Recently, machine learning has become a popular tool to quantify galaxy mor-

phologies and identify mergers – exploiting the often disturbed and asymmetric morphological features present in merging galaxies. However, this technique relies on using an appropriate set of training data to be successful. By combining hydrodynamical simulations, synthetic observations and convolutional neural networks (CNNs), I quantitatively assess how realistic simulated galaxy images must be in order to reliably classify real mergers. Specifically, I compare the performance of CNNs trained with two types of galaxy images, stellar maps and images with full radiative transfer through internal dust, each with three levels of observational realism: (1) no observational effects (idealized images), (2) realistic sky and point spread function (semi-realistic images), (3) insertion into a real sky image (fully realistic images). I show that networks trained on either idealized or semi-real images have poor performance when applied to survey-realistic images. In contrast, networks trained on fully realistic images achieve 87.1% classification performance. Importantly, the level of realism in the training images is much more important than whether the images included radiative transfer, or simply used the stellar maps (87.1% compared to 79.6% accuracy, respectively). Therefore, one can avoid the large computational and storage cost of running radiative transfer with a relatively modest compromise in classification performance. Making photometry-based networks insensitive to colour incurs a very mild penalty to performance with survey-realistic data (86.0% with r-only compared to 87.1% with gri). This result demonstrates that while colour *can* be exploited by colour-sensitive networks, it is not necessary to achieve high accuracy and so can be avoided if desired. I provide the public release of the statistical observational realism suite, RealSim, as a companion to this work.

Galaxy kinematics derived from observational integral field spectroscopy (IFS) may offer an orthogonal and highly-complimentary basis to photometry for accurately identifying and characterizing observed galaxy mergers. As with morphology, mergers can trigger kinematic disturbances in galaxies resulting in irregular and asymmetric kinematic structure. However, these *kinematic* disturbances are not always reflected in the morphologies. The current and future state-of-the-art IFS instruments which provide spatially-resolved kinematics for many thousands of galaxies make kinematic merger studies statistically viable. Anticipating the demand for realistic synthetic IFS and kinematic data for calibrating merger classification models with simulations, I present RealSim-IFS: a novel tool that emulates the instrumental response of current and future fibre-based IFS instruments. Components of RealSim-IFS are tested on real IFS data from the Mapping Nearby Galaxies at Apache Point Ob-

servatory (MaNGA) survey to demonstrate the high precision that is achieved by RealSim-IFS. In a further demonstration with RealSim-IFS, I generate realistic synthetic MaNGA kinematic observations for a sample of galaxies from the IllustrisTNG cosmological hydrodynamical simulations. The survey-realistic kinematic maps for post-merger galaxies are compared with non-merging galaxies to illustrate the potential role of kinematics in enabling more accurate identification and characterization of galaxy mergers – either independently or in tandem with photometry.

Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

ADC — Analog-to-Digital Converter

AGB — Asymptotic Giant Branch

AGN — Active Galactic Nucleus

APO — Apache Point Observatory

BOSS — Baryon Oscillation Spectroscopic Survey

CALIFA — Calar Alto Legacy Integral Field Area

CAS — Catalog Archive Server

CCD — Charge-Coupled Device

CFIS — Canada France Imaging Survey

CGM — Circumgalactic Medium

CMB — Cosmic Microwave Background

CNN — Convolutional Neural Network

DAP — Data Analysis Pipeline

DAR — Differential Atmospheric Refraction

DAS — Data Archive Server

DRP — Data Reduction Pipeline

DR[#] — Data Release [Version #]

DN — Digital Number (sometimes also ADU or DU in literature)

EAGLE — Evolution and Assembly of GaLaxies and their Environments

EW — Rest-frame Equivalent Width

FWHM — Full Width at Half Maximum

FIRE — Feedback In Realistic Environments (simulation model)

FOV — Field of View

HWHM — Half Width at Half Maximum

IFS — Integral Field Spectroscopy

IFU — Integral Field Unit

IGM — Intergalactic Medium

IMF — Initial Mass Function

ISM — Interstellar Medium

MAGPI — Middle Ages Galaxy Properties with IFS

MaNGA — Mapping Nearby Galaxies at APO

MCMC — Markov Chain Monte Carlo

MFM — Meshless Finite Mass

MOP — Median Overall Performance

NaN — Not a Number (data type)

NIR — Near Infrared

QSO — Quasi-Stellar Object (Quasar)

RSS — Row-Stacked Spectra

SAMI — Sydney Anglo-Australian Observatory Multi-Object IFS

SDSS — Sloan Digital Sky Survey

SED — Spectral Energy Distribution

SFR — Star Formation Rate

SMBH — Super-massive Black Hole

sSFR — Specific Star Formation Rate

TNG — The Next Generation (simulation model)

UV — Ultraviolet

ACKNOWLEDGEMENTS

I am exceptionally grateful to the mentors, collaborators, and peers who made the completion of this thesis possible and thoroughly enjoyable.

Foremost, my experience benefited greatly from the excellent supervision by Luc Simard and Sara Ellison. I thank Luc Simard for taking me on, for being an indispensable library of technical knowledge, and for valued scientific and career guidance. I am indebted to Sara Ellison for her exceptionally keen scientific insight, active networking on my behalf, and for the inspirational role she has played in my career thus far. To Sara and Luc, I cannot express the depth of my gratitude. I also thank Katherine Elvira for taking part in my supervisory committee.

Furthermore, the research conducted in this thesis benefited from the efforts and insights of my co-authors and collaborators: Maan Hani, Hossen Teimoorinia, Jorge Moreno, Trevor Mendel, Paul Torrey, Chris Hayward, Mallory Thorp, and Lars Hernquist. Thank you all.

Lastly, there is nothing I can think of that was more therapeutic than going head-to-head and shoulder-to-hip with my teammates and brothers at Westshore Valhallians and Rugby Club de Montréal. Nothing clears the head like a concussion.

Chapter 1

The complexity and diversity of galaxies in the Universe

The remarkable diversity in the observed structures and properties of galaxies presents one of the central challenges to galaxy formation and evolution theory. The spirit of this challenge is to use physical principles to link the observed properties of galaxies to their evolutionary histories and origins. In other words, we want a framework which encompasses the processes that drive the transformation of a galaxy from one physical state to another. What events or processes determine the present structure and dynamics of a galaxy? What stifles or rejuvenates the formation of stars? Is the relative importance of a given process stable across cosmic time? Tackling these questions will rely critically on the ability to *identify* and *distinguish between* different phases in the evolution of galaxies through detailed descriptions of their physical states. This task is complicated by the enormous variety of galaxies in the Universe *and*, consequently, the large number of parameters needed to describe the physical state of an individual galaxy.

1.1 The origin of structure in the Universe

The first redshift surveys to map the 3-dimensional positions of $\sim 10^2 - 10^3$ galaxies revealed a vast and highly non-uniform network distributed throughout the local Uni-

verse[1] (e.g. Kirshner et al. 1978; Davis et al. 1982; Geller & Huchra 1989). Shortly thereafter, subsequent redshifts surveys mapping $\sim 10^5$ exposed this *cosmic web* of structure with exquisite detail (see Figure 1.1 and, e.g. Colless et al. 2001; Blanton et al. 2003b). Quantitatively, this non-uniformity (or, equivalently, clustering) of galaxies on large scales can be measured with a 2-point correlation function: defined as the *excess* probability $\xi(r)$ of finding another galaxy at a distance r from some galaxy (relative to a uniform random distribution) averaged over the full ensemble of galaxies. The measurements of $\xi(r)$ from these surveys, their predecessors, and successors provide fundamental constraints on cosmological models – whose merit is partially determined by how well the large-scale structure of the local Universe is reproduced (e.g. Cole et al. 1998; Springel et al. 2006). Similarly, the observed temperature over-density (anisotropy) field measured in the radiation from the cosmic microwave background (CMB, also measured as $\xi(r)$) provides a crucial observational benchmark for the behaviour of cosmological models in the early Universe (e.g. Smoot et al. 1992; Hinshaw et al. 2013; Planck Collaboration et al. 2018). Combined, these local and early-Universe measurements have provided an important basis for the current concordance cosmogony, ΛCDM – a parametrization of the Big Bang cosmological model.

1.1.1 The cosmological constant

In the ΛCDM model, Λ represents a cosmological constant which, unlike the gravitational forces on matter, gives rise to the unremitting and currently accelerating expansion of the Universe. This component of ΛCDM is rooted in early measurements of the Universe's expansion using the distances and relative velocities of galaxies (Hubble, 1929). The first measurements of the *accelerated* expansion of the Universe (and, consequently, evidence of a cosmological constant) came at the turn of the millennium using Type Ia Supernovae data (Riess et al. 1998; Perlmutter 2003; see also Frieman et al. 2008b for a review). The nature of the force driving this expansion and giving rise to the cosmological constant, so-called *dark energy*, is currently unknown. Nonetheless, CMB measurements show that dark energy must dominate the mass-energy budget of the Universe at 68.9% – with only 31.1% attributed to matter

[1]Taken at face-value, it should at least be somewhat remarkable that a cosmic web of structure - galaxies of various sizes, types, and colours, galaxy groups and clusters, filaments, and voids - should be the outcome of the hot and highly-uniform soup of material in the immediate aftermath of the Big Bang.

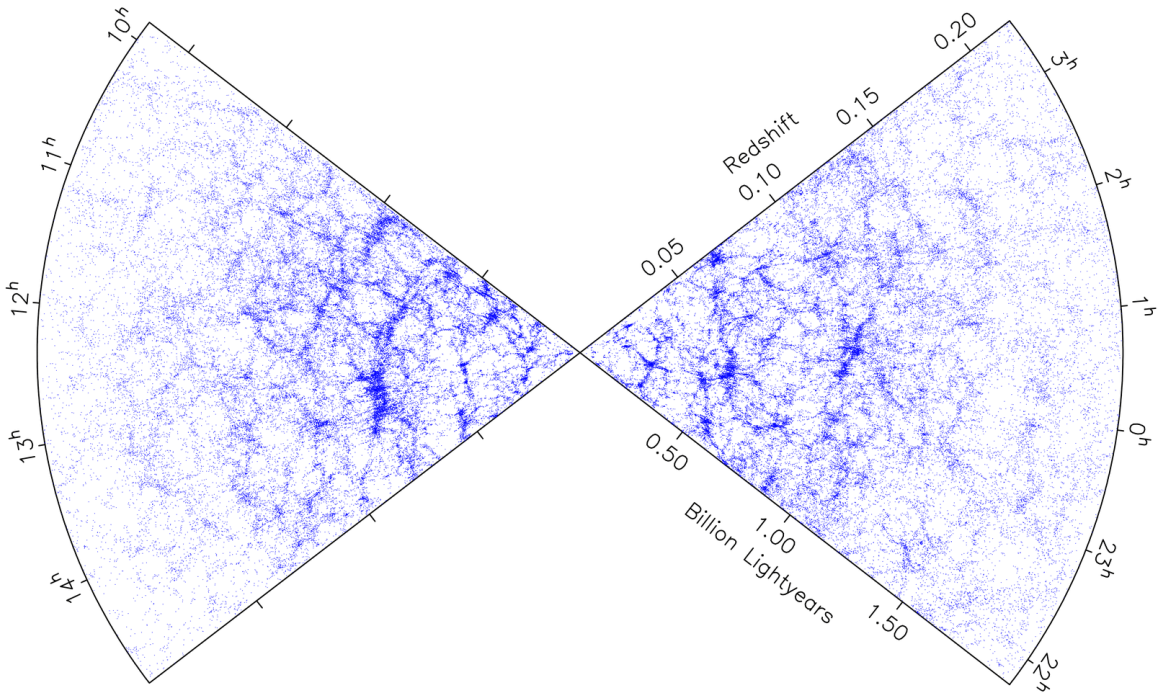

Figure 1.1 "Wedge Plot" of large-scale structure in the local Universe as measured by the 2dF Galaxy Redshift Survey (Colless et al., 2001). Galaxies are distributed in vast filamentary structures, groups, clusters, and superclusters surrounded by voids where few galaxies reside. The clustering statistics from these measurements (among those from other surveys) were used estimate the matter density and baryonic fraction of matter in the Universe (e.g. Percival et al. 2001; Cole et al. 2005). The observer is at the intersection of the two wedges – looking outward in two directions. Each wedge is a 2-dimensional projection of the 3-dimensional galaxy distribution. Each blue dot shows the redshift (also distance in light-years) and right ascension of a single galaxy. The apparent decrease in the numbers of observed galaxies at higher redshifts is due to observational limitations. Credit for this figure goes to (Colless et al., 2001) and the 2dF Galaxy Redshift Survey team (www.2dfgrs.net). This figure has been adapted with permission from Matthew Colless.

(Planck Collaboration et al., 2018). Indeed, only 15.8% of this Universal matter content is the ordinary matter comprising neutrons, protons, and electrons. The other 84.2% is the second core component of the mass-energy budget in ΛCDM – cold dark matter.

1.1.2 Cold dark matter

Dark matter is invoked as a necessary component of the Universe's mass budget to explain gravitational effects on the scales of galaxies and galaxy clusters as well as observations of the CMB. The first line of evidence for a new, exotic type of matter was inferred from dynamics of the nearby Coma cluster of galaxies (Zwicky, 1933). The calculations showed that the velocities of the cluster's constituent galaxies were far too large for them to be bound gravitationally by a mass derived from the visible matter alone. Since then, further indirect evidence for the existence of dark matter through its gravitational influence has expanded to the rotational velocities of disc galaxies (e.g. Rubin & Ford 1970; Rubin et al. 1980; Courteau 1997), the distorted images (lensing) of background sources as their light travels through dense media such as other galaxies, clusters, and the large-scale mass distribution (Fischer et al., 2000; Clowe et al., 2000; Wilson et al., 2001; Refregier, 2003), its effect on the CMB temperature anisotropy (e.g. Hinshaw et al. 2013; Planck Collaboration et al. 2018), and many others. As with dark energy, the exact nature of dark matter is unknown and it has not yet been directly detected. Therefore, *cold* dark matter (CDM) is a distinct hypothetical type of dark matter (other types include *warm* and *hot*) that moves slowly compared to the speed of light and does not interact with itself or anything else except via gravity. In contrast to CDM, warm and hot dark matter models generally do not facilitate the formation of structures that agree with the observed galaxy distribution (e.g. White et al. 1983; Blumenthal et al. 1984; Frenk et al. 1988).

1.1.3 The formation of structure in a ΛCDM universe

Until around 380,000 years after the Big Bang, protons, neutrons, electrons, and photons were held in thermal equilibrium by extreme temperatures and densities. The tight coupling of photons to baryons was primarily mediated by Thomson and Coulomb scattering processes. The matter content of the Universe, expanding adiabatically, was highly homogeneous and smoothly distributed – with small local fluc-

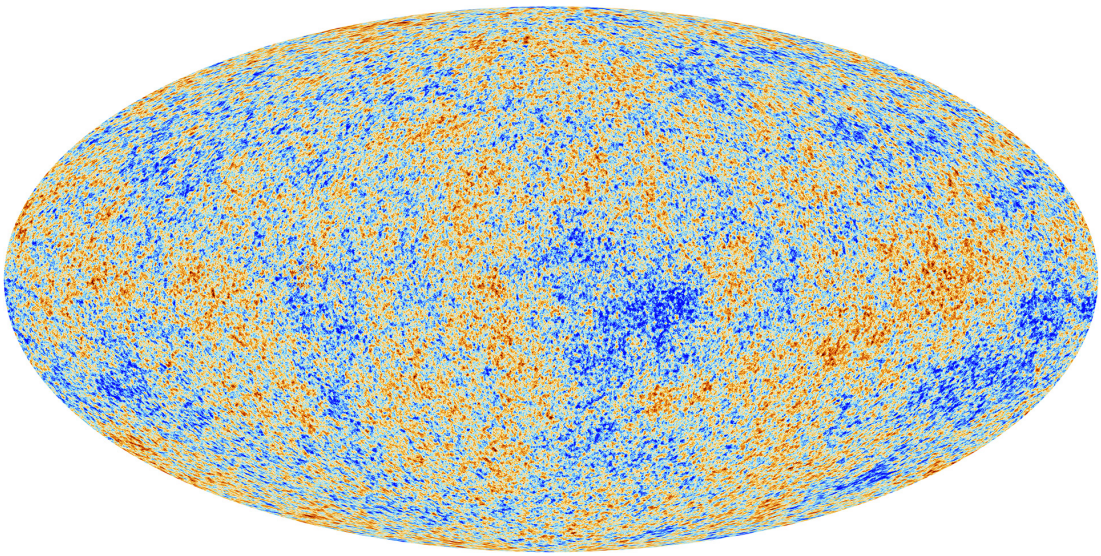

Figure 1.2 All-sky map of the cosmic microwave background (CMB) radiation. The image combines data from the first 15 months of observations with the Planck space telescope (Planck Collaboration et al., 2016). Colours show the small spatial temperature fluctuations in the CMB radiation ($\Delta T_{\rm rms} \sim 10$ μK; red is hot, blue is cold). The temperature fluctuations correspond to regions of local over-density and under-density in the matter distribution around 380,000 years after the Big Bang. These small fluctuations in density are the gravitational seeds of all present day structure in the Universe. Image credit: Planck Collaboration, European Space Agency (https://www.cosmos.esa.int/web/planck).

tuations in the matter density field on the order of $\Delta\delta/\delta \sim 10^{-5}$. The *seeds* for these fluctuations originate in the very early Universe, around $t \sim 10^{-35}$s after the Big Bang, where quantum fluctuations in the primordial fluid were blown up to macroscopic scales by a rapid period of exponential expansion (called *inflation* – Guth 1981; Linde 1982; Albrecht & Steinhardt 1982). With time, these initially weak over-densities grew through gravitational aggregation of dark matter which, unlike baryons, is not opposed by outward radiation pressure. As the perturbations grew through the collapse of dark matter, the baryons – attracted gravitationally to the dark matter but still tightly coupled to the radiation field – also grew over-dense in the regions where dark matter was collapsing (but to a lesser degree than the dark matter itself).

Meanwhile, the Universe was still expanding and cooling in temperature. The mean energy of the radiation was falling. Temperatures were slightly higher in the over-dense regions and lower in under-dense regions. Once the mean photon energy

reached $\sim$ 1eV, too few photons remained with energies necessary to prevent sustained combination of protons and electrons and formation of hydrogen atoms. This epoch is called *recombination*[2] and occurred around 250,000 years after the Big Bang. As neutral Hydrogen atoms formed and the number of free electrons diminished, the travel-time of photons between successive photon-electron scattering events, Γ^{-1}, increased proportionately with $\Gamma = n_e \sigma_e c$, where n_e is the number density of electrons and σ_e is the Thomson scattering cross-section.

Shortly after recombination, the rate of photon-electron interactions, Γ, was overtaken by the rate of expansion of the Universe (i.e. on average, the Universe was expanding faster than a photon could find a free electron from which to scatter). Without free electrons from which to scatter, the photons *decoupled* from the baryonic fluid and streamed freely through the Universe, forming the CMB we observe today (Penzias & Wilson, 1965; Smoot et al., 1992). It should be noted that photons outnumbered baryons by $\sim 10^9$ to 1 before decoupling. Consequently, CMB radiation almost entirely comprises photons that were locked in the photon-baryon fluid before decoupling – with only 1 photon per billion arising from transitions between energy states in cooling Hydrogen atoms. As such, the majority of CMB photons come from the epoch just before decoupling – when every part of the photon baryon fluid (including over- and under-dense regions) was in local thermodynamic equilibrium. And, therefore, the radiation spectrum of every part of the CMB follows a near-perfect blackbody distribution whose peak wavelength/frequency is directly related to its temperature.

Figure 1.2 shows the temperature map of the CMB measured by Planck Collaboration et al. (2016). Red and blue correspond to regions that are over-dense (hot) and under-dense (cold), respectively, in dark and baryonic matter with respect to the average. These small differences visible in the CMB ($\Delta T_{\rm rms} \sim 10$ μK compared to an average temperature of around 2.73 K; Fixsen 2009), are connected to the richness of structures seen later in the Universe through a complex network of processes driven primarily by gravity – including both the large scale structure from Figure 1.1 and galaxies such as those shown in Figure 1.3 (e.g. White & Rees 1978; White & Frenk 1991).

Between recombination and decoupling, the radiation pressure on the baryons dropped precipitously. The over-densities in baryonic matter, now with significantly less radiation pressure to curb collapse and fragmentation, could fall deeper and

[2]A notorious misnomer which implies that electrons and protons had once already been combined.

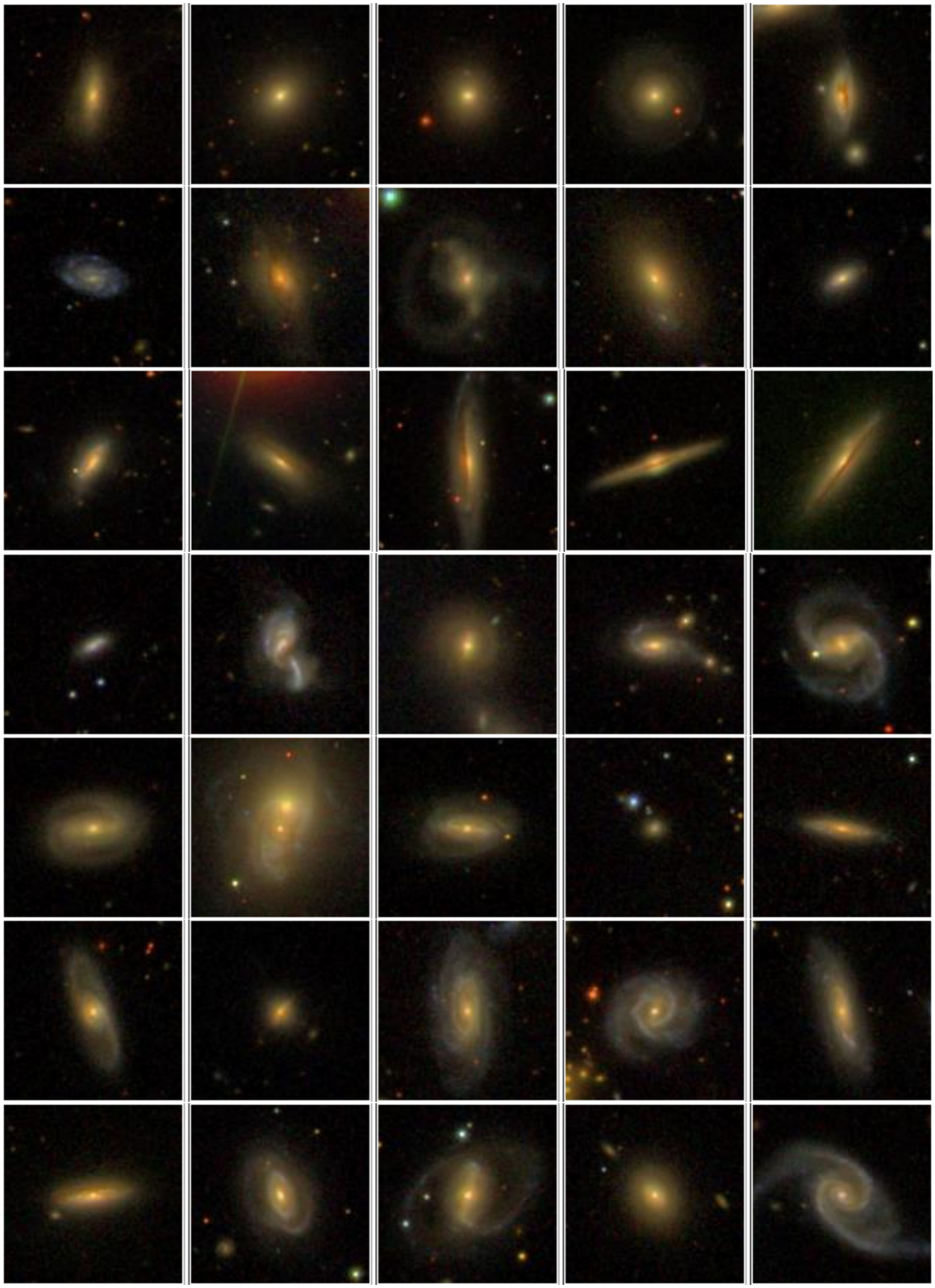

Figure 1.3 Random selection of 35 galaxies with stellar mass estimates, $11.0 < \log(M_\star/M_\odot) < 11.6$, and redshifts, $z < 0.05$. Imaging is from the Sloan Digital Sky Survey (SDSS) DR7 (Abazajian et al., 2009) with composition of *gri* colours following Lupton et al. (2004). SDSS Website: https://www.sdss.org

become gravitationally bound within the corresponding dark matter over-densities (which themselves continue to deepen and fragment into a bottom-heavy mass distribution of quasi-equilibrium dark matter clumps called *haloes*; Zel'Dovich 1970; Harrison 1970; Zeldovich 1972; Gunn & Gott 1972; Press & Schechter 1974). From here, the prevailing theory for large-scale structure and galaxy formation proceeds in two main steps (White & Rees, 1978; Blumenthal et al., 1984; Davis et al., 1985; Frenk et al., 1985; White & Frenk, 1991). First, massive haloes and structures form *"bottom-up"* through the continuous hierarchical assembly of smaller haloes and their constituent baryons across cosmic time (e.g. Peebles 1965; Peebles & Yu 1970). Second, the baryonic content cools and fragments within the gravitational potential of their host haloes to form stars and galaxies (Silk, 1977; Rees & Ostriker, 1977; Blumenthal et al., 1986).

The first of the above steps, the clustering and assembly of haloes, is entirely governed by gravity. The physics of gravity is well-understood and has been invoked extensively in so-called N-body numerical simulations – which model the formation of structure under gravitational forces alone (no baryonic-specific physics such as radiative processes). Indeed, cosmological N-body simulations of representative volumes of the Universe (using initial conditions derived from CMB observations) produce large-scale structure that agrees exceptionally well with observations (e.g. Springel et al. 2005b, 2006). This agreement is expected despite exclusion of baryonic physics because: (a) dark matter dominates the matter content of the Universe by a factor of roughly 6:1; and (b) dark matter only interacts gravitationally.

In contrast, the second step, the cooling and formation of galaxies within dark haloes, requires a framework of processes which govern the evolution of baryonic matter. A highly incomplete list includes gas heating and cooling processes, star-formation, stellar evolution, outflows and chemical enrichment from supernovae explosions, dilution from extra-galactic gas accretion, and black-hole accretion. Many details of these processes are unknown. The properties of galaxies, their morphologies, and dynamics encode information about the processes which dominated their formation and evolution. By studying galaxies and distinguishing between galaxies of different types and evolutionary stages, we can connect the properties of galaxies to these processes.

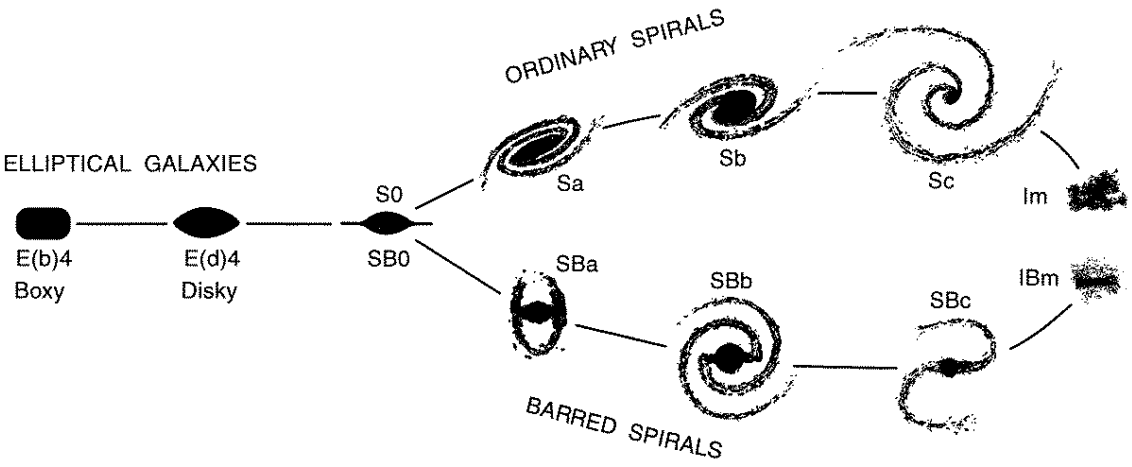

Figure 1.4 A revised adaptation of the Hubble Sequence of galaxy morphology classifications. This figure has been reproduced from Kormendy & Bender (1996) with permission from the lead author.

1.2 The observed structures of galaxies today

The goal of morphological characterization is to establish means of connecting the observed structures of galaxies[3] to the physical processes and events which shape and transform them. One of the most substantial and lasting contributions to morphological characterization of galaxies is the creation of the Hubble Sequence (Hubble, 1926, 1936) – a sequence of visual morphological classifications into which most observed galaxies can be taxonomically organized. While several updates to the original sequence have been suggested to re-contextualize the Hubble Sequence in light of new observational constraints (e.g. de Vaucouleurs 1959a; Sandage 1961, 1975; Elmegreen et al. 1992; Kormendy & Bender 1996; van den Bergh 1998; Kormendy & Bender 2012), many of the core tenets hold.

1.2.1 Visual morphology

Figure 1.4 shows an adaptation of the Hubble Sequence from Kormendy & Bender (1996). On the left are elliptical galaxies which are characterized by smooth, nearly elliptical contours of surface brightness (isophotes), more rounded intrinsic shapes,

[3]Note the shift in terminology here. Here, *structure* is referring to the morphological characteristics of *individual* galaxies. This clarification is made so as not to be confused with the large-scale structure of galaxies in the Universe – which refers to the manner in which galaxies are distributed *collectively*.

and a lack of obvious substructure. Moving to the right are spirals which are characterized by flattened disc-like and spiral arm structures. The spirals are further divided into barred and unbarred groups of classes. At the junction of spirals and ellipticals are lenticular or S0 galaxies. Like ellipticals, S0 galaxies have smooth surface brightness contours (i.e. no spiral arm structures or bright patches where star-formation is actively occurring). However, like spirals, S0 galaxies have a substantial flattened disc component. At the right end of the sequence are irregulars whose appearances are typically dominated by knots of bright star-forming regions with no obvious symmetry. One of the defining features of the Hubble Sequence is that a greater fraction of the total galaxy light originates from a central, spheroidal component as one moves leftward in the sequence.

While the Hubble Sequence classifications are only based on visual morphology, galaxies can also be characterized by other *quantitative* properties which correlate with position in the sequence (though often with considerable scatter[4]). Specific star formation rates (the star-formation rate of a galaxy normalized by its total mass in stars) tend to increase from the left to the right – with the low values in ellipticals and S0 galaxies often branding them as *quiescent* systems. Consequently, the constituent stars are typically younger and bluer in spirals and older and redder in ellipticals. Given that the rate of star formation in a region is a reflection of the availability of cool gas from which it forms (Schmidt, 1959; Kennicutt, 1998b), the mass of gas in neutral and molecular states also typically increases from left to right along the Hubble Sequence (see Kennicutt 1998a and references therein). Morphology also appears to be sensitive to environment. The relative incidence of elliptical and S0 galaxies increases in dense groups of galaxies and galaxy clusters while the incidence of spirals decreases (Dressler, 1980; Butcher & Oemler, 1984; Aragon-Salamanca et al., 1993; Balogh et al., 2004). The connection between physical and environmental properties such as these with morphology provide an observational basis for theoretical models of galaxy evolution.

Figure 1.5 shows examples of how the properties mentioned in the previous paragraph manifest in the spectra of galaxies on the Hubble Sequence. Ellipticals tend to produce relatively more light at longer wavelengths than short. Old and red stars dominate in most elliptical spectra as they are deficient in young stars with bright

[4]It is important to note, however, that part of this scatter comes from the uncertainty in the measurement of a property and does not solely reflect the intrinsic scatter of that property in a galaxy population.

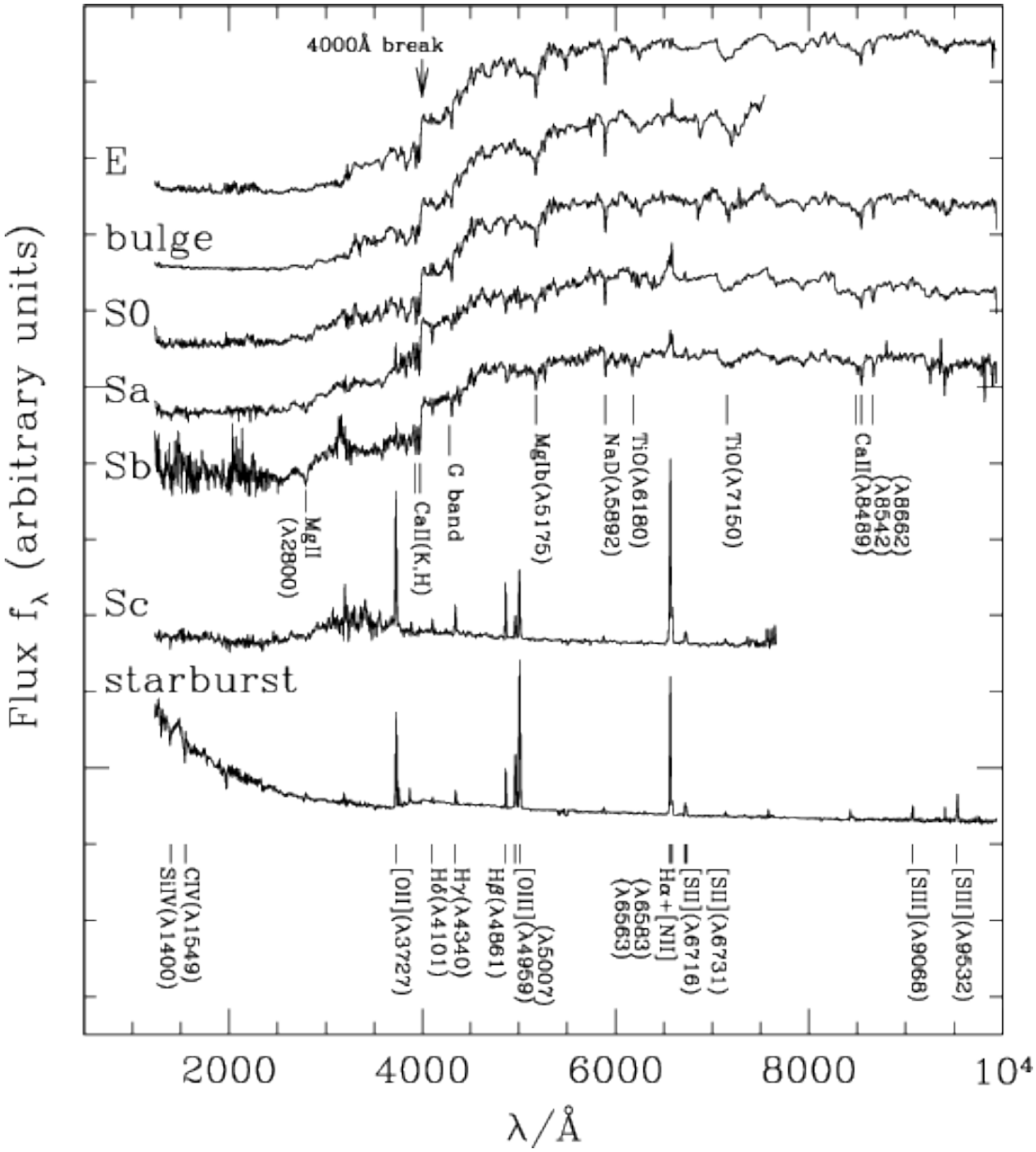

Figure 1.5 Spectral flux density distributions of various galaxy morphological types (energy/time/area per unit wavelength). The vertical offset between each spectrum is only for visualization and does not reflect an intrinsic intensity difference between classes. This figure is a reproduction of Figure 2.12 in Mo et al. (2010) and has been reproduced in this thesis with permission from the lead author. This figure uses data generously supplied by S. Charlot.

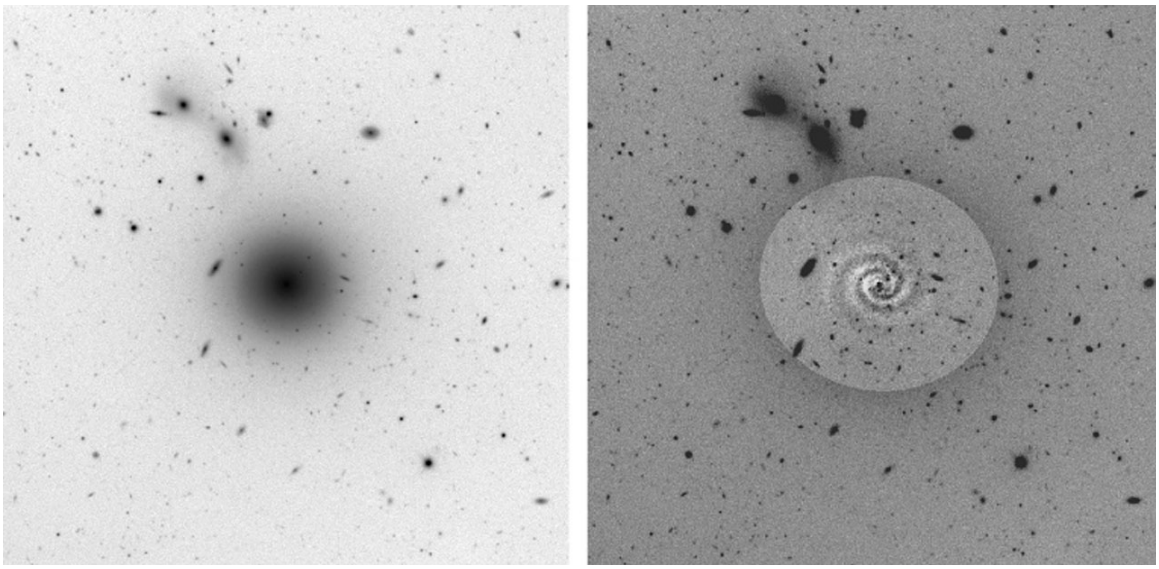

Figure 1.6 Virgo Cluster member IC3328 – previously classified as a dwarf-elliptical. Left: deep r-band image. Right: the same image after removal of the axis-symmetric light from the galaxy via modelling of the galaxy surface-brightness profile using an analytic function and the contrast enhanced greatly in the subtraction region. Figure credit: European Southern Observatory (ESO).

blue continua. The elliptical spectrum is riddled with absorption lines characteristic of the atmospheres of solar (dwarfs) and sub-solar (sub-dwarf) type stars and has no prominent emission lines. Interestingly, many of the characteristic properties of the elliptical spectrum are shared by the central bulges of spiral galaxies – implying the possibility of similar evolutionary histories (more on that later when galaxy interactions are discussed). The drop in flux to the left of 4000Å seen in ellipticals and S0s becomes gradually less pronounced in the spirals. The spiral spectra also have increasingly strong emission lines which arise from heating of interstellar gas by hot young stars.

The observations demonstrate that visual morphology is intricately connected to the physical processes in galaxies. However, characterization of morphology based on visual appearance alone is limited in its capacity for comparison with theoretical predictions. The chief limitation is subjectivity – which is manifested in two ways. First, visual morphology is subjective to the interpretation of the person making the classifications. Solutions to this problem (and the problem that current and future surveys have impractically large galaxy samples for individual classification) are to average over classifications of a large number of participants such as is done in the

Galaxy Zoo project (Lintott et al., 2008) or to use a machine-learning approach (Section 1.4). But second, visual classification is subjective to the properties of the images used for the classifications. The visual appearance of a galaxy is wavelength-dependent (e.g. Bamford et al. 2009; Häußler et al. 2013). Furthermore, Figure 1.6 touches upon the additional complication of contrast. Indeed, in Figure 1.6, it is unlikely that any contrast could have been selected such that the underlying spiral disc in the otherwise smooth ellipsoid would be revealed – that is, not without modelling and subsequent subtraction of the dominant, symmetric part of the light distribution. In general, depending on the choice of contrast (which is often fixed in multi-user participation projects), faint or intermediate-brightness features can be obscured and neglected in classification. These limitations highlight the demand for consistent and quantitative characterization of the morphologies and kinematics of galaxies which allow theoretical predictions to be benchmarked against the real Universe.

1.2.2 Quantitative morphology

The salient feature of most visual classification sequences is that the horizontal position in the sequence can be broadly ascribed to the relative contribution of a disc or ellipsoidal component. In spiral and S0 galaxies, the ellipsoidal component is commonly called the bulge. Indeed, the most prevailing theoretical picture (the one most compatible with observations) is that ellipticals are most likely to be bulges that have outgrown their discs and that interactions between galaxies (mergers) are responsible for this growth (Lynden-Bell, 1967; Toomre & Toomre, 1972; Toomre, 1977; Negroponte & White, 1983; Barnes, 1988; Hopkins et al., 2008b, 2009; Berg et al., 2014). The crucial implication is that the mechanism which induces growth in the bulge also induces the observed changes in the physical properties of galaxies along the Hubble Sequence (or at least correlates with these changes). Accurate characterization of the *structures* of bulges and discs in galaxies is therefore of enormous astrophysical importance insofar as these structures encode the physical processes which dominated their evolution. Quantitative morphologies are well-suited for this task.

The goal of quantitative morphologies is to describe the light distributions of stars in galaxies with either a single estimator or with a model described by a set of parameters. A wealth of tools has been developed (e.g. Abraham et al. 1994; Peng et al. 2002; Simard et al. 2002; Conselice 2003; de Souza et al. 2004; Lotz et al. 2004;

Conselice 2006; Méndez-Abreu et al. 2008; Vika et al. 2013; Ciambur 2015; Robotham et al. 2017). They are split into two categories: parametric and non-parametric. Non-parametric methods can be advantageous in the sense that they assume nothing about the intrinsic shapes of galaxy structures. One example is the Asymmetry parameter from the CAS non-parametric classification system (A, Conselice 2003) – in which an image centred on a galaxy is rotated by 180° and subtracted from the original image. In this way, asymmetric structures in the image are enhanced and quantified by computing the fraction of asymmetric light to symmetric light. The drawback is that, because nothing is assumed about the intrinsic structures of galaxy components, they cannot be used to characterize the structural *components* of a galaxy – only the galaxy as a whole.

In contrast, parametric methods model the light profiles of galaxies with analytic functions. The parameters of the models are optimized based on the stellar light distributions of galaxies in images and can use optimization strategies with varying levels of sophistication. Given the importance of bulges and discs in morphological characterization, one of the hallmarks of parametric morphologies is the bulge-disc decomposition – which simultaneously models the bulge and disc light profiles in a single fit. The profiles of bulges and discs are often very different (de Vaucouleurs, 1959b; Sérsic, 1963; Freeman, 1970; Kormendy & Djorgovski, 1989). Ellipticals and the bulges of spiral galaxies have more cuspy, centrally-peaked profiles but also broad diffuse wings. Discs tend to have exponential profiles. The types of model profiles that are used to describe the separate components are empirically motivated through observational analyses of galaxies in which either the bulge or disc component dominates completely.

Figure 1.7 shows an example of a one-dimensional bulge-disc decomposition for an elliptical galaxy with an intermediate disc from Graham et al. (2016). The left panels show, from top to bottom, the science image, best-fit model, and residual (best-fit model subtracted from the science image). The right panels show the fit to the data in more detail. The top right panel shows the surface-brightness, μ [mag/arcsec^2], as a function of radius from the centre of the galaxy. Circles show the circularized radii and surface brightness of the galaxy's isophotes (contours of constant surface-brightness). The red curve is the best-fitting profile for the bulge. The blue curve, substantially dimmer, is the profile for the disc. The black curve, only barely visible against the circles, is the combination of both model profiles. The panel just below shows the residual surface brightness of these contours after subtracting the model

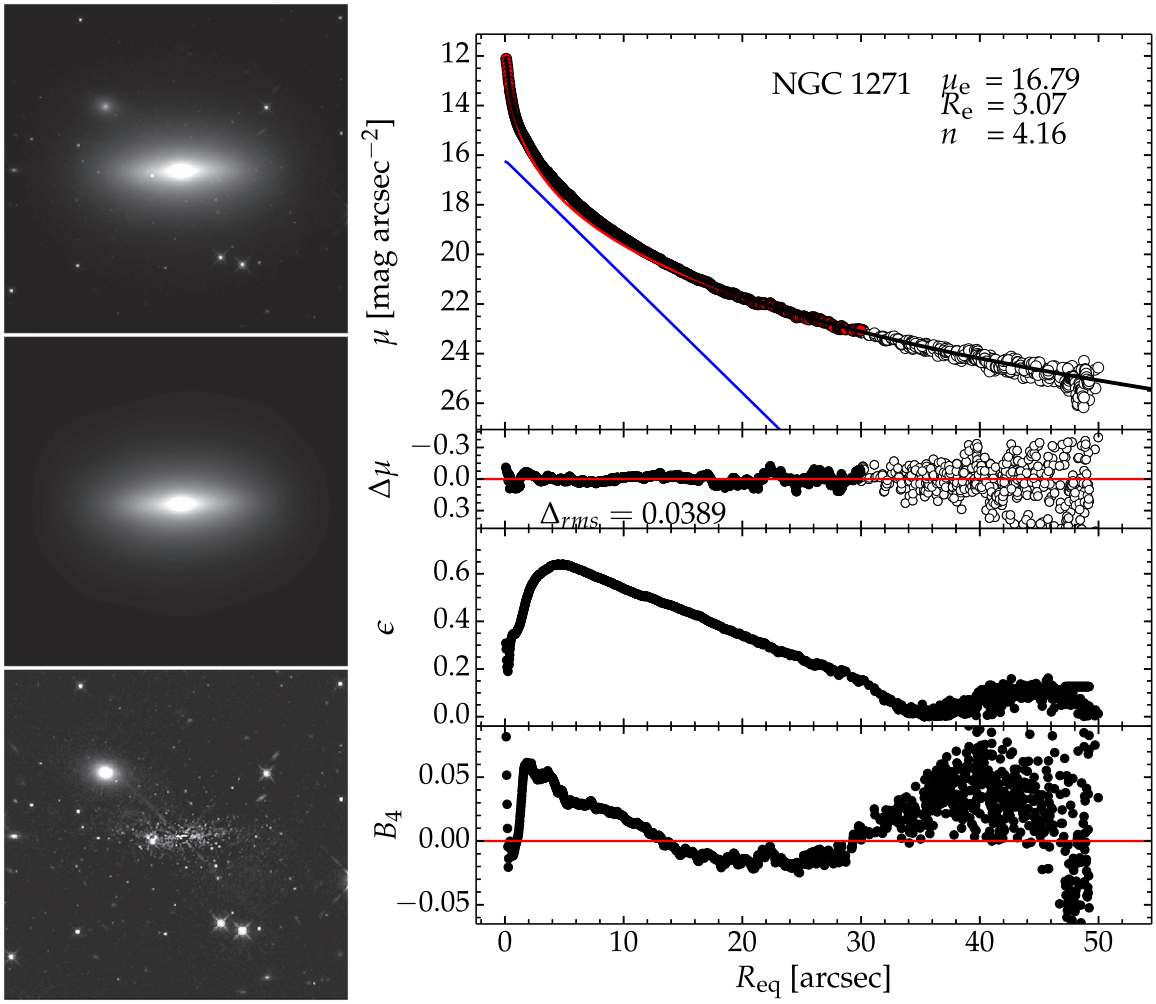

Figure 1.7 Bulge-disc decomposition of NGC1271 using the ISOFIT tool (Ciambur, 2015). On the left are the image, model, and model-subtracted image (residual) for the best-fitting set of model parameters. On the right is a 1-dimensional visualization of the fit to the galaxy light profile. The red line in the top panel is the best-fitting model to the bulge profile. Note its central cusp and extended wings. The blue shows the model for the disc. In a plot of surface brightness μ (a logarithmic quantity) versus radius, an exponential curve is a straight line. The circles are the brightnesses of isophotes extending from the galactic center. Below, the differences between the model brightnesses and the image, $\Delta\mu$, are shown. The root mean squared offset from zero in these residuals Δ_{rms} can be used as an indicator for the successfulness of the model in fitting the data. The ellipticity of the isophotes ϵ and a characterization of whether these isophotes are discy or boxy B_4 are also shown. This figure has been adapted from Figures 2 and 3 from Graham et al. (2016) with permission from the lead author.

profile, $\Delta\mu$. The lower two panels show the ellipticity, ϵ, and boxiness, B_4 of the isophotes as functions of radius.

Automatic bulge-disc decompositions that require no baby-sitting are now used ubiquitously in large galaxy surveys (e.g. Allen et al. 2006; Benson et al. 2007; Simard et al. 2011; Lackner & Gunn 2012; Kelvin et al. 2012; Meert et al. 2015). However, few galaxies are described *perfectly* by simple axisymmetric models for a bulge and disc. Many galaxies contain other structures such as stellar nuclei, bars, isophotal twists, and bright regions of active star-formation. Recent or ongoing mergers between galaxies also produce asymmetric structures such as stellar streams, shells, bridges, and tidal arms (Toomre & Toomre, 1972; Casteels et al., 2013, 2014; Patton et al., 2016). Such structures cannot be accurately described by parametric analytic functions without overcomplicating the models and introducing large covariances and degeneracies between model parameters. Consequently, they often lead to systematics in measurements of bulge and disc properties.

Figure 1.3 shows a few examples of galaxies in the local Universe ($z < 0.05$) which exhibit features of galaxies which are either currently merging or have recently merged. Ongoing mergers (called *pairs*) are often identified by the presence of multiple nuclei within the same galactic body or two distinct galaxies which both exhibit tidally disrupted structures (e.g. [Row, Column]: [4, 2], [5, 2]). Recently completed mergers (called *post-coalescence* galaxies or *post-mergers*) often exhibit similar tidal features to pairs but with a single galactic nucleus (e.g. [2, 3], [7, 3]). While it is difficult to model such features parametrically, they can still be identified by visual inspection or non-parametric measurements. For example, The residual asymmetry (R_A, Elmegreen et al. 1992; Schade et al. 1995) is analogous to the Asymmetry parameter from the CAS system but is measured after a parametric model has already been subtracted from the image (e.g. the residual in the right panel of Figure 1.6). Gini and M20 coefficients (Lotz et al., 2004), outer asymmetry (Wen et al., 2014), and shape asymmetry (Pawlik et al., 2016) are other non-parametric estimators which have been used to quantitively identify galaxy mergers.

Ultimately, the quality of bulge-disc decompositions that makes them desirable for comparison with theoretical predictions is that they provide repeatable quantitative measurements of *bulge* and *disc* photometric structures. If the relative characteristics of bulges and discs encode vital information about a galaxy's evolutionary history (such as their merger histories), it is important to know that these quantitative measurements are robust to the quality of images in which the measurements are made.

One particularly widely used catalog of bulge-disc decompositions is the catalog of 1.12 million galaxies' structural parameters for the Sloan Digital Sky Survey (SDSS) (Simard et al., 2011) – a wide-field imaging and spectroscopic survey. Recently, several repeated observations of a particular section of this survey, Stripe 82, have been stacked to produce significantly deeper images. In Chapter 2 of this thesis, I exploit these deep images to examine the reliability of the Simard et al. (2011) structural measurements. The improved characterization of structure offered by nearly two magnitudes (a factor of ~ 5 improvement in signal-to-noise) allows me to test: (1) the robustness of the decompositions to a better photometric characterization of substructure; (2) the ability to discriminate between bulge and disc light; and, subsequently, (3) discrimination of bulge-disc galaxies from single-component galaxies.

1.2.3 Connection to galaxy kinematics

Using spectroscopy, a crucial new dimension is added to the picture of how galaxies form and evolve: the kinematics of gas and stars in galaxies. From the Doppler effect, the line-of-sight velocity distributions (LOSVDs) of stars and gas are encoded in the continuum, emission, and absorption lines of a galaxy's spectra (see Figure 1.8 for example). The velocities of distant galaxies then comprise two components: (1) a systemic velocity due to the expansion of the universe and the galaxy's bulk motion; and (2) the *internal* velocities of stars and gas in the reference frame of the galaxy. Therefore, if spatially resolved spectra can be obtained across the visible light distributions of galaxies, the internal motions of the stars and gas can be derived.

Spatially resolved spectroscopy of galaxies has a massive historical significance in astronomy and physics. Using spectra for sixty-seven HII regions (regions of active star formation where gas is being excited by young stars) over a broad range of radii extending from the centre of the nearby spiral galaxy, M31, Rubin & Ford (1970) derived the mean the velocity of the gas as a function of radius. These gas rotation curves showed that the gas velocities at large radii were too high and thereby required more enclosed mass to explain rotational stability. As referenced in Section 1.1.2, this discovery is now taken as one of the first in a long line of observational evidence for the existence dark matter. Long-slit spectrographs aligned with the long-axis of nearby disc galaxies have been used to support these findings by comparing with simulations (e.g. Rubin et al. 1980; Courteau 1997; Oman et al. 2015; and many others).

Today, the current state-of-the-art instruments and surveys such as CALIFA

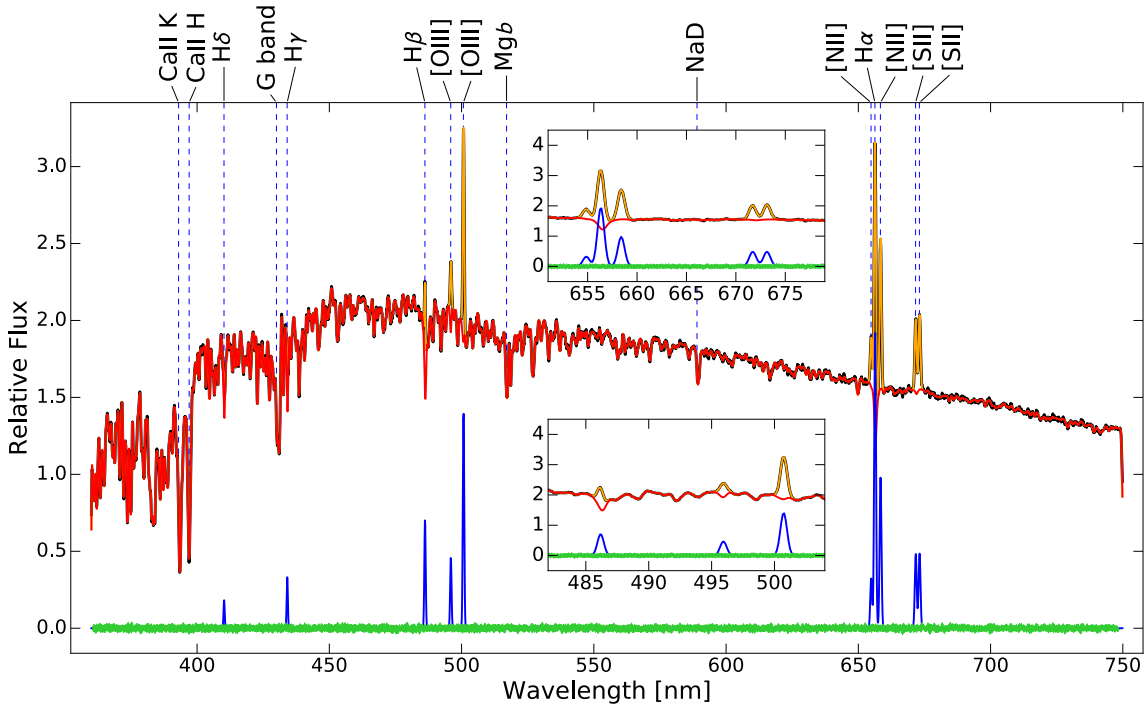

Figure 1.8 The Sun's spectrum with added gas emission lines modelled with the Penalized-Pixel Fitting software (PPXF, Cappellari 2017). The black line (hidden mostly by the model) is the observed spectrum with emission lines added. The red line shows the model fit to the continuum and absorption lines. The orange lines show the model fit to the emission lines – which are modelled simultaneously. The green line shows the residual flux from modelling the continuum and absorption. The blue curves show the continuum-subtracted emission lines. Inset plots cover specific regions of emission and demonstrate that the widths and positions of these lines are accurately measured – vital to measurement of the LOSVD of gas. Similarly, modelling of the continuum and absorption features gives the stellar LOSVD. This figure is a reproduction of Figure 1 in Cappellari (2017) and has been reproduced in this thesis with permission from the author.

(Sánchez et al., 2012), MaNGA (Bundy et al., 2015), MUSE (Bacon et al., 2010), WEAVE (Dalton et al., 2012), KMOS (Sharples et al., 2004), and SAMI (Croom et al., 2012) are used to obtain 2D maps of spatially resolved spectra across whole galaxies and large samples thereof. The 2D velocity maps produced by these integral-field spectroscopy (IFS) instruments and their predecessors have given dynamical context to galaxy morphology. Stellar and gas discs are dynamically supported against inward gravitational collapse by their *rotation* (e.g. see Kormendy & Fisher 2008 and references therein).

The 2D maps in the new era of instruments go further – being able to characterize discs over a wider range of inclinations as well as misalignments between the rotational axes of gas and stellar discs (Barrera-Ballesteros et al., 2014; Taylor et al., 2018; Bryant et al., 2019). Such misalignments can be induced by external factors such as ram pressure stripping (Kronberger et al., 2008) or gas-rich mergers (Naab & Burkert, 2003), rapid gas accretion (van de Voort et al., 2015), or internal instabilities such as bars or spiral arms (Binney & Tremaine, 2008; Sellwood, 2013). On the other hand, elliptical galaxies and the bulges of spiral galaxies can possess little or no rotation. These systems are supported by random motions of stars – also called *dispersion* support.

As with morphology, the transition between rotation- and dispersion-supported dynamical states is heavily dependent on galaxy environment (e.g. Cappellari et al. 2011) – with mergers expected to play a crucial role (Lynden-Bell, 1967; Toomre & Toomre, 1972; Toomre, 1977; Negroponte & White, 1983; Barnes, 1988; Hernquist, 1992; Hopkins et al., 2008b, 2009). But what exactly does this evolutionary transition look like? And how might we identify galaxies in these transitionary states? The Universe offers a frustratingly static view of galaxy interactions – yet merging is one of the core tenets of morphological and kinematic transformation in a ΛCDM Universe.

Similar to photometric imaging, a number of approaches are used to reduce the (often complex) kinematic structures of galaxies to a small number of parameters. For example, a single parameter, $\lambda_{\rm R}$, can be used to describe the degree of ordered rotation in a galaxy relative to the dynamical support it receives from both rotation and dispersion (Emsellem et al., 2007; Jesseit et al., 2009). This parameter is used, for example, to distinguish between galaxies that may be visually similar but dynamically very different. In particular, the distinction between spirals and fast- and slow-rotating elliptical galaxies gives rise to a strong relationship between the envi-

ronment of a galaxy (the number density galaxies in its local neighbourhood) and its morphological and kinematic properties (Cappellari et al., 2011). These observational results imply that dispersion-supported (dynamically hot) ellipticals are the final stage of galaxy evolution and are yielded through the strong gravitational and tidal forces in mergers and galaxy group and cluster environments. Consequently, one should expect also to see disturbances in the velocity and velocity dispersion fields of merging and recently merged galaxies.

Asymmetries and disturbances in the spatially resolved kinematics measurements may therefore provide an orthogonal basis to photometry for identifying galaxies in transformational stages such as mergers (Rampazzo et al., 2005; Kronberger et al., 2007; Yang et al., 2008; Shapiro et al., 2008; Oh et al., 2016; Bloom et al., 2017, 2018). One approach for *quantifying* such kinematic disturbances is the KINEMETRY tool (Krajnović et al., 2006). KINEMETRY is an algorithm that is designed to quantify the degree to which the kinematic maps of galaxies deviate from regular rotation or dispersion support through decomposition of the velocity and velocity dispersion profiles as harmonic series. As with morphological measurements of structural disturbance, KINEMETRY can be calibrated to delineate between the disturbances that are typical of merging galaxies from non-merging galaxies. However, calibration of the threshold for this distinction is of crucial importance. Currently, these thresholds are often calibrated based on visually classified samples of observed galaxies – which themselves carry many biases.

To better understand the role of mergers in the evolution of galaxy morphologies and kinematics (and galaxy evolution in general), it is useful to turn to numerical simulations in which the formation and evolution of galaxies can be monitored explicitly. Therein, with a sufficiently accurate model for galaxy formation and evolution and care in the way that physical properties and morphological and dynamical states are interpreted, important connections to observations are possible.

1.3 Numerical simulations of galaxy formation

Many strategies have been developed in the past several decades to numerically model the formation and evolution of galaxies with the overarching goal of reproducing observed properties, structures, and kinematics (see Somerville & Davé 2015 for a recent review). Crucially, validation of the models is determined through comparison with observations. (e.g. Abadi et al. 2003a; Brooks et al. 2011; Agertz et al. 2011;

Guedes et al. 2011; Christensen et al. 2014; Agertz & Kravtsov 2015; Furlong et al. 2015; Lange et al. 2016; Bottrell et al. 2017a). In this section, I briefly describe why hydrodynamical simulations are particularly useful from the perspective of studying structural and kinematic transformations in galaxies.

1.3.1 Hydrodynamical simulations

The most direct way to derive quantitative predictions from a theoretical model is to numerically track the coevolution of dark and baryonic matter across cosmic time starting with slight density inhomogeneities in the early Universe as observed in the CMB. This tracking is accomplished with hydrodynamical simulations (e.g. Katz et al. 1992; Navarro & White 1994; Katz et al. 1996; Weinberg et al. 1997; Murali et al. 2002; Springel & Hernquist 2003; Kereš et al. 2005; Ocvirk et al. 2008; Crain et al. 2009; Croft et al. 2009; Schaye et al. 2010; Oppenheimer et al. 2010; Vogelsberger et al. 2012). Self-consistent tracking of the motions of both dark and baryonic matter allows predictions to be made about the structures and kinematics of the stellar and gas components and their relationships to other galaxy properties (Abadi et al., 2003a,b; Governato et al., 2004; Agertz et al., 2011; Sales et al., 2012; Marinacci et al., 2012; Kereš et al., 2012; Torrey et al., 2012b) – which can then be benchmarked against observations.

Numerical simulations are therefore useful tools for interpreting the observed properties of galaxies because they offer *foreknowledge* of the evolutionary histories that are encoded in galaxy morphologies and kinematics. The latest generations of cosmological hydrodynamical simulations (e.g. Illustris - Vogelsberger et al. 2014b, EAGLE - Schaye et al. 2015, FIRE - Hopkins et al. 2014, APOSTLE - Sawala et al. 2016, IllustrisTNG - Pillepich et al. 2018a) and high-resolution galaxy interaction simulations (e.g. Moreno et al. 2015; Sparre & Springel 2017; Moreno et al. 2019) are designed specifically to allow straight-forward comparisons between observed and simulated galaxies.

In particular, with respect to galaxy mergers, the simulations allow one to connect morphological and kinematic disturbances to a specific stage in a merger ($t - t_{\rm coalesce}$, for example, where $t_{\rm coalesce}$ is the time at which the central black holes of two galaxies coalesce). Explicitly tracking the morphological and kinematic evolution of simulated galaxies and mergers provides a basis for: (a) identifying interacting galaxies observationally; (b) constraining the time window in which markers for interaction

are observable; and (c) linking observed galaxies to particular stages in their interactions. These points may require strict continuity in the way that morphologies and kinematics are characterized in simulated and observed galaxies. I introduce the topic of continuity in analyzing simulations and observations in the next section. In Chapter 3, I address the subject in detail.

1.3.2 Synthetic observations

Today, a long-standing gap between observational and theoretical galaxy astronomy narrows. Methods which place simulations directly in the context of observational astronomy make it possible to characterize observed and model-generated galaxies with unprecedented even-handedness. In particular, sufficiently realistic synthetic observations of galaxies from simulations can, in principle, be used to *calibrate* observational methods and models – given that the status and histories of simulated galaxies are known *a priori*.

The capacity to generate synthetic observations of galaxies from hydrodynamical simulations was revolutionized by modern radiative transfer codes (e.g. Jonsson 2006; Jonsson et al. 2010; Robitaille 2011; Baes et al. 2011; Camps & Baes 2015). Within a hydrodynamical simulation, groups of stars and gas are represented by respective particles. Each stellar particle can be assigned a spectral energy distribution (SED) similar the one shown in Figure 1.8. However, the spectrum that would be given to the particle in the simulation accounts for the light from a full stellar population rather than a single star. Rays can then be projected from every star particle to a camera placed anywhere in the simulation. Features such as absorption by interstellar dust grains, emission lines, absorption lines, and the effect of gas velocities on these lines can be incorporated in the ray tracing to produce realistic spatially resolved spectra (a synthetic observation) – similar to the ones obtained using modern IFSs.

To produce images, the spectrum in each spatial pixel (spaxel) of the spectral datacube can be convolved with the response function for a desired telescope's optical system to give integrated intensities (e.g. Lotz et al. 2008; Torrey et al. 2015; Rodriguez-Gomez et al. 2019). Producing synthetic images of galaxies from simulations makes it possible to analyze simulated galaxies using the same image-based quantitative characterization techniques used by observers (e.g. non-parameteric: Lotz et al. 2008; Snyder et al. 2015; Rodriguez-Gomez et al. 2019; parametric: Bottrell et al. 2017a,b; Rodriguez-Gomez et al. 2019). As such, radiative transfer has

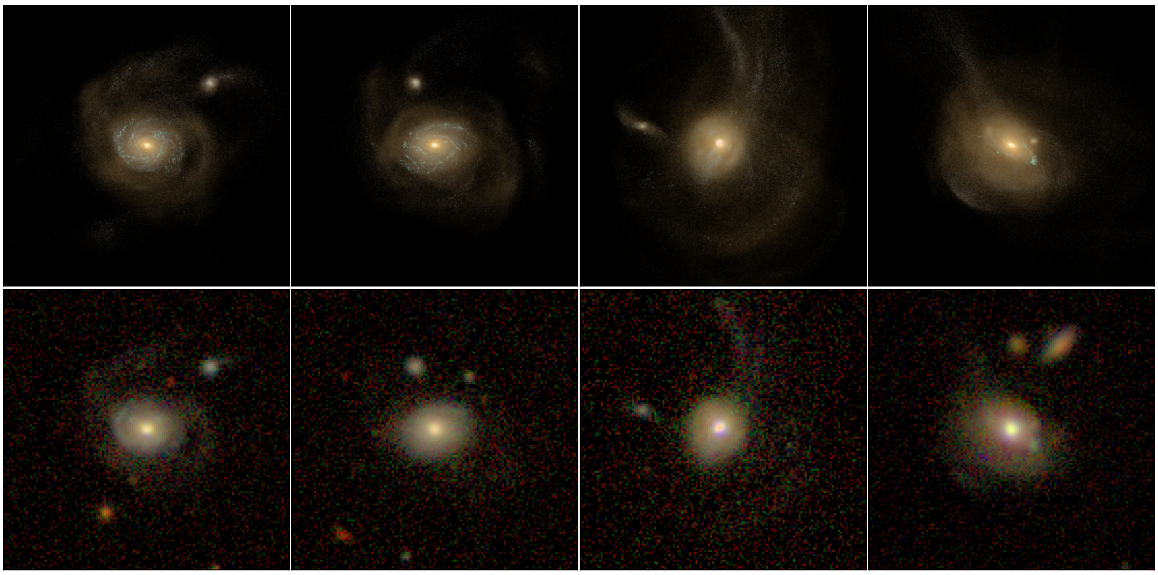

Figure 1.9 Colour-composite synthetic images of galaxies from a hydrodynamical simulation. The upper row shows *idealized* images of a single galaxy in different stages of its evolution produced using radiative transfer and the SDSS broadband imaging filters. The lower row shows corresponding survey-realistic synthetic images produced using a software I developed for Chapter 3. The synthetic images in the lower row incorporate real skies, other sources in the fields of view, and blurring from the atmosphere. An enlarged version of this figure and many other colour-composite synthetic images constructed for Chapter 3 of this thesis can be found in Appendix F.

revolutionized the capacity with which the structures of simulated galaxies can be consistently compared with observations.

Further *realism* can be incorporated to the images to ensure consistent observational limitations and biases. Synthetic images can be inserted into real or Gaussian skies and spatially convolving with an appropriate Point-Spread Function (PSF), for example. Figure 1.9 shows the effects of incorporating real skies and atmospheric blurring on the visible features of galaxies. The strategy of using *realistic* synthetic observations has particular importance in the study of galaxy interactions. For example, the defining theoretical predictions and visually affirmable features of galaxy pairs and merger remnants include stellar tidal arms, counter arms, streams, bridges, and tails (e.g. theory: Toomre & Toomre 1972; Barnes 1988; Barnes & Hernquist 1992; Howard et al. 1993; Gerber & Lamb 1994; Barnes 2011; observation: Brinchmann et al. 1998; Abraham et al. 2003; Conselice 2003; Lotz et al. 2004; Bundy et al. 2005; Casteels et al. 2013). In principle, these features would make it possible to identify

interacting galaxies and recent merger remnants both visually and quantitatively. In practice, the features are often faint with respect to their host galaxies and sensitive to limitations from sky brightness, spatial resolution, and crowding by nearby projected contaminants in an image. Therefore, the level of realism that is incorporated into synthetic observations of galaxy interactions is exceptionally important insofar as (1) the *observability* of simulated features and their observable timescales can be determined and (2) a model that is calibrated based on simulations can be justifiably applied to observations.

As stated in Section 1.2.3, spatially resolved kinematic measurements provide a complementary basis for characterizing galaxies and their interactions. Synthetic kinematic measurements can be generated for both gas and stars using radiative transfer or, less rigorously, by projecting the positions and line-of-sight velocities of the respective particles onto a two-dimensional grid. A recent study of evolution of stellar and gas kinematics in simulated galaxy mergers has shown promising results – with KINEMETRY measurements demonstrating strong sensitivity to merger stage, the relative masses of interacting galaxies (mass ratio), and the orbits of the merger (Hung et al., 2016). However, apart from this study, there is a paucity of studies that aim to systematically constrain the triggers and time-scales on which kinematic indicators of galaxy interactions are visible. Furthermore, Hung et al. (2016) do not attempt to produce *survey-realistic* synthetic observations (i.e. maps that would be observed as though the simulated objects had actually been observed with a real survey instrument). Consequently, the degree to which one could expect to see such sensitivity in real observations is unclear. The limitations of this *realism* bias are so far completely undetermined. Only by producing kinematic maps of the same interactions using both survey can the observability of kinematic indicators of interaction stage and other properties of mergers be tested.

The main take-away of this section should be that making even-handed comparisons to observations *is not* the only application of forward-modelling simulated galaxies. The other exceptionally interesting application is to use synthetic observations to establish connections between the *observable* features of galaxies and their evolutionary statuses and histories. In particular, Chapter 3 of this thesis explores whether the stages of galaxy mergers can be inferred from realistic images. Intuitively, making models for such complex connections would benefit from exploiting functions which are highly flexible and non-linear.

1.4 Characterizations with deep learning

In the previous sections of this chapter, I have focused on how galaxies can be characterized visually and quantitatively based on the topology of their projected surface brightness and velocity distributions. I have emphasized that (1) the morphologies and kinematics of galaxies are connected intricately to their evolution and (2) these connections can be informed through realistic synthetic observations of galaxies from hydrodynamical simulations. In this section, I discuss how deep learning (a branch of machine learning) can be used to facilitate such connections.

1.4.1 Machine learning primer

In principle, the task of machine learning is no different from any other statistical method. The goal is to construct a model which relates a set of outcomes to a corresponding set of data. However, machine learning is distinguished by the focus that a *trained* model, which relates *known* outcomes to some data, should be used to infer outcomes from data to which (1) the model has not yet been exposed and (2) the outcomes are not known *a priori*[5]. Breaking the process into two steps, the first task in machine learning is to learn the function, $\mathbf{y} = f(\boldsymbol{\theta}, \mathbf{X})$, that relates the dataset, $\mathbf{X}$, to the set of known targets, $\mathbf{y}$. For example, in linear regression of a dataset of d dimensions (often called *features* in machine learning parlance), $f(\boldsymbol{\theta}, \mathbf{X}) = \mathbf{X}\,\boldsymbol{\theta}$ in matrix notation – where $\boldsymbol{\theta}$ is a one-dimensional vector of length d (or $d+1$ with a normalization term). The *learning* is the optimization of the parameters $\boldsymbol{\theta}$ of the model to produce a set of optimized estimators, $\hat{\boldsymbol{\theta}}$. In the second step, the learned function is deployed on new data to make predictions:

$$\hat{\mathbf{y}} = f(\hat{\boldsymbol{\theta}} | \mathbf{X}_{\rm train}, \mathbf{X}) \tag{1.1}$$

where $\mathbf{X}_{\rm train}$ specifically denotes the training data and $\mathbf{X}$ is some other data upon which the trained model is to be applied and for which predictions are to be made.

[5]Here, I am focusing on a specific class of machine learning approaches called *supervised learning* – in which a model is trained using some data with a set of given target outcomes. This type of learning is distinct from *unsupervised learning* in which the model learns from data that has not been labeled, categorized, or classified. The core goal of making predictions for new data is shared. However, given the differences in these approaches, the types of problems that supervised and unsupervised learning can address are therefore also distinct. Supervised learning is divided into classification and regression problems – whereas unsupervised learning is used for clustering and association problems.

In each case, the hat symbol generally denotes a prediction or *estimator*. Just as the $\hat{\mathbf{y}}$ is the prediction of the optimized model for some data, $\hat{\boldsymbol{\theta}}$ is a prediction for the optimal model parameters given the training data.

This general formalism encapsulates the arbitrarily complex models that can be optimized through machine learning. For example, $\hat{\boldsymbol{\theta}}$ may describe the weights on each feature in linear regression, the cuts which give the maximum information-gain in a decision tree, or the highly non-linear combination of parameters in an artificial neural network. Such data-driven models and other information-driven models (such as models driven by simulated data) can be used to tackle problems covering an enormous range of complexity. In particular, given that we are entering an era of extraordinary data-abundance in which human-interaction with full datasets is impractical, machine learning is now being used ubiquitously in astronomy: image quality assessment (e.g. Teimoorinia et al. 2020a); star and galaxy classification (e.g. Nakoneczny et al. 2019; Walmsley et al. 2019; Cheng et al. 2020; Lucey et al. 2020); detection of weak and strong gravitational lensing (e.g. Jacobs et al. 2019; Ribli et al. 2019b; Teimoorinia et al. 2020b); transient object detection (e.g. Muthukrishna et al. 2019); and inference of cosmological parameters (e.g. Ribli et al. 2019a; Giusarma et al. 2019) to name a few.

1.4.2 Convolutional neural networks

Deep learning is a branch of machine learning in which models generally comprise a series of layers which extract increasingly high-level features of the raw data (Dechter, 1986; Lecun et al., 2015). This feature extraction is facilitated through the connections created between the output from each layer and the functions deployed by the next layer in the network. A class of deep learning models that has been particularly widely adopted in astronomy is the convolutional neural network (CNN, LeCun et al. 1995). Given CNNs are used extensively in Chapter 3 of this thesis, I describe the salient properties of CNNs here.

Figure 1.10 shows a simple CNN architecture. In general, CNNs comprise a series of layers which can be broadly divided into *feature extraction* layers and *classification* layers. Feature extraction layers perform operations which are designed to identify topological features in the data that are useful to the classification task. This is accomplished first through convolution, activation, and extraction operations applied to the data. For example, the first layer (C1) of the CNN in Figure 1.10 is a convolution

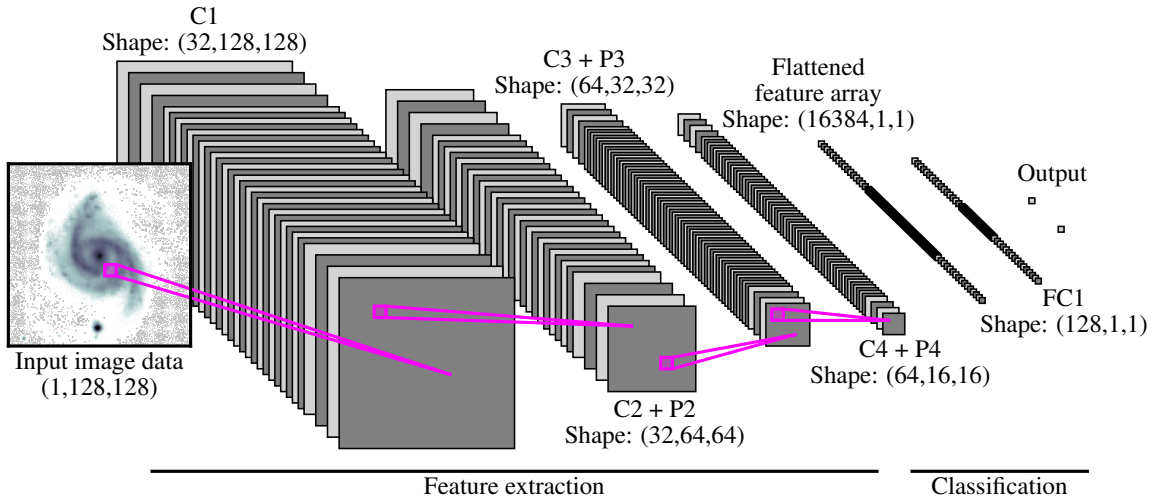

Figure 1.10 An example schematic of a simple convolutional neural network (CNN) architecture for a binary classification task. CNNs comprise a series of layers divided broadly into feature extraction layers (left group) and classification layers (right group). The classification layers are also widely called *dense* and *fully-connected* layers. The names of each feature extraction layer denote which operations are performed: Convolution (prefix C); Pooling (prefix P). The output from the feature extraction layers are flattened into a one-dimensional "feature array" for which every element is connected to every neuron in the Fully Connected (FC) layer. Similarly, every neuron in the FC layer is connected to each possible output class. The shape of the output of each layer's operations is indicated below each layer's name. The output classes for this binary classifier could be, for example, *spiral* and *elliptical* morphological classifications. The input image is of a galaxy in the Canada France Imaging Survey.

layer which comprises 32 convolution kernels which are applied to the input data. The elements of the kernel matrices are typically randomly initialized but are refined through training to amplify useful features, spatial correlations, and anti-correlations in the data for subsequent layers. The output of each convolution is typically followed by the application of a non-linear activation function. One common function used for non-linear activation is the Sigmoid function, $A(x) = (1 + e^{-x})^{-1}$, which has the potentially desirable property of squashing the input to a value between zero and one. Another is the rectified linear unit (ReLU), $A(x) = \max(0, x)$, which has been shown to greatly accelerate training speeds due its very simple form (Krizhevsky et al., 2012). These activations functions allows the model to create highly non-linear connections between the data and the targets.

In the second layer of the CNN in Figure 1.10, C2+P2, a new set of 32 convolution kernels is applied to each image in the data volume extracted by the previous layer. The resulting images are combined linearly and then passed to the non-linear activation function. If this process were to continue with convolutions and activations only, the model would have limited capacity to *extract* the dominant features for subsequent layers that are useful for connecting input data to targets. The Pooling (P) operation which commonly follows a convolution and activation operations in feature extraction layers serves two crucial functions in CNNs: (1) extracting features that have been amplified by the convolution and activation operations and, subsequently, (2) discarding or suppressing data which is deemed uninformative – thereby decreasing the computational demand by reducing the dimensionality data. Pooling operations can be accomplished by extracting the maximum (Max Pooling) or average (Average Pooling) of every $(n \times n)$ spatial block of elements in the data volume. For example, P2 takes the (32, 128, 128) output from the C2 convolution and activation operations and pools every (2×2) spatial block to reduce the output data volume to (32, 64, 64) – a factor of n^2. In summary, the process of feature extraction is the combination of many layers of convolution, activation, and reduction operations aimed at identifying and extracting features of the data which are valuable to the classification component of the CNN. The output from the feature extraction layers is typically flattened into a one-dimensional feature array which is passed to the classification component.

The set-up of the classification component is the same as a traditional artificial neural network such as a multi-layer perceptron. For the CNN in Figure 1.10, this component comprises the flattened output from the feature extraction layer, a single Fully Connected layer (FC), and the output layer (also called the *classifier* in the

case of a classification model). Fully connected layers comprise a series of *neurons* that are each connected to every element of its input and perform linear combination and non-linear activation operations. in Figure 1.10, the single fully connected layer comprises 128 such neurons. Finally, each neuron in the output layer also performs a linear combination of the output generated by the previous fully connected layer. This is followed by some form of normalization such that the output values for these neurons encode the "likelihoods" that the model assigns to each possible class based on its training.

A CNN's trainable parameters generally include the elements of the convolution kernels, weights in linear combinations, and parameters of the activation functions. These parameters are adjusted during training through *backpropagation* (LeCun et al., 1989). In training, stochastically selected batches of labelled training images are passed through the model. The difference between the predicted outcomes and the true outcomes (the error) can be quantified through some loss function (e.g. mean squared error for regression and cross-entropy for classifications). Given that all linear and non-linear functions used in CNNs are chosen to be differentiable (with computational ease), the gradients of this loss with respect to each parameter can be propagated back through the model from the output layer. These gradients provide the basis for parameter adjustments. In theory, increasing the number of feature extraction and classification layers in a CNN should allow one to asymptotically approach some minimum loss allowed by the noise in the data and other choices in the model architecture (such as the dimensions of the convolution kernels, for example). In practice, the propagation of gradients through very deep models leads to a "vanishing gradient" problem in which the error grows with model depth. This problem can be solved with more sophisticated model architectures such as Residual CNNs (He et al., 2015).

CNNs are now being widely adopted for galaxy classification based on photometric images (e.g. Huertas-Company et al. 2015; Domínguez Sánchez et al. 2018; Fischer et al. 2019; Domínguez Sánchez et al. 2019; Cheng et al. 2020). The training of these CNNs is made possible by the extra-ordinary foundational work of expert visual classifiers (e.g. Nair & Abraham 2010; Kartaltepe et al. 2015) and by the collective efforts in citizen science projects (e.g. Galaxy Zoo; Lintott et al. 2008; Willett et al. 2013; Simmons et al. 2017). The main body of literature using deep learning for morphological classifications is focused on one particular application of these models: to mitigate the need for human visual classifications in the face of the unremitting

growth of modern galaxy surveys. For example, the forthcoming Large Synoptic Survey Telescope will collect roughly as much imaging data in a single night as SDSS did in its operational lifetime (Ivezić et al., 2019). Consequently, there is certainly demand to deploy trained models on these enormous volumes. However, it is equally (if not more) interesting to see how far the connections made by deep learning models can be pushed.

One particularly interesting and novel application of deep learning is to model the complex and highly non-linear connections between the *observable* properties of galaxies (e.g. their morphologies and kinematics) to information about their evolutionary statuses and histories. For example, can the observed morphological or kinematic features of a galaxy be linked to its merger status or history? Furthermore, does the observationally accessible information provide a basis from which the *details* of recent and ongoing merger events be recovered: such as the mass ratio between the merger constituents, their gas fractions, initial morphologies, and orbital parameters? Given that none of this information is *a priori* known observationally, building and calibrating such deep learning models may benefit greatly by drawing upon state-of-the art numerical simulations of galaxy formation and evolution – where the histories and statuses of galaxies are known truths. In this approach, a deep learning model would be calibrated using synthetic observations and known targets from the simulations. Then, given that the same morphological and kinematic features should be expected in both synthetic and real observations, the model can be ported to real observations to estimate the same targets for real galaxies.

1.5 Thesis outline

This thesis is concerned with the morphological and dynamical indicators which (1) link galaxies to their evolutionary histories and (2) can be used to identify galaxies currently undergoing transformations.

Numerical simulations show that presence and relative prominence of stellar bulge and disc components in galaxies are connected to particular evolutionary pathways and events. Observationally, these structures can be characterized quantitatively through bulge-disc surface brightness decompositions. One of the most widely-used observational catalogue of bulge and disc structural measurements is that of Simard et al. 2011 for 1.12 million galaxies in the SDSS Legacy Survey images. Given that bulges and discs encode information about a galaxy's evolution, it is important to

know whether these measurements are robust. Chapter 2 of this thesis exploits a section of the SDSS with greatly improved photometry to rigorously test the robustness of bulge, disc, and other structural measurements against photometric depth.

The strong gravitational forces involved in mergers drive transformations in the morphologies, dynamics, and internal properties of galaxies. However, the *relative* role of mergers in driving these changes requires pure and complete samples of galaxies undergoing interactions. Galaxy merger samples that are generated through visual classification or quantitative analysis may suffer from a number of biases – most of which derive from the absence of *a priori* knowledge whether a galaxy is undergoing a merger or not. Chapter 3 combines hydrodynamical simulations, synthetic observations, and deep learning to examine the whether the morphological structures in galaxy images can be used to estimate whether a galaxy is (a) isolated; (b) currently undergoing an interaction; and (c) recently merged. In particular, I focus on the sensitivity of the estimates to the realism in the synthetic observations.

As outlined above, the effects of an on-going and recent merger can also be imprinted in a galaxy's kinematics. Kinematics may therefore provide an orthogonal basis for the identification and characterization of galaxy mergers (e.g. by stage and/or mass ratio). To calibrate a model for selecting mergers based on kinematic features, it is again useful to turn to simulations – in which these properties are known. However, there is currently no framework for generating synthetic kinematic observations of galaxies from hydrodynamical simulations that directly incorporate the instrumental and atmospheric responses. Chapter 4 presents a code for generating highly-realistic synthetic integral-field spectrograph observations of galaxies from simulations. The code is designed to produce survey-realistic velocity and velocity dispersion maps of galaxies from simulations (e.g. compared to the MaNGA and SAMI IFS surveys). With this framework, a deep learning model can be trained to exploit observable kinematic features to estimate merger stage and potentially other merger properties such as mass ratio.

This thesis is concluded in Chapter 5 with a summary of (1) the salient results of these studies, (2) the remaining challenges to be overcome and (3) the roles of current and future instruments in facing these challenges.

Chapter 2

Stripe 82 deep morphologies

Preamble

The work presented in this chapter concerns the quantitative, morphological characterization of galaxies from images. Specifically, I investigate the role of imaging depth (surface brightness limits) on morphological measurements by comparing measurements derived from deep images with those from relatively shallow images for the same set of galaxies. The work presented in this chapter is published in Bottrell et al. (2019b)[1] but includes more detailed descriptions of the deep image construction (Section 2.2.3) and corresponding propagation of errors (Section 2.2.4).

2.1 Introduction

Morphology is the visual connection to the formation and evolutionary histories of galaxies. From a theoretical perspective, observed stellar structure is tied intricately to these histories. Stellar discs are built-up from gas that either collapsed gravitationally from a protogalactic cloud in a rotating dark matter halo (White & Rees, 1978; Fall & Efstathiou, 1980) or was accreted at least semi-coherently with the rotation of a pre-existing galaxy (Kereš et al., 2005; Sales et al., 2012; Stewart et al., 2013; Danovich et al., 2015). Amongst stellar bulges, the observed agreement between classical dispersion-supported bulges and elliptical galaxies across a number of key scaling relations (e.g., Kormendy 1977; Faber & Jackson 1976; Djorgovski & Davis 1987; Dressler et al. 1987; Fisher & Drory 2008; Gadotti 2009; Fisher & Drory

[1] https://doi.org/10.1093/mnras/stz855

2010; Kormendy & Bender 2012), the black-hole mass vs. bulge velocity dispersion relation (Kormendy et al., 2011; Kormendy & Ho, 2013), and the similarity of their stellar populations (Moorthy & Holtzman, 2006) suggest a similar and coordinated formation mechanism. Mergers are widely regarded as key suspects due to broad theoretical compatibility with these observed similarities (Lynden-Bell, 1967; Toomre & Toomre, 1972; Toomre, 1977; Negroponte & White, 1983; Barnes, 1988; Hernquist, 1992; Hopkins et al., 2008b, 2009; Taranu et al., 2013; Rodriguez-Gomez et al., 2017; Martin et al., 2018). Meanwhile, a class of morphologically and kinematically disc-like pseudo-bulges which do not neatly subscribe to trends shared by classical bulges and ellipticals (see Kormendy & Kennicutt 2004 for a review) are observed as downward dips in the stellar velocity dispersion at the centres of galaxies and are often linked to secular formation through disc instabilities (Emsellem et al., 2001; Márquez et al., 2003; Falcón-Barroso et al., 2006; Peletier et al., 2007). Accurate characterization of morphology is therefore of enormous astrophysical importance because the allocation of stars to the bulge and disc and their structures encode the dominant physical processes in a galaxy's formation (e.g., Berg et al. 2014; Bluck et al. 2019).

The dominance of a stellar disc or bulge forms the basis of most visual morphological classification schemes while additional stellar components can motivate splitting of the main track into parallel sequences (Hubble, 1926, 1936; de Vaucouleurs, 1959a; Sandage, 1961). However, purely visual morphology classification has limited capacity for comparison with theoretical predictions. Intrinsic subjectivity in visual classification combined with the enormous scale of modern galaxy surveys demands robust and quantitative characterization of galaxy morphology. This demand has been met with a wealth of tools for accurately characterizing the global and/or component structures based on galaxy surface-brightness distributions (Peng et al., 2002; Simard et al., 2002; de Souza et al., 2004; Lotz et al., 2004; Conselice, 2006; Méndez-Abreu et al., 2008; Vika et al., 2013; Ciambur, 2015; Robotham et al., 2017). These tools enable routine and reliable separation of bulge and disc light in tens to hundreds of thousands of galaxies (e.g., Allen et al. 2006; Benson et al. 2007; Simard et al. 2011; Lackner & Gunn 2012; Kelvin et al. 2012; Meert et al. 2015). Indeed, marriage of visual and quantitative morphologies on such large scales may provide even more accurate descriptions of galaxy structures that go beyond the bulge and disc. Recent work by Kruk et al. (2018) has shown that it is possible to use statistics for visual morphologies harvested from the Galaxy Zoo citizen-science project (Lintott et al., 2008; Willett et al., 2015) as priors on the forms of models used in quantitative photo-

metric decompositions. Given that the same quantitative and visual methods can now be employed on realistic mock-observations of galaxies produced in large-volume cosmological hydrodynamical simulations (e.g., Vogelsberger et al. 2014b; Schaye et al. 2015; Pillepich et al. 2018b), the capacity with which observed structures may be compared with theoretical predictions has never been better (e.g., Snyder et al. 2015; Bottrell et al. 2017a,b; Dickinson et al. 2018; Rodriguez-Gomez et al. 2019).

But just how robust are these quantitative measurements? If the bulge-to-disc (B/D) or bulge-to-total (B/T) light ratios encode the above-mentioned formation histories, then accurate measurements of these quantities and their limitations within large galaxy surveys is vital. A classic test of the sensitivity of measured parameters to various metrics for image quality is to insert analytic galaxy bulge+disc models into real survey fields and to evaluate their recovery (e.g., Simard et al. 2002). However, these tests do not capture the systematic and random errors that arise from additional substructure in the bulge and disc components or, indeed, structures that are neither bulge nor disc. Instead, comparisons of structural measurements obtained for real galaxies in images of varying quality do capture these biases and enable a more critical evaluation of the sensitivity of measured quantities (e.g., Lotz et al. 2004, 2006; Lisker 2008). Thus far, no such image-quality comparison has been performed to evaluate the limitations for the measured properties of galaxies in the bulge/disc decomposition analysis of 1.12 million galaxies by Simard et al. (2011) using SDSS Legacy photometry (Abazajian et al., 2009). The deep co-added images offered for the SDSS Stripe 82 (Annis et al., 2014) provide the ideal basis for such a test – extending 1.6-1.8 magnitudes deeper than the Legacy single-pass images. Having been combined from images taken with the same instrument and resulting in images with similar spatial resolution, the deep co-adds offer a direct look at how the measured properties of galaxies respond to depth alone. Furthermore, one may expect that faint substructures that were previously inaccessible in Legacy are revealed in the deep images.

In this chapter, I characterize the extent to which structural measurements for galaxies derived from parametric two-dimensional surface brightness decompositions are robust to photometric improved depth. To do so, I compare structural measurements taken from existing catalogs using SDSS Legacy photometry (Simard et al., 2011; Mendel et al., 2014) to those in a new set of publicly available catalogs for galaxies in the SDSS Stripe 82 deep co-add images. Surface-brightness decompositions are performed using the GIM2D IRAF software (Simard et al., 2002). As such,

my comparison of *ugriz* structural measurements in deep and shallow images is entirely self-consistent. This chapter is structured as follows. Section 2.2 provides a description of the salient features of the SDSS image acquisition, features, and definitions that are relevant to creation of the co-adds. I also describe the galaxy sample selection and provide a characterization of typical spatial resolution in the co-adds. Section 2.3 describes the decomposition models, methods, and catalogs. Section 2.4 compares the global and component structures, statistics, and asymmetries of galaxies in the deep and shallow images. My main results are summarized in section 2.5. All rest-frame quantities and distance measures computed for the catalogs and for this chapter assume a $(H_0, \Omega_m, \Omega_\Lambda) = (70 \text{ km/s/Mpc}, 0.3, 0.7)$ cosmology and magnitudes are quoted in the AB magnitude system (Oke & Gunn, 1983).

2.2 Data

In this section I describe the imaging methods, construction of the co-add images, corresponding propagation of uncertainties, and galaxy selection criteria. I also provide a brief characterization of image quality in the co-add images with respect to the single-epoch images.

2.2.1 The Sloan Digital Sky Survey

SDSS imaging was performed with the dedicated 2.5-m wide-field telescope at Apache Point Observatory (APO) in New Mexico. A total of around 45,000 square degrees of science quality images across five broad bands were produced during operation. Taken together, the images cover a unique footprint of 8,423 square degrees of sky. The telescope and distinctive terminology used to describe its imaging are presented in Gunn et al. (2006) and Stoughton et al. (2002), respectively. Certain features and definitions relevant to the co-add image construction and characterization warrant description here.

The SDSS APO telescope focal plane is comprised of 30 charge-coupled device (CCD) cameras organized in six camera columns each 13.5 arcmin wide and separated by 11.7 arcmin from neighbouring columns. Five CCD detectors, one for each of the *ugriz* filters described by Fukugita et al. (1996), make a camera column (*camcol*). Each CCD detector's dimensions are 2048x2048 pixels with pixel scale 0.396 arcseconds/pixel. Imaging *strips* are taken in drift-scan mode (time-delayed integra-

tion, TDI) along great circles forming six parallel *scanlines*, each 13.5 arcmin wide. Two interleaving strips, filling in the gaps and overlapping slightly between scanlines, make a single *stripe* that is 2.5 degrees in width (York et al., 2000).

Operating in drift-scan mode, it takes an object 53.9 seconds to move from one end of a CCD detector to the other. This translates to the effective exposure time for every $0.396''$x$13.5'$ sliver of sky along a scanline in each bandpass. Scanlines are arbitrarily divided lengthwise into 9.8x13.5 arcmin^2 segments called *frames* with 128 pixels overlapping between neighbouring frames. A frame corresponds to the readout from a unique CCD detector (bandpass) in a camera column over 53.9 seconds. A *field* consists of the frames from the five *ugriz* filters that cover the same section of sky. A *run* is a single continuous image observing scan along a length of a strip. Lastly, an imaging run can be processed using more than one software version and/or calibration denoted by the *rerun* identifier. Putting it all together, a processed SDSS image produced by a unique CCD detector is identified by five properties: the observing *run* in which it was taken, the *rerun* processing ID, the *camcol* (or scanline) of the run, the *field*, and the *ugriz* bandpass used in the imaging.

The SDSS image processing pipeline is performed in two steps. First, astrometric calibrations are determined and assigned to the images (Pier et al., 2003). Second, the brightnesses, shapes, and positions of detected objects are determined (Lupton et al., 2001, 2002; Stoughton et al., 2002). The measured shapes of stars in a field and its four flanking fields (two along the scanline in which the field resides and one from each neighbouring scanline) are used to construct a model for the spatial variation in the PSF. Each image field then has a corresponding unique PSF file from which the PSF can be reconstructed for any bandpass and pixel location. In Data Release 6 (DR6) and earlier, photometric calibration is performed on the basis of 158 primary standard stars first measured using the US Naval Observatory 40-in telescope in Flagstaff, Arizona (Smith et al., 2002) and a set of 1520 41.5x41.5 arcmin^2 secondary standard *patches* with four patches allocated to each stripe. The primary standard stars and secondary standard patches are observed with the SDSS 0.5-m Photometric Telescope (PT) at the APO site (Tucker et al., 2006; Ivezić et al., 2004). Primary standard stars are used to determine the atmospheric extinction for each observing night and so must be measured coincidently with observing scans.

Photometric zero-points[2] are determined by calibrating the the survey images to the secondary standard patches which can be measured on any photometric night.

Data Release 7 (DR7) and later releases use a simplified and improved calibration strategy. First, the relative photometry over the entire survey is determined using the overlap regions between adjacent observation runs in a process called "ubercalibration," (Padmanabhan et al., 2008). Relative photometry in each of the SDSS filters is determined independently. Second, the *ugriz* photometric zero-points are computed by forcing stellar aperture magnitudes to agree on average with those in DR6. Relative calibration errors from ubercalibrated zero-points are $\sim 1\%$ in *griz* bandpasses and $\sim 2\%$ in the *u*-band.

DR7 imaging is complimented by a 640-fibre-fed pair of multi-object double spectrographs with coverage from $3800 - 9200$Åat a resolution of $\lambda/\Delta\lambda \simeq 2000$. The SDSS DR7 Main Galaxy spectroscopic sample consists of a magnitude-limited sample of 928,567 galaxies complete to a Galactic extinction-corrected Petrosian magnitude limit $m_{\rm r,Petro} = 17.77$ (Schlegel et al., 1998; Strauss et al., 2002).

2.2.2 Stripe 82

The SDSS Stripe 82 resides at $-50^\circ < \alpha < +60^\circ$ and $-1.^\circ25 < \delta < +1.^\circ25$ along the celestial equator in the Southern Galactic Cap. Repeated imaging runs of Stripe 82 were carried out to enable stacking of images and reach fainter magnitudes. Images from Stripe 82 runs $125 \leq$ run ≤ 5924 were used in the Annis et al. (2014) stacks – numbering 123 runs in total (Legacy: York et al. 2000; Hogg et al. 2001; Supernova Survey, SNeS: Frieman et al. 2008a). This selection rejects a large number of the original 303 total runs that were taken under poor seeing, bright/moonlit sky, and/or non-photometric conditions. Individual fields of poor quality were then rejected based on quantitative cuts to the $r-$band seeing, sky brightness, atmospheric transparency, and number of standard calibrating stars[3]. The matching *ugiz* frames which did not satisfy the corresponding $r-$band quality cuts were also rejected. As such, the

[2]A photometric zero-point is the flux (or corresponding magnitude) at which exactly 1 count [digital number, DN] per second is registered by the detector. By calibrating zero-point fluxes against standard sources of known flux, they are used to convert detector counts to physical fluxes.

[3]These data are available for all fields covering Stripe 82 in the DR7 Stripe 82 Catalog Archive Server (CAS). The DR7 Stripe 82 CAS (along with any other Data Release) can be conveniently queried using Structured Query Language (SQL) in Python 2 with the `sqlcl` tool. Queries to the CAS for specific data releases can be made by changing the `default_url` parameter (e.g. stripe 82: `default_url` = '`http://cas.sdss.org/stripe82/en/tools/search/x_sql.asp`'

effective number of images used in the co-add stacks varies along the Northern and Southern strips from 15 to 34. The depth is therefore not perfectly homogeneous and varies at most by around 0.4 mag between the shallowest and deepest co-add fields. On average, the stacks go $\Delta(u, g, r, i, z) = (1.6, 1.8, 1.6, 1.6, 1.8)$ magnitudes deeper than the single-epoch SDSS Legacy images based on 50% completeness limits for point sources. For reference, the average 50% completeness limits for point sources in the single-pass Legacy images are $(u, g, r, i, z)_{\rm Legacy} = (22.3, 23.2, 23.0, 22.6, 21.1)$ magnitudes. These limits are determined by inserting and measuring recovery of artificial point-source models into the respective single-epoch and stacked image fields (Fliri & Trujillo, 2016).

Stacks were also constructed by Jiang et al. (2014) and Fliri & Trujillo (2016) (hereafter IAC) which each achieve 1.7-1.9 mag and 1.7-2.0 mag improved depth compared to the single-epoch images, respectively (again with stellar 50% completion limits as the metric) – improving slightly on the depth achieved in the Annis et al. (2014) stacks. Unlike Annis et al. (2014), the Jiang et al. (2014) stacks use all Stripe 82 runs with less strict cuts on input image quality and a slightly modified weighting scheme. In particular, use of images with significant moonlight or strong variations in sky brightness along the drift direction without overly degrading quality in the stacks was facilitated by their background subtraction method – which efficiently handles such cases. Fliri & Trujillo (2016) take a completely different approach in the IAC stacks. Instead of attempting to remove the sky, they deliberately place special emphasis on preserving low surface brightness structures on all spatial and intensity scales. Rejecting about 1/3 of the Stripe 82 runs with the poorest quality, their stacks offer the depth and resolution to examine stellar streams, ultra-diffuse galaxies, intracluster light, and even the dust filamentary structure of our Galaxy via optical cirrus.

The quality of the stacked images is sensitive to the combination procedure and therefore may be optimized for specific science objectives. In particular, the Annis et al. (2014) stacks were designed to enable processing of the co-adds using the SDSS standard measurement code, `PHOTO` (Lupton et al., 2001, 2002, 2012). For my purposes, this choice offers three main advantages over other stacking strategies for a small cost in depth: (1) the SDSS collaboration has thoroughly tested the `PHOTO` pipeline and characterized the uncertainties in the measurements made by its algorithms; (2) consistency between the methods and measurements for the single-epoch images and stacks; and (3) the data access architecture in the Catalog Archive Server

(CAS) and Data Archive Server (DAS) are conveniently consistent with the single-epoch images. Annis et al. (2014) also imposed strict quality cuts and an unforgiving quality-weighting scheme for the images considered in the co-adds (described in the next section). These factors facilitate the most straight-forward comparison between quantitive morphologies of galaxies in the Annis et al. (2014) stacks and the Legacy images.

2.2.3 Construction of co-add images

Construction of the Stripe 82 co-add images is described by Annis et al. (2014). However, several salient features of the combination procedure relevant to the propagation of errors used in the decomposition analyses warrant explicit description here.

All frames used in the stacks are scaled to a standard photometric zero-point[4], $\texttt{aa} = -m_{zp} = -23.90$ mag. The magnitude zero-point offset, Δm_{zp}, of a frame relative to this standard zero-point is related to the relative atmospheric transparency, T, by:

$$\Delta m_{zp} = (m_{zp} - \texttt{kk} \times \texttt{airmass}) - 23.90 = 2.5 \log_{10} T \tag{2.1}$$

where `kk` is the atmospheric extinction coefficient and `airmass` is the relative atmospheric path length with respect to the zenith. Thus, a frame may be re-scaled to the standard photometric zero-point magnitude, $m_{zp} = 23.90$, by multiplying the inverse of its relative transparency, T_i^{-1}. The number of counts measured in an unscaled and corresponding scaled input images are F_i and $F_{s,i}$, respectively, where $F_{s,i} = T_i^{-1} F_i$. Input images (scaled and unscaled) measure integrated counts over an exposure time, $t_{\rm exp} = 53.907$ seconds. The zero-point corrected fluxes are:

$$\left(\frac{f}{f_\circ}\right)_i = \frac{F_i}{t_{\rm exp}}\, 10^{0.4(\texttt{aa}_i + \texttt{kk}_i * \texttt{airmass}_i)} \tag{2.2}$$

$$\left(\frac{f}{f_\circ}\right)_{s,i} = \frac{F_{s,i}}{t_{\rm exp}}\, 10^{0.4(-23.90)} \tag{2.3}$$

Note that by equating 2.2 and 2.3 and using $F_{s,i} = T_i^{-1} F_i$ one arrives naturally at the definition for transparency from Eq. 2.1. Scaled input images are then combined

[4] `aa` is the value stored in the DAS for a given single-epoch field or co-add field and so its relation to the zero-point provided by $\texttt{aa} = -m_{zp}$ is explicitly shown for clarity. Though potentially leading to confusion, the supposed utility of storing the negation of the zero-point (rather than the zero-point itself) is that the flux calibration in Eq. 2.2 requires no negative sign in the exponent.

in a weighted mean to construct the stacks using a version of `Swarp` (Bertin et al., 2002) in conjunction with the existing astrometry from the `astrom` pipeline (Pier et al., 2003). The weights on each scaled input image are:

$$W_i = \frac{T_i}{\mathrm{FWHM}_i^2 \sigma_{sky,i}^2} \tag{2.4}$$

where FWHM_i and $\sigma_{sky,i}$ are the full-width at half-maximum (FWHM) of the PSF and sky noise per pixel in an input image. The weights are computed individually in each bandpass, ensuring that the frames with the highest quality seeing, atmospheric transparencies, and skies contribute most to the co-add stacks in each filter. The zero-point corrected flux in the co-add image is then:

$$\left(\frac{f}{f_\circ}\right)_c = \frac{F_c}{t_{\mathrm{exp}}}\, 10^{0.4(-23.90)} \quad \text{where} \quad F_c = \frac{\sum W_i F_{s,i}}{\sum W_i} \tag{2.5}$$

The co-additions produced 800 stacked fields in five bandpasses along each of the 12 scanlines (6 Northern and 6 Southern) for a total of 48,000 deep images. The co-adds were processed using a modified version of the standard SDSS measurement pipeline, `PHOTO`. The main modifications to the standard procedure were the methods for fitting the spatial variation of the PSF and computation of the gain, G_{eff}, in the co-add images. Using the calibrations, the co-add image counts are converted to count rates and re-scaled to a uniform photometric zero-point $\mathtt{aa} = -30$ mag. In SDSS calibrated linear flux units (maggies), this adopted zero-point gives 1 picomaggy per unit of intensity in the co-add images (the count rate in digital counts per second, [DN/s]). The count rate in the scaled co-add image, F_{sc}, is related to the zero-point calibrated flux by:

$$\left(\frac{f}{f_\circ}\right)_{sc} = F_{sc}\, 10^{0.4(-30)} \tag{2.6}$$

Equating 2.5 and 2.6 we arrive at the ratio of digital intensities in the scaled and unscaled co-add images:

$$\frac{F_{sc}}{F_c} \equiv c_\circ = \frac{1}{t_{\mathrm{exp}}}\, 10^{0.4(30+\mathtt{aa_c})} \quad \text{where} \tag{2.7}$$

$$\langle c_\circ \rangle = \frac{1}{t_{\mathrm{exp}}}\, 10^{0.4(30-23.90)} \approx 5.1092 \tag{2.8}$$

Because all input images are first calibrated using the same pipeline to $m_{zp} = -\texttt{aa} = 23.90$ mag, the extension of Eq. 2.7 to 2.8 is highly precise for all fields but the ratio is exactly determined using the calibrated zero-point $\texttt{aa}_c$ from `PHOTO` in Eq. 2.7. This calibration is not available in the CAS. However, its main use is to compute the effective gain for each co-add field which is accessible in the corresponding `fpAtlas` image header.[5]

2.2.4 Error propagation and effective gain

Accurate modelling and meaningful uncertainties in model parameters rely crucially on a correct interpretation of noise. As with a single-epoch image, the total variance in a calibrated co-add image pixel is well approximated by:

$$\sigma^2_{sc,\mathrm{tot}} = \sigma^2_{F_{sc}} + \sigma^2_{sc,\mathrm{sky}} = F_{sc}/G_{\mathrm{eff}} + \sigma^2_{sc,\mathrm{sky}} \tag{2.9}$$

where F is the background-subtracted source intensity measured in the pixel [DN/s] and G_{eff} is the (effective) gain [electrons/DN] of the image. The variance in a co-add image pixel intensity derives from noise contributions from the individual input images which are generally non-homogeneous in S/N. The sky variance, $\sigma^2_{sc,\mathrm{sky}}$, can be measured directly and locally around a particular galaxy in the co-add images. I use the approach of Simard et al. (2011) to estimate σ^2_{sky}. SExtractor (Bertin & Arnouts, 1996) is used to delineate between objects and sky as well as to deblend pixels from overlapping sources. The SExtractor parameters were calibrated in Simard et al. (2011) using several sensitive tests based on the results of subsequent bulge+disc decompositions with gim2d: the size luminosity relation of discs and the colour-magnitude diagrams and fibre colours of pairs in the Patton et al. (2011) pair catalog. The residual sky level is computed from a SExtractor segmentation image corresponding to the sky-subtracted co-add corrected image. A minimum of

[5]The *effective* gain is the conversion between the digital intensities stored in an image (which could be any linear unit of flux or intensity, calibrated or uncalibrated) and the number of photo-electrons which correspond to that digital intensity. Measurements of flux and their uncertainties for CCDs are driven by the statistics of photons and corresponding photo-electrons – not the digital units to which they are converted. Consequently, for digital images, the effective gain is necessary to exploit these statistics. The *true* gain is a property of the analog-to-digital converter (ADC) paired to a CCD and is often known or easily measured. However, flux calibration and stacking of images taken with different CCDs and ADCs (each with a given *true* gain) complicates the calculation of the effective gain of a stacked image. In Section 2.2.4, I address this topic in detail with respect to the Stripe 82 co-adds.

20,000 sky pixels are used to estimate the local background and variance, excluding pixels within four arcseconds of any source's pixels. I assume that the readout noise is accounted for in the Poisson noise of the background.

The per-pixel variance where a source's flux dominates, $\sigma^2_{F_{sc}}$, also follows Poisson statistics and can be determined using the pixel intensity, F_{sc}, accounting for the effective gain, $G_{\rm eff}$, for converting electrons to pixel counts. Computing the effective gain in a co-add image is complicated by the calibrations applied to the input images and their inhomogeneous weights in the stacking. Starting from the final stacks and working down, the scaled co-add image has units [DN/s] and calibrated photometric zero-point $\mathtt{aa}_{sc} = -30$ mag. The variance in source intensity in the scaled co-add can be equated to the variance in the unscaled co-add (i.e. the co-add image before it was re-scaled to $\mathtt{aa}_{sc}$), where the partial derivative term is the constant, $c_\circ$, from Eq. 2.8:

$$\sigma^2_{F_{sc}} = \left(\frac{\partial F_{sc}}{\partial F_c}\right)^2 \sigma^2_{F_c} = c_\circ^2 \sigma^2_{F_c} \tag{2.10}$$

The variance in the unscaled co-add image, $\sigma^2_{F_c}$, can also be expanded. Its variance is the sum in quadrature over the errors in the scaled input images.

$$\sigma^2_{F_c} = \sum \left(\frac{\partial F_c}{\partial F_{s,i}}\right)^2 \sigma^2_{F_{s,i}} = \sum w_i^2 \sigma^2_{F_{s,i}} \tag{2.11}$$

$$\text{where} \quad \left(\frac{\partial F_c}{\partial F_{s,i}}\right) = \frac{W_i}{\sum W_i} \equiv w_i \tag{2.12}$$

The flux in the unscaled co-add image is the mean of the scaled input image fluxes, $F_c = \sum w_i F_{s,i}$. Therefore the partial derivative for each term in the sum $\partial F_c / \partial F_{s,i}$ is equal to the normalized weight on the scaled input image, w_i. Finally, the variance in an individual scaled input image can be expressed in terms of the variance in the corresponding unscaled input image. But first, note that the source variance, $\sigma^2_{F_i}$ in an unscaled input image is equal to the background-subtracted intensity in counts, F_i, divided by the CCD gain, g_i [electrons/DN], analogous to Eq. 2.9.

$$\sigma^2_{s,i} = \left(\frac{\partial F_{s,i}}{\partial F_i}\right)^2 \sigma_i^2 = T_i^{-2} \sigma_i^2 = T_i^{-2} F_i / g_i = \frac{T_i^{-1} F_{s,i}}{g} \tag{2.13}$$

where the partial derivative in Eq. 2.13 is identified as the inverse transparency from Eq. 2.1 such that $F_{s,i} = T_i^{-1} F_i$. The gain for an input image is also constant for

a given CCD (bandpass) and camera column so $g_i = g$ for any set of images being co-added. The last equality from Eq. 2.13 can be incorporated into Eq. 2.11:

$$\sigma^2_{F_c} = g^{-1} \sum w_i^2 \; T_i^{-1} \; F_{s,i} \tag{2.14}$$

$$\approx g^{-1} \langle F_{s,i} \rangle \sum w_i^2 \; T_i^{-1} = g^{-1} F_c \sum w_i^2 \; T_i^{-1} \tag{2.15}$$

The intensities in the scaled input images, $F_{s,i}$, are calibrated to the same standard photometric zero-point, $\mathtt{aa}_c = -23.90$ and so are roughly equal. Their expectation value is the intensity in the unscaled co-add image, F_c. Inserting Eq. 2.15 into 2.10 and recognizing $F_{sc} = c_\circ F_c$ from Eq. 2.8:

$$\sigma^2_{F_{sc}} = c_\circ^2 \frac{F_c}{g} \sum w_i^2 \; T_i^{-1} = c_\circ \frac{F_{sc}}{g} \sum w_i^2 \; T_i^{-1} \tag{2.16}$$

Finally, the last equality of Eq. 2.16 can be inserted into $\sigma^2_{F_{sc}} = F_{sc}/G_{\rm eff}$ to obtain $G_{\rm eff}$:

$$G_{\rm eff} = \frac{g}{c_o \sum w_i^2 T_i} \approx \frac{g}{5.1092 \; \sum w_i^2 \; T_i^{-1}} \tag{2.17}$$

I computed the effective gain in each corrected co-add image directly using the input image lists from the co-add headers and the input image properties accessible from the CAS. A corrected image header contains *all* of the input image fields used to construct a co-add, but typically there are two images per run that each cover part of the co-add image field-of-view. The properties of neighbouring fields in a run are likely to be highly correlated. Therefore, I used the image properties from the field in each run that had the greatest overlap with the co-add field. The effective gain in the co-add images roughly scales as $N_{\rm images} \times g$, as expected for a mean, but are generally slightly smaller. The gains I have computed differ slightly (by at most 2%) from those provided in the atlas image headers. I interpret the small differences in my gain calculations from those in the headers as arising from: (1) the choice of constant $c_\circ$ as opposed to the exact value derived from the calibration; and (2) the input image property acquisition procedure outlined above.

2.2.5 Spatial resolution

The modified version of `PHOTO` that was used for the Stripe 82 co-adds models the spatial variation of the PSF using the suitably weighted sum of the model PSFs from

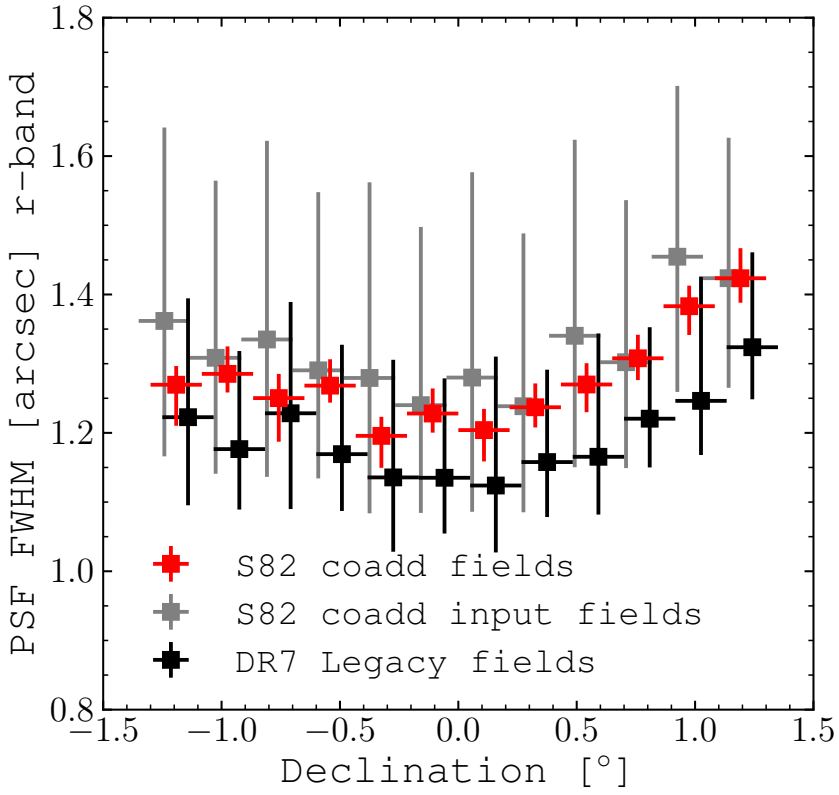

Figure 2.1 Characterization of spatial resolution in r-band co-add images. PSF FWHM are obtained from the respective co-add and single epoch FIELD tables and are measured from 2D gaussian fits to the PSF in each (co-add) image field. Points show the medians and $16^{\text{th}} - 84^{\text{th}}$ percentiles of PSF FWHM for all co-add stacks (red), single-epoch DR7 Legacy (black) fields, and co-add input fields in each of the 12 scanlines. Black and grey markers are offset by 0.05° right and left of their respective bin centres for clarity. Input co-add fields (grey) include also SDSS-II Supernovae Project runs which have poorer photometric standards standards than Legacy fields.

the input image fields – for which the spatial variations were already separately determined. The weighting scheme is designed such that images with the best seeing have higher weights in the co-add construction. Figure 2.1 compares the median spatial resolution along each scanline in the co-add images (red markers), the DR7 Legacy images used in the Simard et al. (2011) decompositions (black markers), and the full ensemble of input images used in the co-add stacks (grey markers). Measurements for the PSF were taken from the SDSS FIELD tables for Stripe 82 and DR7. The Simard et al. (2011) decompositions used Stripe 82 runs: (2583, 2662, 3325, 3388, 2738, 2659) from Legacy. The spatial resolution in the Stripe 82 co-adds is ~ 0.1" poorer than in the Legacy fields but is more homogenous along each scanline; reducing the typical scatter from ~ 0.3" to ~ 0.1". The median seeing and its scatter are improved in the co-adds with respect to the full set of input images – as expected from the image weighting strategy. The main difference between the Legacy and co-add input fields is the co-add inputs include images from the SDSS-II SNeS project runs which were generally taken under poorer photometric conditions. Fields from these runs generally have small weights in the co-add construction but are more numerous than Legacy fields; offsetting the median seeing to poorer quality.

2.2.6 Galaxy sample selection

I select galaxies from the SDSS DR7 spectroscopic galaxy catalog that are previously analyzed by Simard et al. (2011) and which reside in Stripe 82. This provides a sample of 16,908 galaxies for which decompositions in the deeper images can be directly compared with the more shallow reference decompositions from Simard et al. (2011). I avoid misclassified non-galaxy sources in my galaxy sample by selecting targets determined to be galaxies from SDSS fibre spectroscopy. Consequently, I have spectroscopic redshifts for all galaxies in the sample. Furthermore, galaxies from the Simard et al. (2011) decompositions all have Petrosian magnitudes $m_{\rm r,Petro} \leq 17.77$ mag – the faint limit of the SDSS Legacy spectroscopic sample. My selection, combined with $\sim 1.6 - 1.8$ magnitudes deeper imaging for this large galaxy sample, enables at least two unique experiments using bulge+disc decompositions. First, I can investigate the impact of added depth on global and component properties. Second, components and asymmetric structures that may previously have been too faint for accurate characterization, or indeed detection, in the single-epoch images may be better characterized in the Stripe 82 co-adds.

2.3 Decompositions and catalogs

2.3.1 Photometric Decompositions

To be consistent with the analyses of Simard et al. (2011) and Mendel et al. (2014) in shallower images, I use the same analysis pipeline for my decompositions in Stripe 82. Quantitative morphologies were obtained using the parametric surface-brightness decomposition tool GIM2D. An important feature of the GIM2D software is that it uses a Bayesian Metropolis-Hastings maximum-likelihood optimization algorithm (but see also PROFIT; Robotham et al. 2017). As such, the best-fitting model parameters are largely insensitive to the volume of parameter space in which optimization iterations begin. The nature of the Metropolis-Hastings algorithm is that there is always a finite probability of leaving a position of greater likelihood for a position of lesser likelihood. Therefore, the algorithm need not settle at a local maximum but can escape to (hopefully) converge on the global maximum. Each galaxy was fitted using three models:

- (`ps`) single-component Sérsic profile
- (`n4`) two-component $n_b = 4$ bulge and exponential disc
- (`fn`) two-component free-n_b bulge and exponential disc

The parameters of these models are described in Simard et al. (2011). As in Mendel et al. (2014), I perform *ur*, *gr*, *ir*, and *zr* pairwise simultaneous surface-brightness decompositions. The $r-$band mask from SEXTRACTOR is used for all images to avoid inconsistencies in pixel allocations in each band. Similarly, following Mendel et al. (2014), structural parameters in each band of a pairwise simultaneous fit are forced to be the same. These tied structural parameters are the bulge effective radius, bulge ellipticity, bulge Sérsic index, disc scale length, disc inclination, and bulge and disc position angles. Total galaxy magnitudes, bulge-to-total fractions, and centroid offsets are free to vary in each band of a pairwise fit.

In addition, following Mendel et al. (2014), the structural parameters for *ur*, *ir*, and *zr* pairwise fits were fixed to values first derived in *gr* decompositions (with the exception of Sérsic index in the `ps` decompositions – which is allowed to vary freely in each pairwise fit; see Appendix A of Mendel et al. 2014 for discussion). As such, the only truly independent structural measurements are those derived in the *gr* simultaneous decompositions – the structural measurements of all other pairwise

decompositions for a given model type (`ps`, `n4`, `fn`) are held fixed to the *gr* values. Consequently, this method does not offer the flexibility required to capture the known wavelength dependence of galaxy structural parameters that has been identified using more rigorous multi-wavelength models (e.g., SIGMA– Kelvin et al. 2012; and MEGAMORPH– Häußler et al. 2013; Vika et al. 2013). Such a model is not employed here because my primary goal is to characterize the impact of improved imaging depth on the Mendel et al. (2014) and Simard et al. (2011) decomposition measurements. Exploiting the same model and analysis pipeline is of fundamental importance to the nature of this task. So while the wavelength dependence of galaxy magnitudes and bulge-fractions in my models can be largely considered independent for each band, the structural parameters that are held fixed across bands cannot. Mendel et al. (2014) adopted this (albeit restrictive) multi-wavelength strategy to suitably construct a robust broadband SED for each galaxy which would not be strongly affected by the covariances between model parameters for sources that are faint in any particular band. In summary, any changes to *uiz* model structural parameters are *entirely* due to changes in the *gr* decomposition results (by construction). But total magnitudes, bulge and disc fractions, and corresponding component magnitudes are independent in each band. Any changes to these features are reflections of the improved constraints provided by the deeper images.

The fiducial `n4` model adopts an $n_b = 4$ Sérsic profile for the bulge. The choice of Sérsic index for bulge component models in galaxy surface-brightness decompositions has been a long-standing debate in the literature – ultimately resolving that a continuous range of bulge profile shapes must exist (Andredakis & Sanders 1994; Andredakis et al. 1995; de Jong 1996; Khosroshahi et al. 2000; Graham 2001; Möllenhoff & Heidt 2001; MacArthur et al. 2003; Graham & Guzmán 2003; Balcells et al. 2003; Kormendy & Kennicutt 2004; Fisher & Drory 2010; Laurikainen et al. 2010; see also Graham 2013 §2 and §4 for a review). At the heart of the remaining challenges to bulge characterization is whether Sérsic index can be used to classify and discriminate between bulges of different formation mechanisms. The challenge is exacerbated by the presence of additional structural components such as bars, excess nuclear light, and internal lenses (rings that do not derive from gravitational lensing) which are known to affect structural measurements – in particular the bulge-to-total fractions, B/T (Graham & Guzmán, 2003; Laurikainen et al., 2005; Peng et al., 2010; Laurikainen et al., 2013). Without independent priors for the presence of such features (e.g., by thorough visual inspection of each galaxy's surface brightness profile), adding

model components will only serve to create degeneracies between model parameters and produce meaningless results. Recently however, Kruk et al. (2018) have shown that visual classification from the Galaxy Zoo citizen science project (Lintott et al., 2008) can serve as useful priors for using additional components in the decompositions. Calibration of such priors and their implementation is beyond the scope of the present work.

The Stripe 82 co-adds present an opportunity to investigate whether added depth enables better characterization of the bulge profile. Simard et al. (2011) showed that the majority of galaxies in the single-epoch images had insufficient S/N and/or spatial resolution to discriminate between bulge profiles. Figure 2.1 shows that improved characterization of the peak due to better seeing is not expected. The seeing in a co-add image is not improved in quality (indeed, it is slightly poorer) compared to a corresponding single-epoch Legacy image. As such, changes to measurements due to seeing are generally not expected. As for depth, the 1.6-1.8 magnitude improvement (factor of 4.4-5.2 in S/N) should increase the number of pixels assigned to objects in the deblending – as more will meet the intensity threshold with respect to the sky noise. Larger footprints and more degrees of freedom make the fits slower, but enable better characterization of the low surface-brightness limits of galaxy profiles. With respect to the bulge, the increase in S/N should improve delineation of the sky from the wings of the bulge profile – which is also important for accurate bulge characterization.

2.3.2 Tables & catalogs

Catalogs for the photometric decompositions are made available in plain text and Structured Query Language (SQL) format[6]. The catalogs have the same naming convention and structures as the Simard et al. (2011) public catalog. The salient features of each table are shown in Table 2.1. The basic structure and parameters of each catalog are shown in Table B.1 in Appendix B. Uncertainties on model parameters are the 16% and 84% confidence limits acquired by sampling the convergence region about the best-fitting model after convergence. The statistical uncertainties on observed fluxes from GIM2D are lower limits, which do not account for additional uncertainties arising from relative and absolute SDSS photometric calibration and sky subtraction.

[6]The catalogs are available here: `http://orca.phys.uvic.ca/~cbottrel/share/Stripe82/Catalogs/`

Mendel et al. (2014) used Monte Carlo simulations of analytic bulge+disc models in SDSS fields to confirm that the GIM2D statistical uncertainties and the rms noise in the sky background are consistent with the total uncertainties in the recovery of simulated galaxy fluxes. Additionally, an uncertainty of $\sim 3\%$ in *griz* and $\sim 5\%$ in u bandpasses arises from the relative and absolute photometric calibrations. Taken together, the combination of these three sources of uncertainty account for the total uncertainties in the model fluxes.

2.4 Comparison with Legacy results

The galaxy sample selection and consistency in methodology enable direct comparison of decomposition results from the co-adds and shallower Legacy images. Changes to *integrated* galaxy properties (accounting total flux contributions from all model components – whether single-component or multi-component) such as total galaxy magnitudes and galaxy half-light radii are not expected to change drastically among galaxies with surface brightnesses that are well within the Legacy detection limits of each bandpass. For example, Figure 2.2 shows a mosaic of `n4` decomposition results for a single randomly selected galaxy in the range $15 < m_r < 16$ covering each band for both the Stripe 82 co-adds and corresponding Legacy images. Despite a visibly significant improvement in image quality in the co-adds, total magnitude and half-light radius measurements in each band are largely unchanged – even in the $u-$ and $z-$ bands where the starkest contrast in image quality is seen.[7] In the next section, I will show that this consistency holds for most galaxies in my comparison. However, galaxies at the thresholds of the Legacy detection limits stand to see improved constraints on their integrated properties. As for the components, it is of particular interest whether the added depth can better discriminate the light emerging from the bulge and disc. In this section, I compare the `n4` decompositions for galaxies in Stripe 82 co-adds and Legacy images.

[7]Although, note that if the *gr* structural parameters are not changed, then the reason that *uiz* half-light radii do not change *can* be partly by construction – the bulge and disc structures in these bands are tied to the *gr* bulge and disc structures. However, the magnitude of each component is fully independent in each band. Therefore, one can still have band-specific changes in half-light radius even where the structures of bulge and disc components are fixed.

Table 2.1 Photometric decomposition catalogs. Fits using each of the three models are performed simultaneously in two bandpasses – one of which is always the $r-$band. Structural parameters (bulge effective radius, bulge ellipticity, disc scale length, disc inclination, and bulge and disc position angles) are either free or fixed to the gr results as indicated. The model's (bulge) Sérsic index is also either free or fixed to the gr results as indicated (though it is always free in the single Sérsic component model decomposition). The number of successful fits to each bandpass pair, $N_{\rm success}$, is shown in the final column.

Catalog Name	Decomposition Model	Bandpasses	Structural Parameters	(Bulge) Sérsic index	$N_{\rm success}$
`sdss_s82_morph_gr_n4`	$n_b = 4$ bulge + exp. disc	g and r	Free	Fixed $n_b = 4$	16,822
`sdss_s82_morph_ur_n4`	$n_b = 4$ bulge + exp. disc	u and r	Fixed to gr	Fixed $n_b = 4$	16,672
`sdss_s82_morph_ir_n4`	$n_b = 4$ bulge + exp. disc	i and r	Fixed to gr	Fixed $n_b = 4$	16,663
`sdss_s82_morph_zr_n4`	$n_b = 4$ bulge + exp. disc	z and r	Fixed to gr	Fixed $n_b = 4$	16,704
`sdss_s82_morph_gr_ps`	single free-n Sérsic	g and r	Free	Free	16,892
`sdss_s82_morph_ur_ps`	single free-n Sérsic	u and r	Fixed to gr	Free	16,764
`sdss_s82_morph_ir_ps`	single free-n Sérsic	i and r	Fixed to gr	Free	16,724
`sdss_s82_morph_zr_ps`	single free-n Sérsic	z and r	Fixed to gr	Free	16,729
`sdss_s82_morph_gr_fn`	free-n_b bulge + exp. disc	g and r	Free	Free	16,822
`sdss_s82_morph_ur_fn`	free-n_b bulge + exp. disc	u and r	Fixed to gr	Fixed to gr	16,707
`sdss_s82_morph_ir_fn`	free-n_b bulge + exp. disc	i and r	Fixed to gr	Fixed to gr	16,681
`sdss_s82_morph_zr_fn`	free-n_b bulge + exp. disc	z and r	Fixed to gr	Fixed to gr	16,664

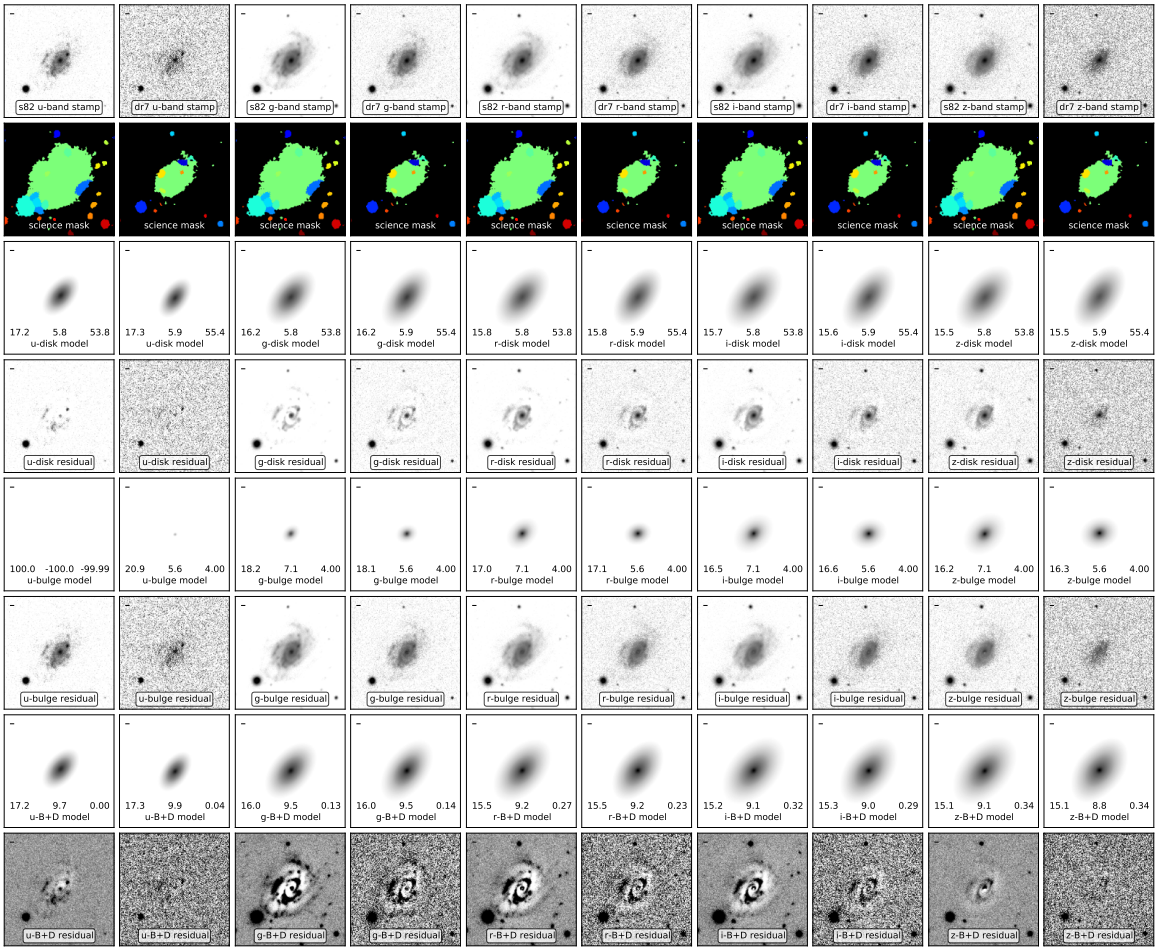

Figure 2.2 Bulge+disc decompositions in SDSS *ugriz* bands for a single randomly selected galaxy in the sample. Each pair of columns (reading the figure in landscape) shows the decomposition results for a corresponding band in Stripe 82 (left columns) and DR7 Legacy (right columns). The galaxy models in the right columns are constructed from the Simard et al. (2011) results. Each image, with the exception of the mask, includes a ruler on the upper left measuring 3 arcsec. *First row*: science image cut-outs; *Second row*: mask image generated with SExtractor from the r-band image frame (pixels with the same colour as the central pixel belong to the source, black=sky, and all other colours are deblended sources); *Third row*: `n4` disc model with disc apparent magnitude, disc scale length (arcsec), and disc position angle; *Fourth row*: exponential disc model residual; *Fifth row*: `n4` bulge model with apparent bulge magnitude, bulge effective radius (arcsec), and bulge Sérsic index (always 4.0 in this case); *Sixth row*: bulge model residual; *Seventh row*: full `n4` model with total apparent magnitude, half-light radius (arcsec, derived from the combination of the bulge and disc model), and bulge-to-total fraction; *Eighth row*: `n4` model residual. The bulge+disc model residual images use a linear greyscale in which the contrast limits are set to $(-5, 5)\sigma_{\rm sky,S82}$. All other images use a logarithmic greyscale in which the contrast limits are set to $(0.01, I_{10})$ nanomaggies where I_{10} is the maximum intensity within 10 pixels of the target galaxy barycenter.

2.4.1 Total magnitudes

The total flux from all model components should always be conserved in any decomposition that uses the same degrees of freedom and model parametrization. In other words, similar models fitting the same set of pixels should recover the same total flux – even when other best-fitting model parameters may differ (e.g., degeneracy). The number of degrees of freedom in a 2D surface brightness decomposition is equal to the number of pixels used in the fit, $N_{\rm pixfit}$, subtracted by the number of free parameters in the model. This quantity is not conserved in my comparisons, as the increased S/N in the co-adds extends the source masks when deblending objects from sky – increasing $N_{\rm pixfit}$ often substantially (for example, see the second row of panels in Figure 2.2). However, since the total magnitudes are computed by integrating the model fluxes to infinity, increases to the sizes of the object footprints should not change their measured fluxes unless the signal at extended radii necessitate changes to the structural parameters. It is one of the goals of the current work to examine whether the added depth imparts such changes.

Figure 2.3 shows that the total galaxy light is largely conserved. The median offsets in total apparent magnitude for all galaxies in each band are $\Delta m^{50\%}_{u,g,r,i,z} = (+0.037, +0.022, +0.021, +0.020, +0.019)$ mag (DR7-S82). The most notable exception is the $u-$band which has the poorest response and in which many galaxies are also intrinsically fainter. Mild median systematic trends in $\Delta m_{x,\rm DR7-S82}$ are clear in each bandpass. The brightest galaxies get boosts to their total fluxes in the co-add measurements with respect to Legacy. At the faint end, a systematic trend is notable in the $u-$band – where the median systematic magnitude offsets first increase slightly (brighter in the co-adds) at $m_u \lesssim 19.5$ mag then drop rapidly at $m_u \lesssim 21.2$ mag (becoming fainter in the co-adds). To understand this systematic, Figure 2.3 also shows percentiles in the DR7 (solid) and Stripe 82 (dotted) mean local sky uncertainties, $\sigma_{\langle \rm sky \rangle}$, expressed as magnitudes (rather than relative magnitude *offsets* as used in, for example, Appendix B). These are computed:

$$\sigma_{\langle \rm sky \rangle}(\rm mag) = -2.5 \log_{10}\Big(\sigma_{\rm sky}(\rm maggies/arcsec^2) \times \sqrt{N}_{\rm pixfit} \times \texttt{pixel\_scale}^2\Big) \tag{2.18}$$

where $\sigma_{\rm sky}$ is the local *sample* sky surface brightness uncertainty that is measured (see Appendix B, Eq. B.1), $N_{\rm pixfit}$ is the number of pixels used in the fit, and `pixel_scale`

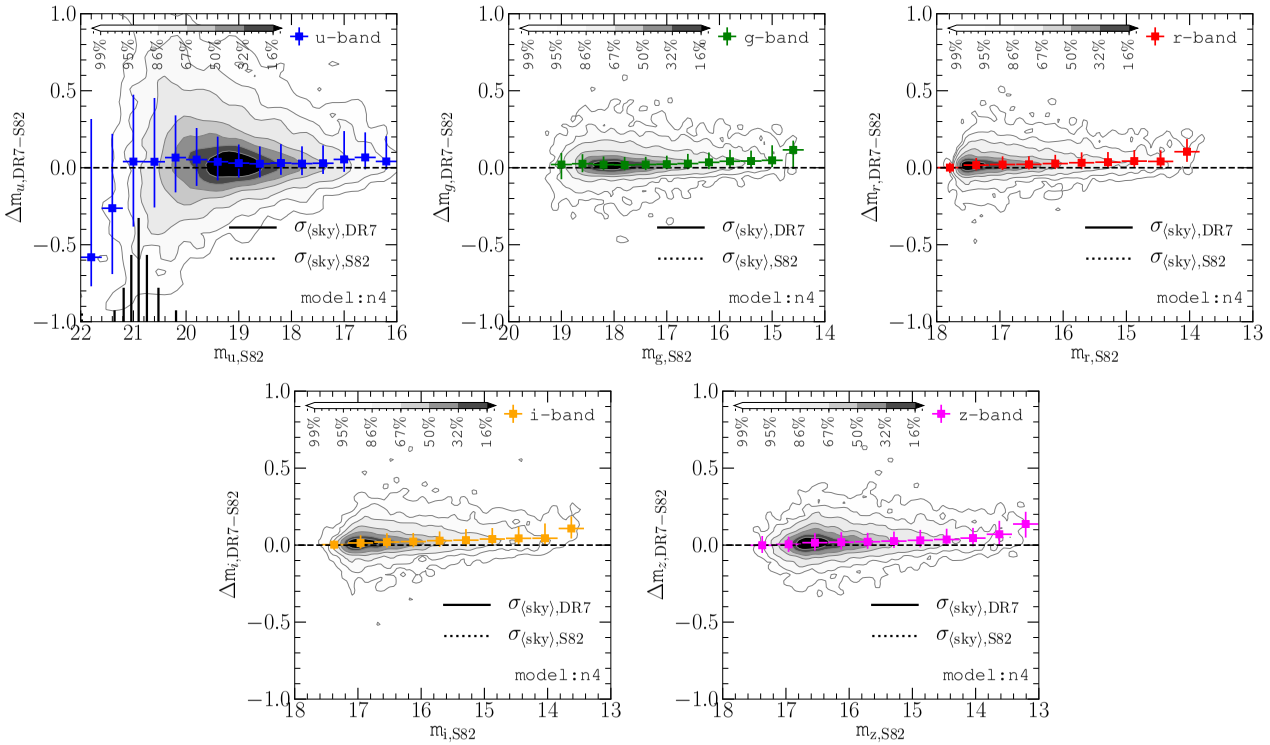

Figure 2.3 Difference in total apparent magnitudes of galaxies obtained from Stripe 82 co-adds and DR7 Legacy images with the `n4` decomposition model. The $r-$band magnitudes are taken from the gr decomposition catalogs. Contour lines show percentiles of the magnitude difference distributions, $\Delta m_{x,\mathrm{DR7-S82}}$, in each panel. Dashed lines denote $\Delta m_{x,\mathrm{DR7-S82}} = 0$. Coloured markers show the median in each bin with error bars corresponding to the $16-84$ percentile range in bins of 0.4 mag along $m_{x,\mathrm{S82}}$. Lines extending from the magnitude axis denote (from left to right) the {95%, 84%, 68%, 50%, 32%, 16%, 5%} percentiles of the local *mean* sky measurement uncertainties, $\sigma_{\langle \mathrm{sky}\rangle}$ (mag) in the Stripe 82 fits (dotted, not visible in any panel) and Legacy fits (solid, only visible in the $u-$band). As a pedagogical example of how to interpret these lines, 5% of DR7 u-band mean sky surface brightness uncertainties correspond to magnitudes brighter than $m_u \approx 20.2$ mag. Note that the Legacy 50% $\sigma_{\langle \mathrm{sky}\rangle,u}$ peak coincides with the sudden decrease in Stripe 82 $u-$band brightnesses relative to Legacy. Legacy brightness measurements have difficulty penetrating the sky noise at these limits.

is the CCD plate scale of 0.396 arcsec/pixel. While these uncertainties are too small to be seen in the panels for other bandpasses, the sudden systematic drop in median Stripe 82 u-band brightness at $m_u \approx 21$ mag coincides with the median in $\sigma_{\langle \mathrm{sky} \rangle,\mathrm{DR7}}$ in that band. So, while the Legacy flux measurements for the faintest targets in the $u-$band can be limited in their capacity to penetrate the sky noise, Stripe 82 measurements go deeper and are able to properly characterize these faint sources. In other words, given that the Legacy $u-$band mean sky uncertainty measurements lay closer to the distribution of galaxy u-band brightnesses, the u-band magnitudes stand to be most improved by the added depth of the Stripe 82 co-adds.[8] Visual inspection of the science images and decomposition results for faint $u-$band galaxies with $m_{u,\mathrm{DR7}} > 20.5$ mag in the co-adds and Legacy revealed that the vast majority of these galaxies are indeed barely (if at all) discernible from the sky background in the Legacy $u-$band images. Figure 2.4 highlights a few such cases in a mosaic of decompositions for five individual galaxies in the $u-$band. Cases in which a galaxy is barely detectable can generate both negative *and* positive scatter in $\Delta m_{x,\mathrm{DR7-S82}}$, but have a stronger tendency towards negative offsets as the model likely attempts to fit features of the noise.

Figure 2.5 shows two examples that are representative of many faint galaxies whose fluxes were increased in the co-add measurements. Despite the differences to the model parameters, the residuals are identical in quality. This result suggests that the total brightnesses of these faint systems are not incorrectly over-estimated in Stripe 82. Rather, the enhanced fluxes in the models reflect true changes to the faint end of the galaxy surface-brightness distributions with respect to what is accessible in Legacy. It is possible that changes to the model parameters are partially due to degeneracies between structural parameters at these limits of resolution and S/N. But if degeneracy was the dominant factor, the total magnitudes would be unchanged as long as the residuals are consistent. A statistical approach to characterize the role of degeneracy in the decompositions follows in Section 2.4.5. For now, I assert that the total fluxes in the co-adds exhibit reduced systematics on model parameters due to surface-brightness limits relative to the Legacy images for my galaxy sample. I further investigate this assertion by directly examining the total photometric uncertainties for measurements of total and component magnitudes and bulge-to-total fractions in

[8] One should recall, however, that *structural* parameters in ur, ir, and zr in the `n4` fits are fixed to the gr results in the experimental design. So these parameters are not affected by any improvements in the photometry of uiz bandpasses.

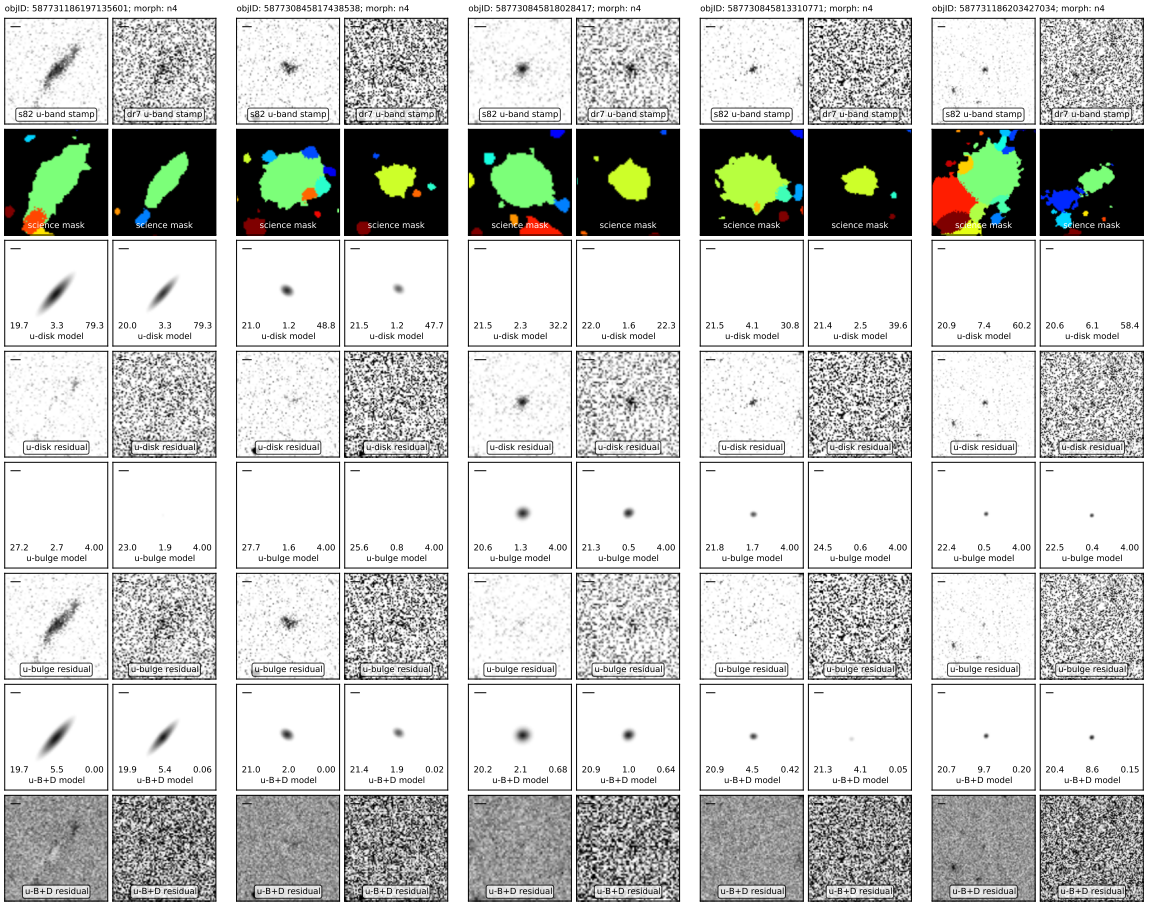

Figure 2.4 Similar to Figure 2.2 but now each pair of columns pertains to an individual galaxy's $u-$band Stripe 82 and Legacy decompositions. The first pair of columns shows a galaxy that is generally representative of the positive $\Delta m_{u,\mathrm{DR7-S82}}$ scatter at faint $u-$band magnitudes in Figure 2.3 owing to substantially improved visibility in the Stripe 82 co-adds. Despite this improvement, the change in total $u-$band apparent magnitude for this galaxy is relatively small $\Delta m_{u,\mathrm{S82-DR7}} \approx 0.1$ mag. The second, third, and fourth pairs of columns show galaxies with $m_{u,\mathrm{DR7}} > 20.5$ mag and larger positive offsets in $\Delta m_{u,\mathrm{DR7-S82}}$. In the second and fourth pairs of columns, the galaxies are virtually invisible in the Legacy images. In the third case, the galaxy is visible but is still noise-dominated – resulting in similar under-estimated total apparent magnitude. The fifth pair of columns shows an example characteristic of the downward systematic in the $u-$band panel of Figure 2.3. It shows a galaxy for which the Stripe 82 total apparent magnitude is significantly fainter than in Legacy and practically invisible in the shallower image.

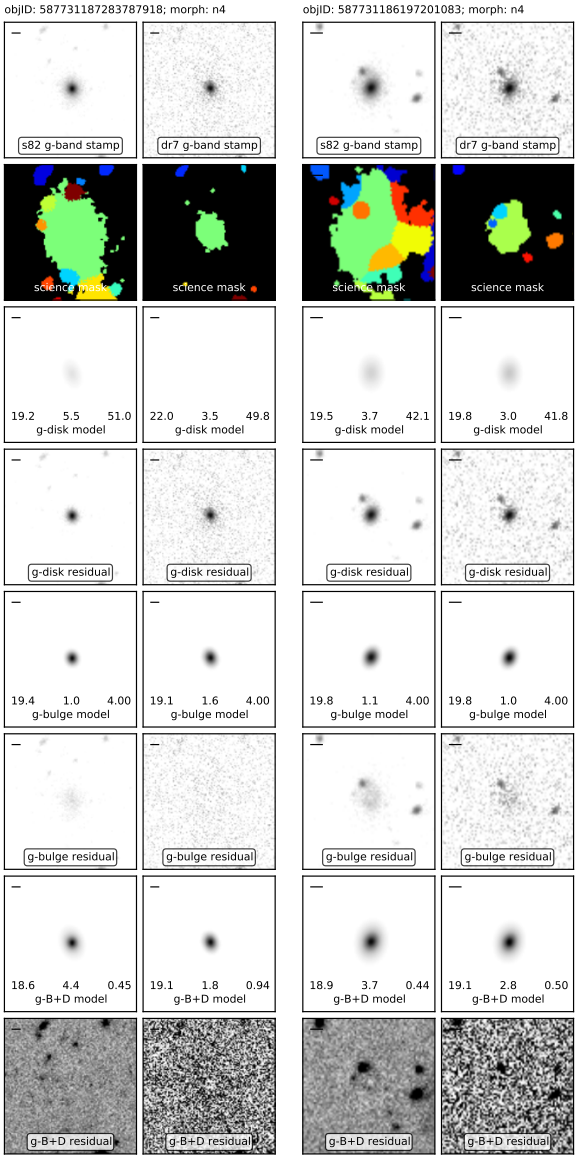

Figure 2.5 Decomposition mosaics for two galaxies whose faint $g-$band total magnitudes were brightened by the co-adds. Improved characterization of the faint structure surrounding the central galaxy is afforded by the co-adds in each case. The galaxy in the left two columns switches from bulge-dominated in Legacy to a composite bulge+disc system in the co-adds. An extended disc is assigned to characterize the faint structure in the co-add images. This new extended structure accounts for $\Delta m_{g,\mathrm{S82-DR7}} = -0.5$ mag difference between the best-fitting bulge+disc models. The panels on the right also shows a galaxy for which the disc becomes brighter and larger in order to characterize the fainter structure. The bulge model remains largely consistent. Additionally, a possible tidal stream is revealed between the central galaxy and the interloper to its upper left. It is most visible in the disc-subtracted co-add image.

Appendix A.

The Stripe 82 magnitudes of the brightest galaxies in each band can be up to 0.1 mag ($\sim 10\%$) brighter than the DR7 magnitudes. This systematic difference in total flux is likely due to changes to the measured light profiles at extended radii – afforded by the increased S/N and consequent galaxy footprint size in the mask image. Given that GIM2D galaxy magnitudes are computed by integrating the best-fit models out to infinity, an improved characterization of low-surface brightness wings in the light profiles of bright galaxies may easily add up to a 10% total flux difference for bright galaxies. Another likely culprit is that the local sky level around each galaxy is more accurately estimated in the Stripe 82 images than in Legacy. The increased footprint sizes of sources in the mask images (by an average of 246% relative to Legacy) mean that the local sky statistics are computed from an ensemble of pixels that are farther from *all* sources in an image. Given the environments that the brightest and largest galaxies tend to live in and their tendency to have highly extended stellar halos (e.g., Tal & van Dokkum 2011), it is important to find balance between a obtaining a truly local estimate of the sky and ensuring that the sky estimate is not overly contaminated by the extended flux from bright targets or their neighbours.[9] From Figure 2.3, I highlight while this contamination is undoubtedly the source of suppressed fluxes for the brightest galaxies in Simard et al. (2011)'s Legacy decompositions relative to Stripe 82, its effect is limited to ~ 0.02 mag with the exception of the brightest bins in each band.

2.4.2 Galaxy colours

The scatter in $\Delta m_{x,\mathrm{DR7-S82}}$ is related to the photometric depth in each bandpass and to the intrinsic brightnesses of galaxies in each bandpass. The scatter is greatest in the $u-$band, where the Legacy photometry is the poorest. Brighter than $m_{u,\mathrm{S82}} \approx 20$ mag, this scatter is symmetric and the median systematic offset is $\Delta m_{u,\mathrm{DR7-S82}} \lesssim 0.02$ mag. Where it is symmetric, the scatter in $\Delta m_{x,\mathrm{S82-DR7}}$ should be the sum in quadrature of the total random and systematic measurement uncertainties from the respective co-add and Legacy decompositions with respect to the galaxies' intrinsic magnitudes

[9]To justify this claim, I have separately confirmed that $\Delta m_{x,\mathrm{DR7-S82}}$ and the Legacy residual sky offset between the local GIM2D sky and the full-frame SExtractor sky, *db*, are positively correlated for galaxies with $m_r < 14.5$ mag. However, the systematic suppression of total flux is limited primarily to targets in this magnitude range and that *also* have high bulge-to-total fractions (high Sérsic indices) as predicted by Mendel et al. (2014) using artificial galaxy simulations.

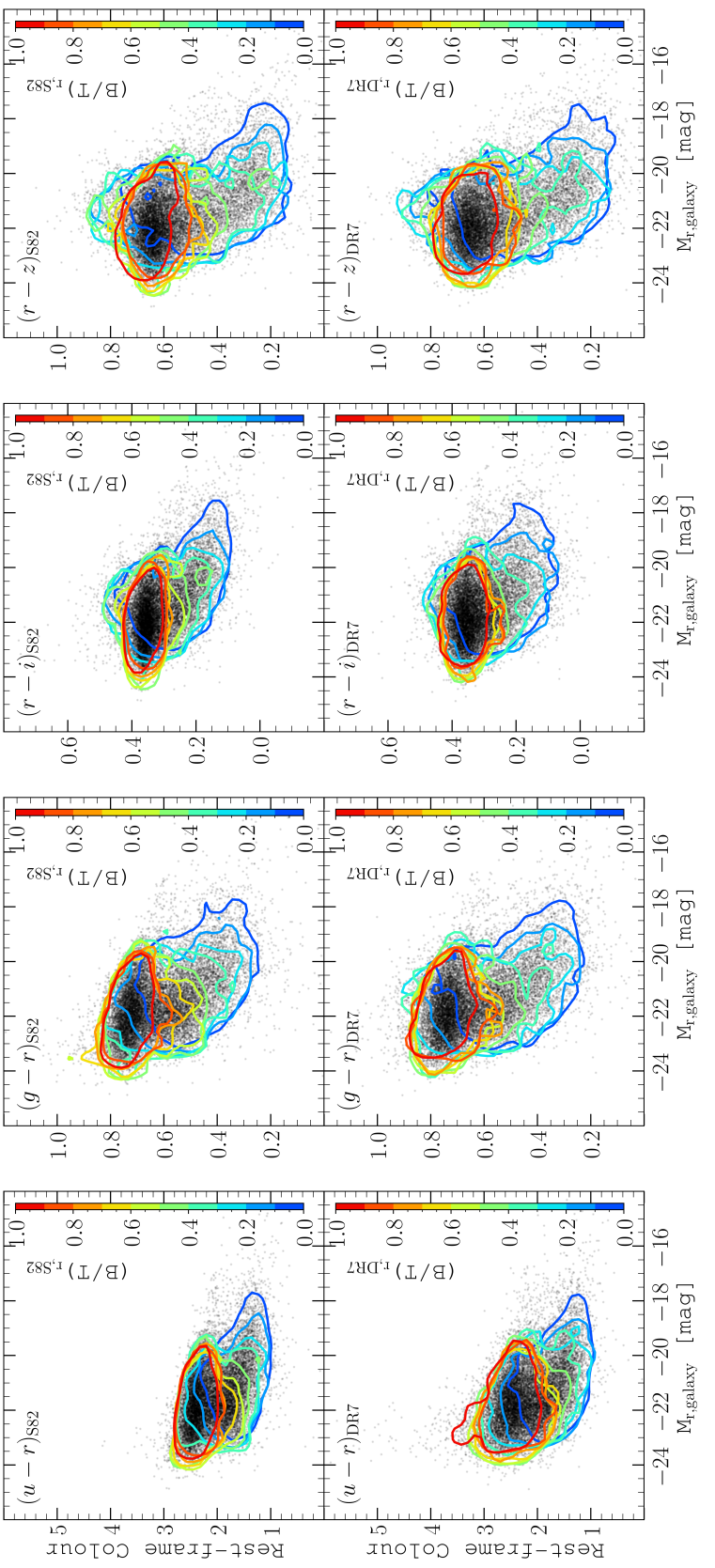

Figure 2.6 Rest-frame Colour-Magnitude Diagrams from **n4** decompositions of galaxies in the co-adds (upper row of panels) and DR7 Legacy images (lower row of panels). Colours from each simultaneous xr decomposition are only correlated insofar as forcing the structural parameters to be the same as those derived from the gr fits. Brightnesses of each component in each bandpass are free to vary independently. Black markers show the full distribution of galaxy colours each panel. Coloured lines show the 86% density contours of these distributions in 10 discrete bins of $(B/T)_r$ following the colourmaps displayed in each panel. For example, the dark red line shows the 86% density contour for galaxies in the last B/T bin the range $0.9 \leq (B/T)_r \leq 1$.

– which are not known.

To illustrate which of the Stripe 82 and Legacy decompositions contribute most to the scatter shown in Figure 2.3, I compare their respective rest-frame colour-magnitude diagrams (CMDs) in Figure 2.6. The tightness of the red-sequence is a common metric for the constraints on global galaxy colours and the precision of total flux measurements (e.g., Simard et al. 2011). Figure 2.6 shows the distributions of Stripe 82 and Legacy galaxies for four colours. Coloured lines show contours of these distributions in bins of $r-$band B/T. Most bulge-dominated galaxies live on the red sequence. But a large number of red discs and red two-component systems are also found there. The tightness in the red sequence is most improved in the $u-r$ colours in the co-adds compared to the Legacy colours. Note the large number of outliers with high $u-r$ in Legacy that are no longer outliers using the co-add measurements. In the $g-r$, $r-i$, and $r-z$ colours, the visible tightness of the red sequence is mildly improved in the co-add decompositions. In general, these results are consistent with a scenario in which the scatter in the magnitude offsets, $\Delta m_{x,\mathrm{DR7-S82}}$, is dominated by larger measurement uncertainties in the Legacy decompositions. Both sets of decompositions, however, identify the well-established trend that galaxies are typically redder as they become more bulge-dominated (Strateva et al., 2001; Kauffmann et al., 2003b; Blanton et al., 2003a; Baldry et al., 2004; Balogh et al., 2004; Driver et al., 2006; Faber et al., 2007) – consistent with quenching scenarios which revolve around growth of a compact stellar component (Kauffmann et al., 2003a, 2006; Schiminovich et al., 2007; Bell, 2008; Cheung et al., 2012; Fang et al., 2013; Lang et al., 2014; Omand et al., 2014; Bluck et al., 2014; Woo et al., 2015; Bluck et al., 2016; Teimoorinia et al., 2016).

The tightness of the red sequence is also a general indicator of the effectiveness of deblending algorithms in masking the light emanating from nearby sources. If the deblending is poor and the colours of nearby objects differ from those of the target galaxies, then the scatter in the red sequence will be inflated. The visibly improved tightness of the red sequence in the co-add colours indicates that the increased S/N in the co-adds may improve deblending of interloping light for crowded targets. This result may be particularly useful for analysis of galaxy pair colours and the masses that can be inferred from these colours.

2.4.3 Galaxy sizes

As with the total magnitudes, total galaxy sizes (e.g., Petrosian and galaxy half-light radii) should be mostly conserved barring galaxies at the detection limits. However, even for galaxies with good photometry in Legacy, half-light radii in the co-adds could benefit from improved constraints on the extended light from the bulge or disc. In the previous section, I asserted that the systematic increase in galaxy fluxes measured in the co-adds was due to the increased S/N at faint surface-brightnesses. We should then expect to see a corresponding systematic increase in galaxy half-light radii. Figure 2.7 shows this expected increase in galaxy sizes using $r-$band galaxy half-light radii, $r_{\rm hl,galaxy,r}$, from the gr `n4` decompositions. Half-light radii are computed as the semi-major axis radius of the ellipse in which half of the total model flux is contained (circular aperture half-light radii are also available in my catalogs). Similar to Figure 2.3, I plot the offset of the co-add sizes from the Legacy sizes. The median systematic peaks at the bright end with $\Delta r_{\rm hl,galaxy,r} \approx -0.03$ dex or $\sim 7\%$ and shallows at fainter magnitudes. Figure 2.7 is focused on the r-band, but the same trend was found in each other band (though, as for the magnitude differences, with larger scatter). On some level the consistency of this trend across each band is by construction in the `n4` fits – as the structural parameters of the bulge and disc components are fixed to the gr results. However, because the magnitudes of the components (and consequently B/T) are free to vary independently in each band, half-light radii (which measure half-light radius of the full model) can still change in each band due to varying bulge and disc fractions.

The scatter in $\Delta r_{\rm hl,galaxy,r}$ will be partially driven by differences in the segmentation maps. As discussed in previous sections, the sizes of the SExtractor footprints for galaxies in the co-adds are larger on average than in the Legacy images. There is an important consequence of having larger maps and higher S/N in the deep stacks on galaxy size. The increased S/N enables detection and deblending of sources previously unidentified in the Legacy segmentation maps. By deblending faint sources from the galaxy flux, their typically positive systematic on galaxy sizes can be reduced.

2.4.4 Bulge-to-total light fractions

So far I have focused on characterizing the integrated properties of galaxies. I now shift attention to the the bulge and disc components starting with bulge-to-total fraction, B/T – a useful quantitative indicator of galaxy morphology. Figure 2.8

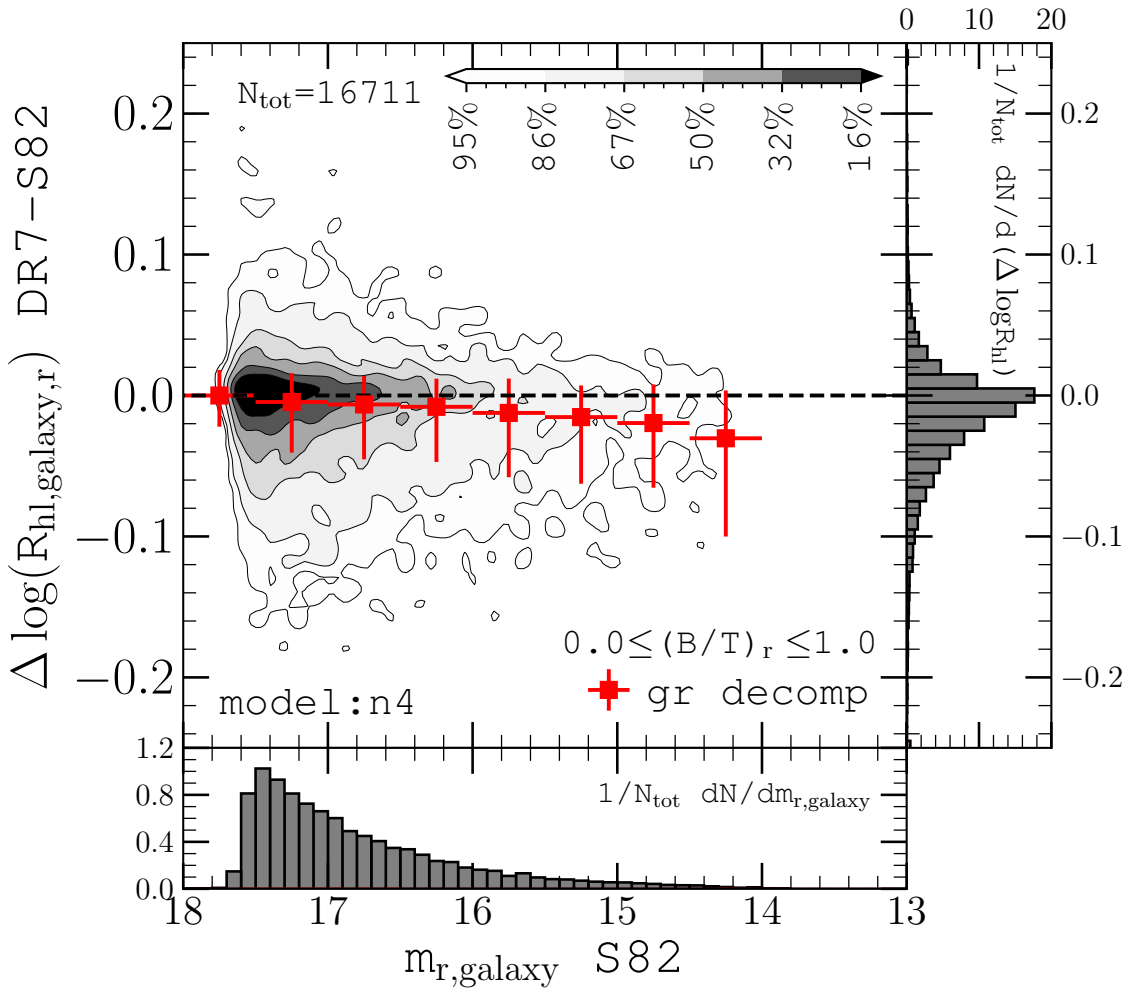

Figure 2.7 Galaxy half-light radii measured the $r-$band from the `n4` decompositions. Median systematic offsets for all galaxies taken together are typically small, but can be larger for particular B/T ranges. The largest median systematics were found to be -0.05 dex at the bright end of 0.4 <B/T< 0.6.

shows the changes in B/T in the co-add and Legacy images. The lower panel shows the B/T distribution for each set of decompositions taken from the respective `n4` *gr* tables. Interestingly, the peak at $B/T = 0.5$ in the Legacy images is largely removed in the co-add results. The upper panel shows the difference between the co-add and Legacy bulge-to-total fractions in each band, $\Delta(B/T)$, plotted against the Legacy B/T estimates in the r-band. Here we see that the B/T values which previously resided at the peak are enhanced in the deeper images and that they fill up the deficit between $0.6 \lesssim B/T \lesssim 0.95$. These results support the idea that the peak at $B/T = 0.5$ in the Legacy decompositions is an artifact that arises wherever there are very poor constraints on a galaxy's structure in the Legacy photometry. Such cases will naturally tend to $B/T \sim 0.5$ because it is the median value of a flat B/T posterior probability distribution (i.e. unconstrained B/T).

Note also the large number of galaxies with $(B/T)_{\rm DR7} = 1$ in the upper panel of Figure 2.8 which now have negative offsets as large as $\Delta(B/T) = -0.4$. While it should be noted that values of $\Delta(B/T) > 0$ are by definition impossible for $(B/T)_{\rm DR7} = 1$, the same is true for $\Delta(B/T) < 0$ where $(B/T)_{\rm DR7} = 0$ and no such strong clustering of positive $\Delta(B/T)$ is seen there. Are these discs being revealed by deeper photometry of predominantly bulge-dominated galaxies? To answer this question, I examine the underlying hypothesis that the deeper images enable better discrimination between single-component and two-component systems.

2.4.5 F-test statistics

Until now we have not considered the decomposition results in the other tables. The integrated quantities (total magnitudes and galaxy half-light radii) for the `ps` and `fn` model decompositions broadly mirror the results from the `n4` decompositions highlighted in the above subsections and compare similarly with the matching values in Legacy. The quality of the decomposition of a galaxy is characterized quantitatively by the reduced χ^2_ν statistic of the model with respect to the data:

$$\chi^2_\nu = \frac{1}{N_{\rm dof}} \sum_{i=1}^{N} \frac{(f_{i,\rm data} - f_{i,\rm model})^2}{\sigma_i^2} \tag{2.19}$$

where $N_{\rm dof}$ is the number of degrees of freedom, N is the number of data points used in the fit, $f_{i,\rm data}$ and $f_{i,\rm model}$ are the data and model values of element i of the data, respectively, and σ_i^2 is the variance in that element. χ^2_ν can vary significantly between

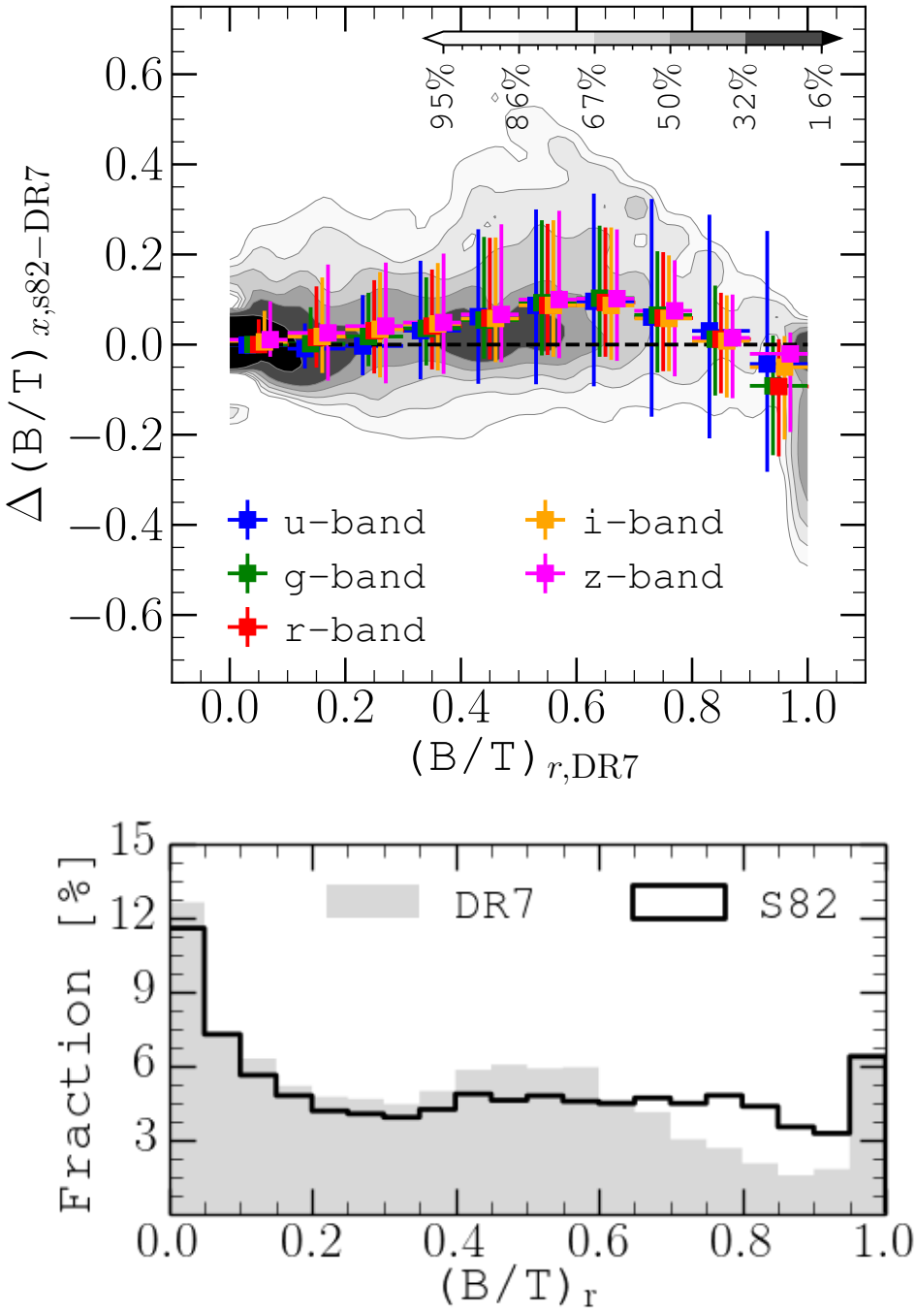

Figure 2.8 Galaxy bulge-to-total light ratios. The upper panel shows the B/T offsets, $\Delta(B/T)$, as a function of the B/T measured using Legacy photometry. The background contour map shows the two-dimensional distribution in the $r-$band. Coloured markers show the median and 16-84 percentile ranges in $\Delta(B/T)$ measured in each bandpass. The lower panel shows the one-dimensional B/T distributions from the Legacy (gray, shaded) and Stripe 82 (black, empty) decompositions in the $r-$band.

models. However, a direct comparison between the χ^2_ν for each model in the hierarchy (`ps`, `n4`, `fn`) will always yield an equal or better χ^2_ν for the model with the greatest number of degrees of freedom[10] (up to Monte Carlo error).

In order to discern whether the data demand a more complex model over a simpler model, I compare the results from different models using the F-statistic (Simard et al., 2011; Meert et al., 2013; Mendel et al., 2014). For each decomposition model, I compute χ^2_ν, where $N_{\rm dof}$ is taken as the number of *resolution elements*, $N_{\rm res} = N_{\rm pixfit}/(\pi\ {\rm HWHM}^2_{\rm psf})$, subtracted by the number of free parameters in the model[11]. The HWHM is the half-width at half-maximum of the PSF. The F value is then the ratio of the χ^2_ν for each pair of models. These F values are converted to corresponding probabilities via the F-distribution. As in Simard et al. (2011) and Mendel et al. (2014), the F values in my calculations convert to probabilities that the more complex model is *not* required to properly model the galaxy structure. In this way, a more complex model is only favoured under the test if the additional degrees of freedom in the model offer *substantial* improvement to the χ^2_ν of a simpler model. The `n4` table therefore includes the P_{pS} probability that an `n4` decomposition is *not* required to properly characterize the surface brightness distribution of a galaxy relative to the `ps` decomposition. Similarly, the `fn` table includes both P_{pS} and P_{n4} probabilities that the `fn` decomposition is *not* required relative to the `ps` and `n4` models, respectively.

Figure 2.9 shows the F-test results for my Stripe 82 galaxy sample in a comparison with Legacy values from Simard et al. (2011). Upper panels show the P_{pS} probabilities computed from the χ^2_ν of the best-fitting `n4` and `ps` from the *gr* decompositions. An ideal indicator of whether a galaxy is a single- or two-component system would yield a binary classification. In practice, this idealized binary classification scenario reduces to assignment of probabilities due to the presence of additional structures in galaxies (bars, stellar nuclei, etc.) and the limitations of the photometry to reveal faint underlying bulge or disc components. The upper left panel of Figure 2.9 shows that P_{pS} from Legacy are centrally peaked and largely concentrated around $P_{pS} = 0.5$. This result demonstrates that Legacy photometry offers little discriminating power between the bulge+disc and single-component models. The ambiguity in Legacy is

[10]Note that this is not a general statement and is specific to cases where $N_{\rm dof}$ is closely approximated by the number of data points, N, used in the fit (i.e. the number of free parameters in the model is very small compared to $N_{\rm dof}$).

[11]The number of degrees of freedom, $N_{\rm dof}$, takes a different definition here than it does in the decompositions themselves. In the decompositions, $N_{\rm dof}$ is the number of target pixels (those flagged as the target galaxy) and sky pixels in the science image cut-out (together, $N_{\rm pixfit}$) minus the number of free parameters in the model.

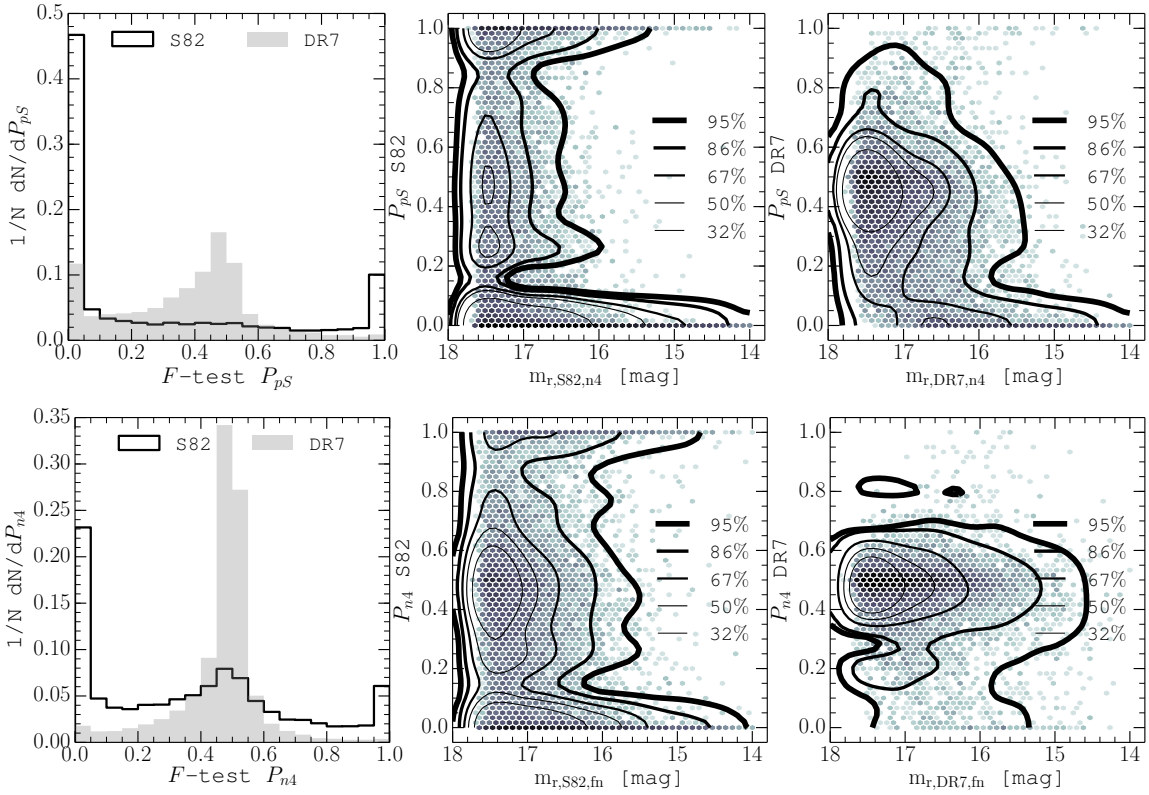

Figure 2.9 F-test statistics comparing between models of increasing number of degrees of freedom. *Upper left*: one-dimensional histograms of P_{pS} for Stripe 82 (solid, empty) and DR7 Legacy (gray, shaded). *Upper center and right*: histograms of P_{pS} in Stripe 82 (left) and DR7 Legacy (right) and their respective galaxy total magnitudes measured in the `n4` decompositions. Percentile legends for the black contour lines are inset on the middle right of each panel. *Lower row*: same as upper row of panels but comparing the P_{n4} statistics derived Stripe 82 and DR7 Legacy photometric decompositions.

largely eliminated in the co-add decomposition results – which offer substantially improved classification of galaxies as a single- or two-component systems. Panels to the right show the sensitivity of P_{pS} discriminating power to total apparent magnitude in the co-adds (middle) and Legacy images (right). Both maps become denser around $P_{pS} = 0.5$ for fainter galaxies. But the improved S/N in the co-adds greatly extends the magnitude range over which the F-test probabilities yield a binary-like classification.

Lower panels of Figure 2.9 show the P_{n4} values computed from the χ^2_ν of the best-fitting bulge+disc models using free and fixed bulge Sérsic indices. Analogous to P_{pS}, the Legacy images offer little discriminating power between bulge+disc models of free and fixed n_b. Legacy decomposition P_{n4} values are concentrated around $P_{n4} = 0.5$ with almost no power at the $P_{n4} \approx 0$ or $P_{n4} \approx 1$. While many of these ambiguous P_{n4} are preserved in the co-add decompositions, a large fraction are not. As with the P_{pS}, the ability to discriminate between bulge models is sensitive to a galaxy's intrinsic brightness. The concentration in the co-add $P_{n4} - m_{\mathrm{r,s82,fn}}$ map is still broad, however. Nearly 50% of galaxies still have $0.25 < P_{n4} < 0.75$ in the co-adds. This number is less than 25% for P_{pS}.

2.4.6 Galaxy and bulge Sérsic indices

I have examined the differences in the distributions of galaxy (`ps` fits) and bulge (`fn` fits) Sérsic indices in the Legacy and the co-adds. First, I note the criterion of my methods that although Sérsic index is allowed to vary in each bandpass pair of the `ps` fits, the effective radius and axis ratio are held fixed to the gr results. Furthermore, since the $r-$band tends to have the best S/N, the optimization of Sérsic index will rely most strongly on the $r-$band photometry in cases when a target is faint in neighbouring bands. Simard et al. (2011) showed that the S/N and spatial resolution in Legacy is often limited in its capacity to characterize galaxy and bulge Sérsic indices – particularly for faint targets (hence a fiducial fixed $n_b = 4$ bulge in their two-component model). Simard et al. (2011) also argued and that the *particularly* large bump in the n_b distribution that they saw at $n_b = 4 - 5$ (compared to n_g which showed a more minor bump) is not some general property of physical bulges but is more likely a statistical artifact of computing the median of bulge Sérsic index, marginalized over all other parameters, in a likelihood space that does not respond sensitively to changes in n_b (i.e. a uniform posterior probability distribution with

hard limits $0.5 \leq n_b \leq 8$). Here, I determine whether the added depth from the co-adds offers an improved characterization of galaxy and bulge Sérsic indices given our foreknowledge of these existing systematics.

Figure 2.10 compares the distributions of galaxy Sérsic indices, n_g, from the `ps` model fits to each pair of bandpasses (ur, gr, ir, zr). The `ps` Sérsic index distributions for the ur, ir, and zr fits each show some modest differences that are qualitatively similar to the trend in B/T reported in Section 2.4.4. Fewer galaxies reside in the bump at $n_g = 4 - 5$ and the distribution is slightly more uniform at $n_g \gtrsim 3$ – with some galaxies moving from the bump to higher and lower n_g in the co-add fits. Interestingly, this qualitative trend is not shared in the gr `ps` fits (upper right panel, green and grey) – despite the fact that the $r-$band is used in every other fit *and* the other structural are fixed to the gr result. Indeed, the gr `ps` distributions for Stripe 82 appear very similar apart from the right-most bins at $n_g \gtrsim 7$ where there are *fewer* targets in the Stripe 82 relative to Legacy – contrary to what is seen in every other panel.

The upper panel from Figure 2.11 explores the apparent discrepancy in gr more closely by plotting the difference in Legacy and Stripe 82 n_g as a function of Stripe 82 `ps` apparent magnitude. At bright magnitudes, the median Stripe 82 `ps` n_g are larger. In particular, this result, combined with the increase in total brightness and sizes seen in Figures 2.3 and 2.7 in the deeper images, supports findings by other comparative works that have suggested that the GIM2D sky estimation method used in Simard et al. (2011) can be sub-optimal for bright targets with highly extended surface brightness profiles (Bernardi et al., 2014; Mendel et al., 2014; Meert et al., 2015). This systematic is less important in Stripe 82 (even though the same background estimation method and segmentation parameters are used) because the increase in S/N expands the sizes of the source masks in the segmentation images – forcing the background estimates to be made from pixels significantly farther from the target and every other source in Stripe 82 images. As shown by Mendel et al. (2014) (in their Appendix B) using artificial galaxy simulations, the systematic is most likely to arise in the Simard et al. (2011) fits to galaxies whose profiles are dominated by a component with intrinsically high Sérsic indices, $n \gtrsim 5$. Combining information from Figures 2.3, 2.7, 2.10, and 2.11 I caution that Simard et al. (2011) measurements for Legacy galaxies with high n_g and $m_r \lesssim 14.5$ mag *can* be compromised by these systematics which I have now quantified using deeper imaging. These results highlight the importance of artificial galaxy simulation recovery analysis and/or deep imaging

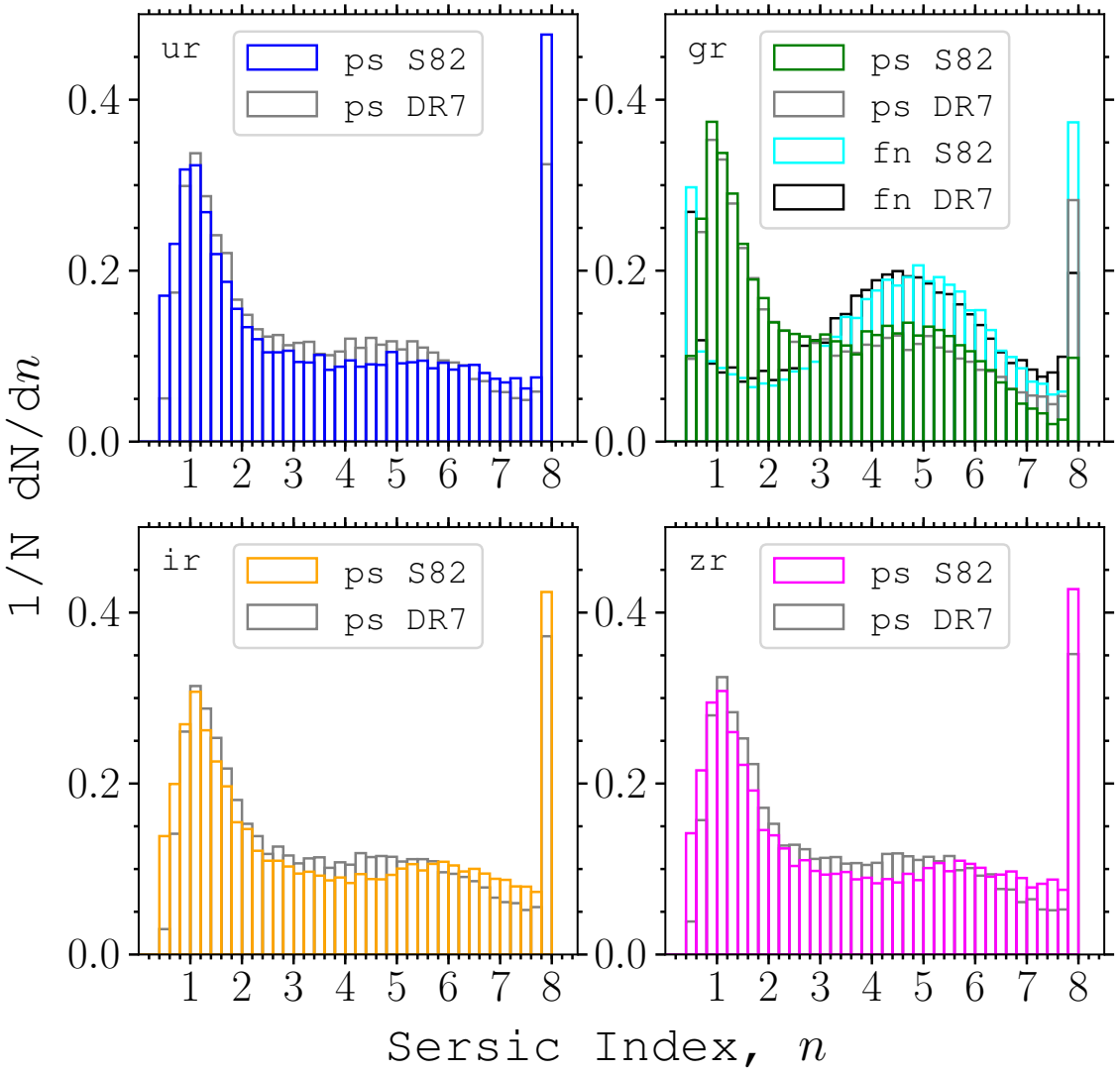

Figure 2.10 Comparison of distributions of Sérsic indices, n, in Stripe 82 and DR7 Legacy. Panels compare distribution functions of `ps` galaxy Sérsic indices, n_g, in the *ur*, *gr*, *ir*, and *zr* fits. Stripe 82 histograms are coloured according to each band pairing and the corresponding DR7 Legacy histograms are behind in grey. The *gr* panel to the upper right also compares the distributions of `fn` bulge Sérsic indices, n_b, from the corresponding *gr* fits (Stripe 82 – cyan; DR7 Legacy – black). The offsets in *gr* Sérsic index are shown as a function of apparent magnitudes in Figure 2.11.

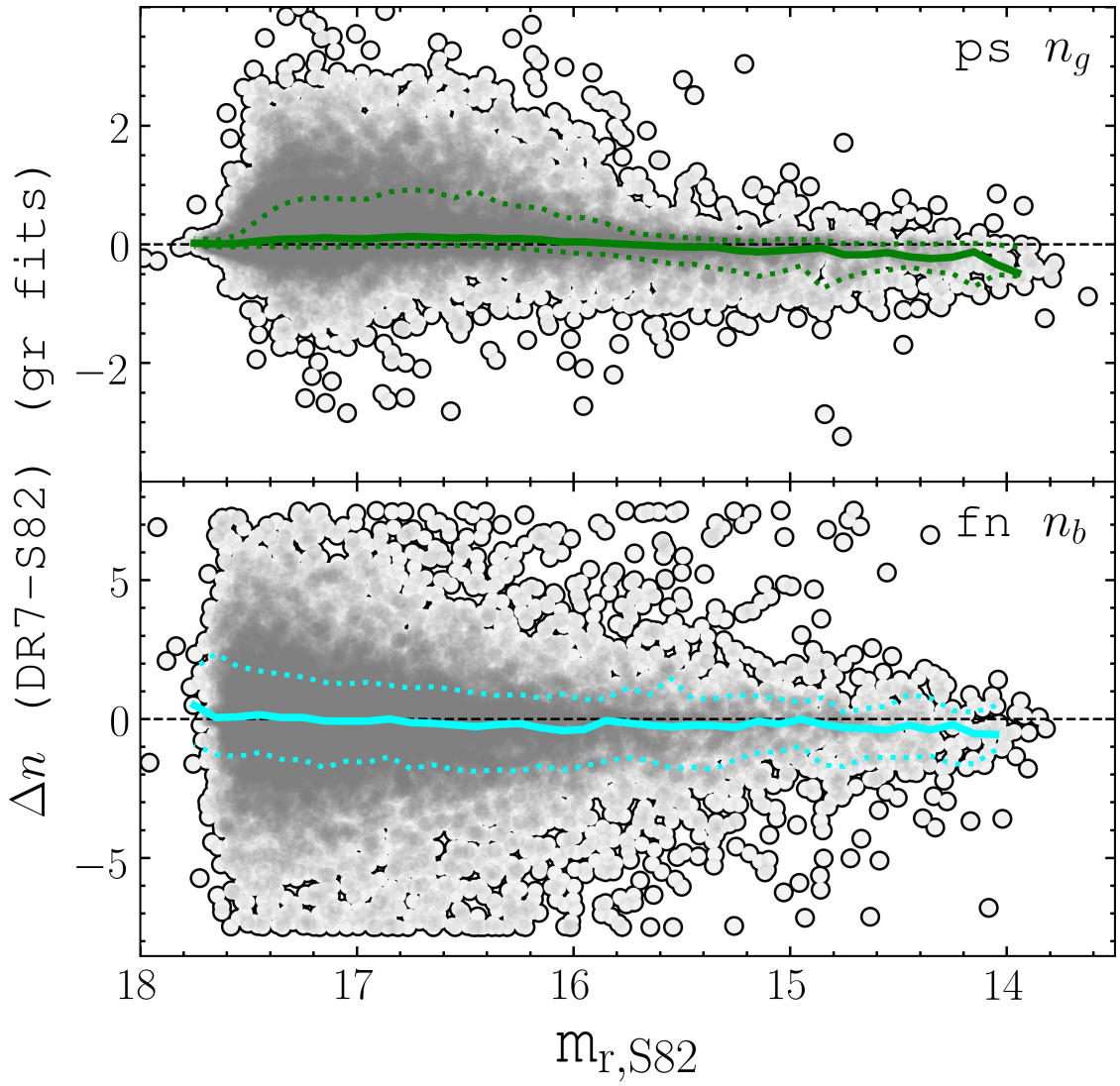

Figure 2.11 Comparison of the Stripe 82 and Legacy *gr* galaxy Sérsic indices from the `ps` fits (upper, green) and bulge Sérsic indices from the `fn` fits (lower, cyan) as functions of *total* $r-$band magnitude from the corresponding model. Solid lines show the median Δn as a function of apparent magnitude. Dashed lines show the 16th and 84th percentiles. Density maps in the background show the full sample distributions.

compliments to morphological analyses.

The upper panel from Figure 2.11 also shows the median trend that fainter targets tend to have mildly lower n_g in the `ps` fits for Stripe 82 relative to Legacy. Judging from the one-dimensional n_g distributions in the upper right panel for `ps`, these must be targets that had $n_g \gtrsim 7$ in Legacy but now have lower n_g in Stripe 82. The asymmetry of the scatter at the faint end supports this assertion. Given that this trend is reversed in the *ur*, *ir*, and *zr* fits, it is unclear what is driving this mild decrease in *gr* n_g for faint targets – particularly because the other structural parameters of the `ps` fits in those other bandpass pairs (effective radius, axis ratio, and position angle) are fixed to the *gr* results. The Legacy distributions, for example, do not differ in this way – they are practically identical in all bandpass pairs. The fact that Legacy n_g distributions do not change in these other bandpass pairs is mostly likely a consequence of fixing other structural parameters to the *gr* results and dominance of the higher S/N $r-$band images to the optimization of n_g in cases where there is low S/N in the u, i, or z images. As such, it could be argued that the improved S/N in the *uiz* images increases the covariance between *uiz* images *and* their $r-$band counterparts in optimizing n_g to both bands – resulting in unique distributions in each bandpass pair. But this argument does not also explain why the trend in the *gr* fits at high n_g is so suddenly reversed relative to fits in the other bandpass pairs. At this time, it is not immediately clear why the fainter sources have suppressed n_g in Stripe 82 relative to Legacy. However, I point out that this suppression is very modest – with the median offset not exceeding $\Delta n_g = 0.15$.

The upper right panel of Figure 2.10 also compares Stripe 82 and Legacy bulge Sérsic indices, n_b, from the `fn` fits to the *gr* images (cyan and black). The differences are minor with the exception of an increased number of objects in the final $n_b \approx 8$ bin. As with the `ps` fits to the *gr* bandpasses, I examine the difference more closely in the lower panel from Figure 2.11. The Δn axis has been broadened to better encompass the scatter (that predominantly arises due to the fact that I have not made any cuts on bulge brightness). I find similar systematics in n_b for `fn` fits to bright sources that were reported and discussed for the `ps` fits. As total galaxy brightness decreases, the scatter increases rapidly but with reasonable symmetry and no other particular systematics between deep and shallow images. Ultimately, I caution that the similarity in one-dimensional distributions of n_b in the `fn` decompositions implies that the majority of n_b measurements from the deep images are likely to be affected by the same systematics that affected n_b measurements in the shallow images (also

supported by the $P_{n4,S82}$ results in the previous section). Cuts in B/T (to focus on bulge brightness) or the more discriminate $P_{n4,S82}$ (to focus on objects for which the free n_b made a significant difference to the fitting result) can be used to suppress this scatter or to glean galaxies that are less affected (e.g., as in Figure 15 of Simard et al. (2011)) but may introduce biases depending on the science case.

2.4.7 Bulge and disc sizes

The F-test statistics show that the co-add images offer improved discrimination between single- and two-component systems compared to Legacy. I now compare the bulge and disc sizes to see how the component properties are affected. The upper left panel of Figure 2.12 shows the change in disc scale-length, $\Delta \log r_d$, plotted against disc apparent magnitude for the full range in B/T. There is excellent agreement between the median co-add and Legacy disc sizes all the way to $m_r \approx 18$ mag. The 16-84 percentile range in $\Delta \log r_d$ for discs brighter than $m_r \approx 18$ mag is also remarkably tight between the co-add and Legacy decompositions. However, at fainter disc magnitudes, the scatter is increased and Stripe 82 disc components are systematically larger than their Legacy counterparts. The median offset at faint disc magnitudes can be as large as $\Delta \log r_d \approx -0.07$ dex.

The lower left panel of Figure 2.12 shows the change in bulge effective radius $\Delta \log r_e$. The red shading in the marginals shows the relative fraction of bulges in each bin whose measured effective radii in the Stripe 82 or Legacy decomposition are less than associated the HWHM of the PSF:

$$F_{\rm bin,red} = \frac{N_{\rm bin}(r_{\rm e,S82} < {\rm HWHM_{S82}}\ {\rm OR}\ r_{\rm e,DR7} < {\rm HWHM_{DR7}})}{N_{\rm bin,tot}} \tag{2.20}$$

The same is shown for the discs, but very few disc sizes from either the co-adds or Legacy are below the HWHM of the PSF. Galaxies with the largest offsets in bulge size (those clipped to the edges of the right marginals) are those that are unresolved and faint (see lower marginals).

Stripe 82 bulge sizes are systematically larger than in Legacy for all bulge magnitudes (with median offsets as large as $\Delta \log r_e \approx -0.1$ dex). The scatter in the bulge size comparison is also far larger than for disc sizes. In general, bulges are intrinsically more difficult to fit than discs due to their high sensitivity to spatial resolution, sky subtraction, and the presence of additional substructures and components (such as

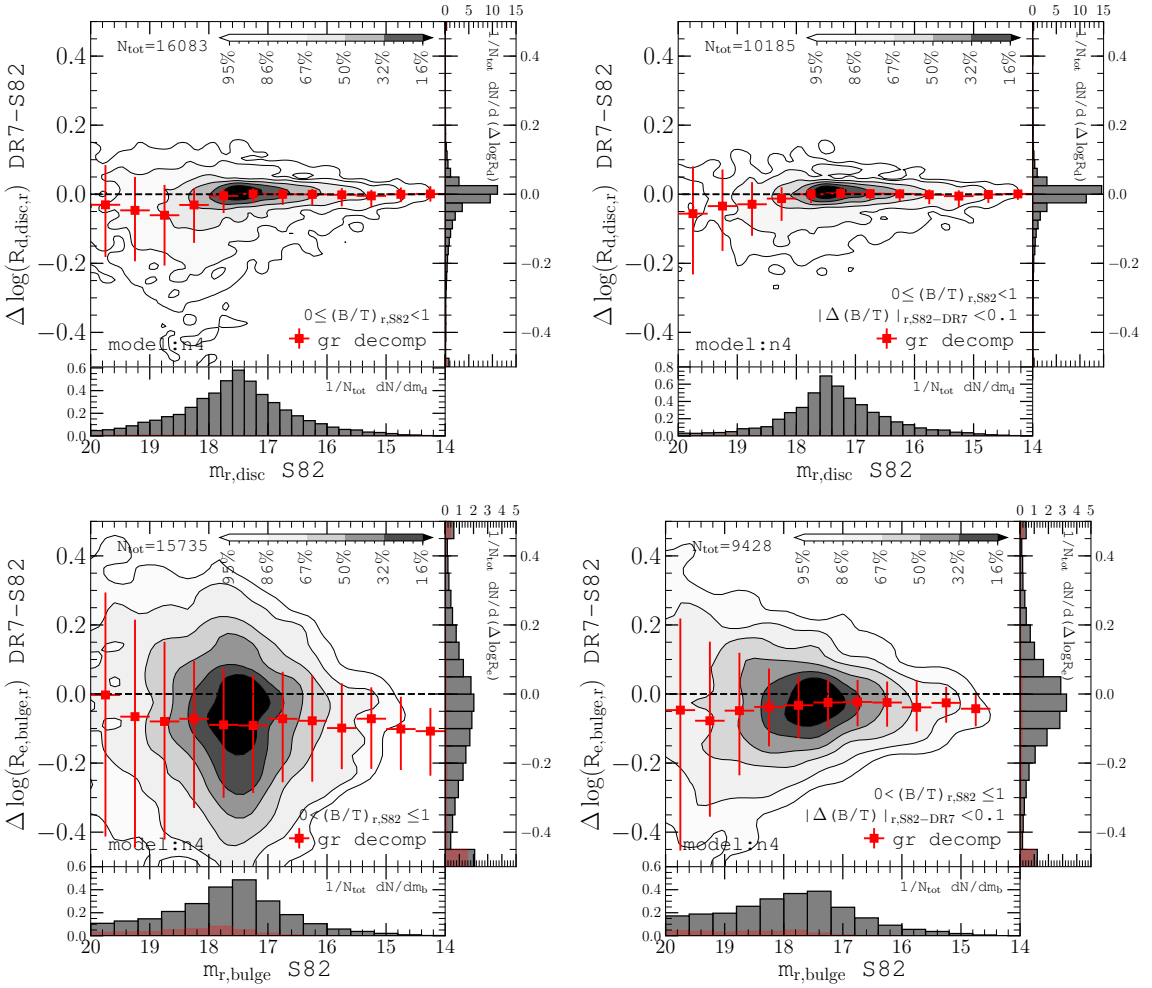

Figure 2.12 Comparison of disc scale lengths and bulge effective radii in Stripe 82 and Legacy. Markers show the median and 16-84 percentile range for $\Delta \log r_{\rm comp}$ in bins of component magnitude. Two-dimensional distributions are flanked by marginal histograms for each axis – where all values extending beyond the binning range have been clipped and added to the corresponding edge of the distribution. Red shading in marginal histograms indicate the relative fraction of galaxies per bin for which the Stripe 82 or DR7 size is smaller than the HWHM of the PSF following Eq. 2.20. All measurements are taken from the `n4` *gr* decomposition table and are by construction identical in each other bandpass pair (excepting their magnitude distributions). Markers show the median and 16th-84th percentiles in each 0.5 magnitude bin. Left panels: $\Delta \log r_{d,\rm disc}$ and $\Delta \log r_{e,\rm bulge}$ plotted against respective Legacy disc and bulge apparent magnitudes. B/T ranges are indicated at the lower left of each panel. Right panels: selection of galaxies from left panels with $|\Delta(B/T)|_{\rm r,DR7-S82} < 0.1$. This selection reduces systematics in bulge sizes and reduces the scatter in both bulge and disc comparisons.

bars). But there are several factors which can drive increased scatter and potentially introduce systematics.

First, changes in the sizes of intrinsically small and faint discs or bulges are exacerbated on a logarithmic scale. Since bulges are generally more compact than discs for the same luminosities (e.g., Graham & Guzmán 2003), the scatter in $\Delta \log r_e$ will generally be larger than in $\Delta \log r_d$.

Second, if a physical component is so small in angular size that it is spatially unresolved then scatter will be introduced by mismatches between the reconstructed PSF with which the models are convolved and the true PSF. Since the spatial resolution is 8% poorer on average in the co-adds than in Legacy, there is a slightly greater likelihood of having such mismatches in the co-adds.

Third, since galaxies in my sample all have $m_{\rm r,galaxy} \leq 17.77$ mag, a bulge (or disc) with brightness $m_{\rm r,comp} > 17.77$ mag often belongs to a system whose surface brightness distribution is dominated by the disc (or bulge). Applying a cut of $0.5 \leq B/T < 0.9$ in the disc comparison or $0.1 < B/T < 0.5$ to the bulge comparison reveals that there is greater scatter in $\Delta \log r_d$ for discs and $\Delta \log r_e$ for bulges that are embedded in two-component systems in which the secondary component dominates. Applying a $0 \leq B/T < 0.3$ cut to select only disc dominated systems reduces the scatter to $\Delta \log r_d$ to 0.05 dex. However, the same cannot be said of the bulge size comparison using a cut of $0.7 < B/T \leq 1$. I found that the scatter and systematics in the lower left panel of Figure 2.12 largely persist despite the cut.

The persistent systematics and biases among bulges raises a fourth driving source of contrast between the co-adds and Legacy component sizes – the increased bulge fractions in the co-adds for many galaxies with intermediate B/T. In the right panels of Figure 2.12, I control for this bias by strictly selecting galaxies which have a change in $|\Delta(B/T)|_{\rm DR7-S82} < 0.1$. Since total galaxy magnitudes are largely the same in the co-adds and Legacy, such a cut is the equivalent of saying that the component brightnesses must be the same as well – and so the result is unsurprising but demonstrates the contribution of this bias to the scatter. Most notably, the systematic in the bulge size comparison at faint bulge magnitudes shown in the lower left panel of Figure 2.12 is almost entirely eliminated in the lower right panel where $|\Delta(B/T)|_{\rm DR7-S82}$ is controlled. Thus the systematic at faint magnitudes and inflated scatter is only the consequence of the different (but typically larger) bulge-fractions measured in the co-adds.

The remaining scatter after my cut in $|\Delta(B/T)|$ is qualitatively consistent with

sources I have already discussed in the previous paragraphs of this section. Now, whether the tendency towards higher bulge-fractions are indeed a *correction* to systematically under-estimated bulge fractions for the fainter Legacy galaxies with intermediate B/T is not certain. But consider again the arguments I made in Section 2.4.4 regarding the artificiality of the bump at $B/T = 0.5$ in Legacy relative to the more uniform B/T distribution found in the co-adds. My results suggest that the bump at $B/T = 0.5$ (the median over the allowed range in bulge fractions) arises from poor constraints on the light profiles of faint galaxies. The characterization of these galaxies and their components is substantially improved in the deep co-adds as shown in the comparison of the F-test statistics. These pieces of evidence suggest correction to a systematic arising from poorer photometric constraints on faint bulges and discs in Legacy rather than a new and unexpected systematic in the co-adds.

As with the sizes, I have compared the brightnesses of the components. My findings are consistent with the analysis of the sizes. Many previously faint bulges get a brightness boost in the co-adds. The disc fractions are suppressed to compensate. It is worth noting that the comparison of component properties essentially functions as a convergence test. I have shown that it is possible to use my tables to estimate the brightness and component contrast (set by a galaxy's intrinsic bulge-to-disc, (B/D), light ratio) for which a bulge or disc component's physical properties hold-up at fainter surface-brightness limits.

2.4.8 Residual asymmetries and non-parametric indices

Six additional morphological indices are computed for each decomposition. The majority of these indices are aimed at quantitatively identifying visual indicators of disturbances in galaxy structures. The first two are the asymmetry, A, and concentration, C, indices from the automatic classification system (CAS) proposed by Abraham et al. (1994, 1996). Two more are D_z and A_z, defined in Section 5.6 of Simard et al. (2002). A_z measures the flux from all pixels that are $n\sigma_{\rm sky}$ higher than their symmetric counterparts when rotated 180° about the target's barycenter, normalized by the total object flux. A_z is computed for $n = 2, 3, 5$ within circular apertures extending either one or two half-light radii from the target centroid. Similarly, D_z is the sum of the fluxes of target object pixels, as determined from the SExtractor segmentation image, whose symmetric counterparts are not also target object pixels. D_z is similar in nature to the shape asymmetry parameter first proposed by Pawlik

et al. (2016) which is sensitive to crowding by neighbouring sources and potentially useful in quantitatively identifying close galaxy pairs.

The remaining two indices are based on the R_T and R_A indices used in local studies of spiral arm patterns by Elmegreen et al. (1992) and first applied to distant galaxies by Schade et al. (1995) as part of the Canada-France Redshift Survey (Lilly et al., 1995). R_T and R_A each quantify the residual light that is not characterized by the symmetric, analytic models used in the decomposition. For both indices, the residual model-subtracted image is either added to (R_T) or subtracted (R_A) by the same image rotated by 180 degrees about the galactic centre.[12] As such, asymmetric features which, in particular, are known to be indicators of galaxy interactions such as tails and bridges may be identified. Calculation of R_T and R_A in the GIM2D pipeline differs slightly from Schade et al. (1995) in that the GIM2D indices are computed within one, two, and three multiples of a galaxy's measured half-light radius. Table values are correspondingly named R{A,T}{1,2,3}_1_{BAND}.

Figure 2.13 compares the R_A indicators measured for the co-add and Legacy decompositions using the difference ΔR_A. The right panel shows the comparison as a function of magnitude. There is a uniform median enhancement in the co-add R_A (by $\Delta R_A \approx 0.07$) across the full magnitude range. However, in the left panel, we see that this enhancement in asymmetry for the co-adds is largely attributed to galaxies with some degree of asymmetry already measured in Legacy. I have confirmed these findings through visual inspection of galaxies with large ΔR_A which either had small or large R_A originally in Legacy. In particular, the co-adds often reveal HII regions, tidal tails, shells, and streams in galaxies that were hardly detectable in the shallower Legacy images. Taken together, these results demonstrate that asymmetric features are generally enhanced in the deeper images insofar as they may be distinguished from the sky background. The standard and residual asymmetric features that are revealed through deep imaging may be exploited to construct more robust samples of galaxies that are expected to contain such asymmetric features such as peculiar early-type galaxies (e.g., Kaviraj 2010) and recent/ongoing mergers (e.g, Ellison et al. 2019).

[12]The R_T and R_A indices are known to be sensitive to the pivot point as for the $C-A$ indices Conselice et al. (2000). I do not include a step to find pivot point in the image about which the asymmetry indices are minimized.

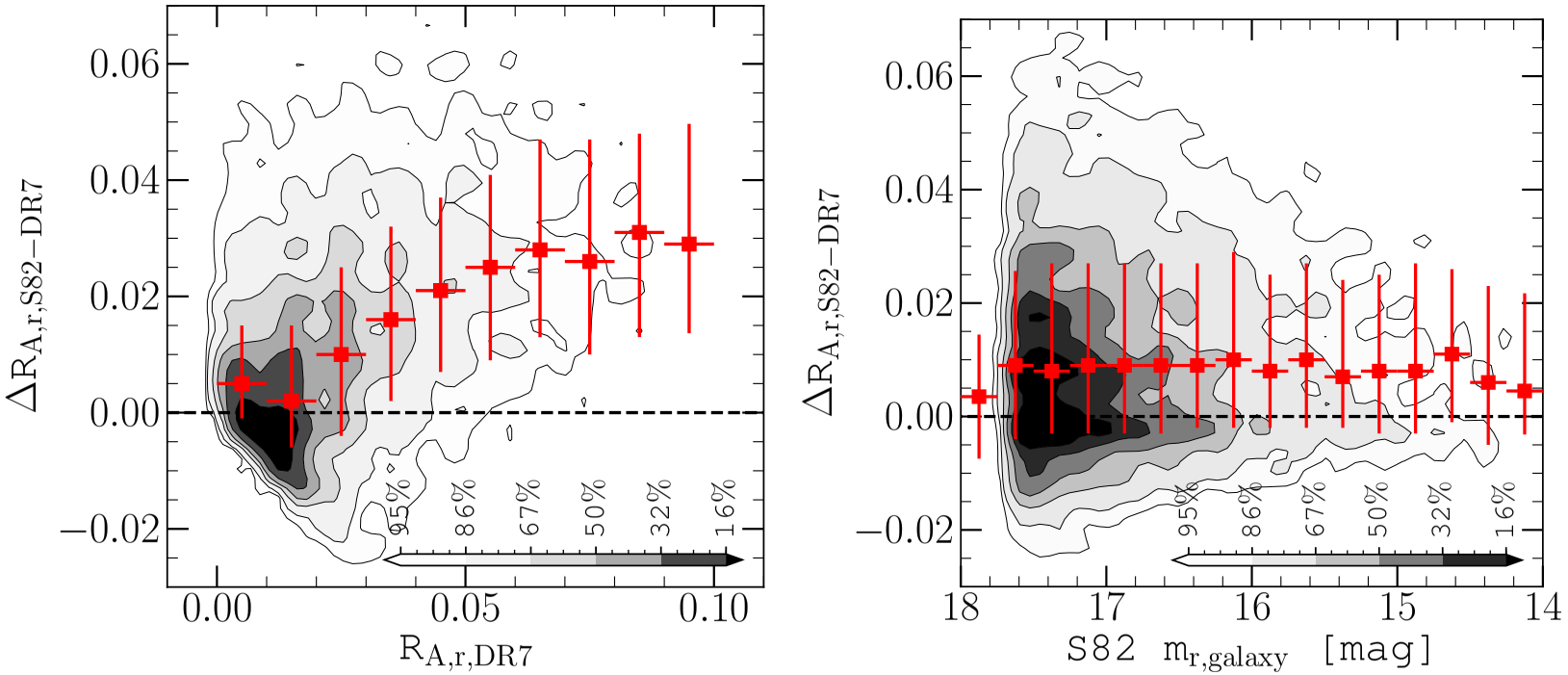

Figure 2.13 Change in residual asymmetries, ΔR_A between Legacy and the Stripe 82 co-adds. R_A is computed within three half-light radii in each case. The left panel shows the enhancement as a function of the original R_A measurement from the Legacy r-band decompositions. The right panel shows the enhancement as a function of Stripe 82 total apparent magnitude. Asymmetries are uniformly enhanced as a function of magnitude, but are preferentially enhanced in targets that had large asymmetries in Legacy images.

2.5 Summary

Using images constructed from multiple exposures of 275 deg^2 in the SDSS Stripe 82, which permit an additional $1.6-1.8$ magnitudes of depth with respect to single-exposure SDSS Legacy images, I have performed free-n_b and $n_b = 4$ bulge+exp. disc (`fn` and `n4`) as well as single-component Sérsic `ps` decompositions on 16,908 galaxies – obtaining parametric morphologies in the u, g, r, i, z bands. I make all of my catalogs publicly available. The catalog structures are consistent with those from Simard et al. (2011), enabling straightforward comparisons of structural measurements using deep and shallow photometry. A sample table structure is provided in Appendix B. My main findings are summarized as follows:

- **Integrated galaxy properties in each band such as size and luminosity are largely unchanged (Fig. 2.3).** For example, the median offsets in total apparent magnitude for all galaxies in each band are $\Delta m^{50\%}_{u,g,r,i,z} = (0.037, 0.022, 0.021, 0.020, 0.019)$ mag (DR7-S82) for the `n4` decompositions. Exceptions arise for the brightest galaxies in each bandpass, where broader target masks and correspondingly improved local sky estimation yields brighter results in the co-adds. Also, many galaxies with u-band intrinsic brightnesses that are poorly constrained or resulted in non-detections in Legacy are actually measured and characterized in the co-adds. This systematic manifests as large negative offsets between the Legacy and Stripe 82 decomposition u-band magnitudes.

- **The colour-magnitude diagrams for the co-adds have a tighter red-sequence in all bands with respect to Legacy decompositions (Fig. 2.6).** The removal of outliers is particularly noticeable in the u-band relative to the Legacy colours. The majority of the outliers in Legacy have large, positive $(u-r)$ colours. Given that their r-band magnitudes are unchanged in the co-add decompositions, the high $(u-r)$ colours indicate that many galaxies' u-band magnitudes were systematically underestimated in Legacy. The reduced scatter in the red-sequences is also an indicator that deblending of light for crowded targets is improved – as satellites and other potential interlopers are now more efficiently masked due to the high S/N in the co-adds.

- **Bulge-to-total fractions for many galaxies with intermediate B/T are enhanced in the co-add decompositions, but are largely unchanged at**

low B/T (Fig. 2.8). Importantly, the changes to B/T appear to remove a systematic aggregation of galaxies around $B/T = 0.5$. I argue that the grouping at $B/T = 0.5$ is not physical, but is a systematic in the Simard et al. (2011) decompositions owing to poor constraints on the structures of galaxies at the limits of Legacy photometry.

- **A number of galaxies with $B/T = 1$ in the shallow images get a boost to their disc fractions in the deeper images (Fig. 2.8).** The median $\Delta(B/T)_{\mathrm{S82-DR7,r}} \approx -0.09$ for galaxies with $B/T \approx 1$ in Legacy. However, this reduction in bulge fraction can be up to 40% for such galaxies.

- **I find that the deep imaging improves the discriminating ability of the F-test in determining whether a more complex decomposition model is favoured with respect to a simpler model (Fig. 2.9).** Where the majority of galaxies in the Legacy decompositions have $P_{pS,DR7} \approx 0.5$ and $P_{n4,DR7} \approx 0.5$ (indicating no statistical preference between decomposition models), the co-add $P_{pS,S82}$ and $P_{n4,S82}$ are bimodal over a large range in galaxy apparent magnitudes. I argue the improved photometry in the co-adds enables better characterization of the components and, in particular, an improved capacity to determine whether a two-component model is favoured over a single-component model.

- **I show that systematics on galaxy and bulge Sérsic index in the `ps` and `fn` fits are suppressed in the deeper images (Figs. 2.10 and 2.11).** More accurate sky measurements and reduced sky uncertainties yield improved characterization of the extended profiles of galaxies with both $m_r < 14.5$ *and* high Sérsic indices whose fluxes, sizes, and Sérsic indices can be suppressed by over-estimated skies in the Simard et al. (2011) Legacy measurements. These findings using deep imaging agree with systematic uncertainty predictions by Mendel et al. (2014) from artificial galaxy recovery analysis. They also agree with results of comparisons between the Simard et al. (2011) Legacy catalogs to other decomposition analyses with alternative sky-estimation procedures (e.g., Bernardi et al. 2014; Meert et al. 2015).

- **Bulge and disc sizes are consistent out to $m_{r,\mathrm{comp}} \approx 17$ (Fig. 2.12).** I assert that this is the magnitude limit at which Legacy measurements of components can be trusted before high measurement uncertainties and systematics

dominate. The scatter in the component sizes increases rapidly with decreasing brightness. I showed that this scatter arises from systems in which the one component dominates over the other and where the photometry of a component is limited in the shallower Legacy images. Using cuts in B/T, the scatter in the disc sizes can be reduced to $\Delta r_{d,disc,r} < 0.05$ dex. The higher bulge fractions in many faint galaxies drive a systematic enhancement of bulges sizes in the co-adds. I caution, however, that users of my catalogs must consider their science case before making such cuts – which may introduces biases.

- **Measures of asymmetry are enhanced in the deeper Stripe 82 co-adds with respect to Legacy measurements (Fig. 2.13).** I compute a range of asymmetry parameters for each galaxy and decomposition that are included in the catalogs. I find that asymmetries are generally enhanced in the Stripe 82 co-adds and that the enhancement is strongest for galaxies that were already asymmetric in Legacy. This result highlights an empirical model (or threshold) for identifying intrinsically asymmetric galaxies (such as ongoing/recent mergers) on the basis of asymmetry measurements are highly specific to the surface-brightness limitations of the corresponding survey. The implication is that such empirical models must be specifically calibrated for any given survey.

In general, the changes to morphological and statistical parameters in the Stripe 82 co-adds highlight the importance of deep imaging compliments to large photometric galaxy surveys. In particular, such deep imaging enables valuable and straightforward validation of galaxy structural measurements and the magnitude limits to which they are robust.

Chapter Acknowledgements

I thank James Annis, Jürgen Fliri, and Ignacio Trujillo for useful discussions on their respective Stripe 82 stacks. I gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). I also thank McGill University and the McGill Space Institute for being generous hosts during the completion of this project. The surface brightness decomposition pipelines used in this project were originally written for the SDSS Data Release 7 by Luc Simard (LS). I adapted these procedures for the SDSS Stripe 82 imaging. LS and J. Trevor Mendel provided valuable technical advice. Sara L. Ellison provided scientific advice. I also

thank the anonymous referee for the corresponding publication to this chapter whose suggestions helped to improve the quality of the work.

This project made use of public data from the Sloan Digital Sky Survey. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Chapter 3

Predicting the stages of galaxy mergers from images

Preamble

In the previous chapter, I showed that depth plays an important role in characterizing the detailed morphological structures in galaxies (e.g. bulges, discs, and asymmetric structures). In particular, I showed that deep imaging provides an enhanced ability to identify asymmetric features. Such features are often associated with mergers, and indeed asymmetry has often been used to identify mergers. In this chapter, I employ tools and develop methods that go beyond individual morphological parameters to identify and *characterize* galaxy mergers. Specifically, I use a combination of hydrodynamical simulations, synthetic observations, and deep learning to build and calibrate a model for identifying and estimating the *stages* of galaxy mergers from images. This chapter also focuses on the importance of *realism* in the synthetic training images. That is, I examine the degree to which realistic observational effects and limitations (such as those explored in Chapter 2) must be incorporated into the synthetic training imaging if a model can be expected to successfully characterize galaxies in real images. The work presented in this chapter is published in Bottrell et al. (2019a)[1].

[1] `https://doi.org/10.1093/mnras/stz2934`

3.1 Introduction

Theoretical predictions and observations alike show that mergers transform galaxies. Stellar bridges and tails observed in interacting galaxy pairs are the relics of the strong gravitational and tidal forces involved in close galaxy-galaxy encounters (Toomre & Toomre, 1972). But the consequences of these forces extend well beyond immediate changes to visual morphology.

Tidal torques and shocks excited by close encounters can rapidly reduce angular momentum in the dynamically cold interstellar medium (ISM) through various channels – all driving inflow of available cold gas towards the centres of interacting galaxies (e.g. Hernquist 1989; Barnes & Hernquist 1992; Mihos & Hernquist 1996; Hopkins & Quataert 2010; Blumenthal & Barnes 2018). There is now a strong numerical and observational framework linking this rapid and central accumulation of gas to boosts in central star formation rates (e.g. Ellison et al. 2008; Patton et al. 2011; Hopkins et al. 2013; Patton et al. 2013; Moreno et al. 2015; Sparre & Springel 2016; Thorp et al. 2019), dilution of central gas phase metallicity (e.g. Kewley et al. 2006; Ellison et al. 2008; Rupke et al. 2010a,b; Sol Alonso et al. 2010; Perez et al. 2011; Torrey et al. 2012a; Moreno et al. 2015; Thorp et al. 2019) and accretion onto central black holes and triggering of active galactic nuclei (AGN, e.g. Keel et al. 1985; Di Matteo et al. 2005; Koss et al. 2010; Ellison et al. 2011; Comerford et al. 2013; Satyapal et al. 2014; Ellison et al. 2015; Barrows et al. 2017; Goulding et al. 2018; Ellison et al. 2019; Foord et al. 2020). Additionally, galactic outflows of gas associated with the enhancements in star-formation rates (e.g. Martin 2005; Rupke et al. 2005b; Strickland & Heckman 2009; Hayward & Hopkins 2017; Müller-Sánchez et al. 2018) and AGN activity (e.g. Rupke et al. 2005a; Veilleux et al. 2013; Zschaechner et al. 2016; Woo et al. 2017; Müller-Sánchez et al. 2018) can also be triggered by mergers – resulting in the growth and enrichment the circum-galactic medium (e.g. Johnson et al. 2015; Hani et al. 2018). Combined with the role of mergers in the assembly of present-day galaxies (e.g. White & Rees 1978; Blumenthal et al. 1984) and transforming their morphologies and kinematics (e.g. Toomre 1977; Negroponte & White 1983; Hernquist 1992; Naab & Burkert 2003; Hopkins et al. 2008b; Berg et al. 2014) these connections make mergers complex but unique laboratories for testing some of the most crucial aspects of galaxy formation physics.

One observationally measurable parameter that is particularly valuable for testing the statistical and cosmological role of mergers in galaxy evolution (and which is di-

rectly comparable to numerical predictions from semi-analytic models or cosmological hydrodynamical simulations) is the galaxy merger rate and its evolution with mass and redshift (e.g. Lacey & Cole 1993; López-Sanjuan et al. 2011; Lotz et al. 2011; Bluck et al. 2012; López-Sanjuan et al. 2013a; Casteels et al. 2014; Rodriguez-Gomez et al. 2015; Martin et al. 2018). Estimating the merger rate requires: (1) a method with which mergers can be distinguished from non-merging galaxies and (2) an estimate of the timescales on which the distinction can be made – which is sensitive to the method used in the former (Hopkins et al., 2008c; Lotz et al., 2008, 2010b,a). However, beyond identifying mergers, we are also particularly interested in predicting merger *stage*. Hydrodynamical simulations of galaxy mergers predict significant evolution in (among others) star-formation rates, ISM content, AGN accretion rates and luminosities, and subsequent stellar and AGN feedback along the merger sequence (e.g. Cox et al. 2008; Hopkins et al. 2008a; Torrey et al. 2012a; Hopkins et al. 2013; Moreno et al. 2015, 2019). Consequently, in order to test the broader and detailed elements of this framework (such as feedback and outflow prescriptions), we must be able to (i) obtain large and reasonably complete observational samples of galaxy mergers and (ii) connect observed galaxy interactions to specific stages in the merger sequence.

Both of these tasks present significant challenges from an observational perspective. For example, while mergers and recent post-mergers can be selected visually on the basis of distinct (but often low surface-brightness) morphological features such as tidal tails, bridges, streams, shells and nearby companions (Darg et al., 2010; Kartaltepe et al., 2015; Simmons et al., 2017), this process is subjective and sensitive to contrast, resolution and surface-brightness limits. For pair candidates, the intrinsic subjectivity of visual classification can be alleviated by obtaining relative velocities with spectroscopy. Spectroscopic pair identification is effective even at high-redshifts (Lin et al., 2007; Wong et al., 2011), but is often incomplete due to the "fibre-collision" problem and sparse sampling – which particularly affect close pair completeness (Patton et al. 2002; Lin et al. 2004; Patton & Atfield 2008, but see also Robotham et al. 2014). Fast and reproducible classifications can be made using automated quantitative morphologies (Conselice, 2003; Lotz et al., 2004; Pawlik et al., 2016; Rodriguez-Gomez et al., 2019). These metrics are designed to exploit the excess asymmetries, disturbed morphologies or multiple nuclei of mergers and merger-remnants relative to non-merging galaxies (e.g. Casteels et al. 2013, 2014; Patton et al. 2016). The main obstacle with quantitative morphologies is defining empirical thresholds which

separate merger from non-merger classes. Like visual classification, these thresholds (particularly for asymmetries) are sensitive to resolution and surface brightness limits (e.g. Ji et al. 2014; Bottrell et al. 2019b) but also, critically, the "training" data with which these thresholds are calibrated.

To calibrate an empirical threshold for a metric which separates merger from non-merger classes, one must have a way of evaluating its performance (e.g. completeness and/or purity). The significant limitation of calibrating on observational data is that the subjective and non-subjective biases afflicting visual classifications and the incompleteness of spectroscopic samples become embedded in the calibration. In other words, observationally, one does not have access to the ground truth. This limitation can be overcome using synthetic images from hydrodynamical merger simulations as the basis for the calibration step – which simultaneously solves the problems above (Lotz et al., 2008, 2010b,a; Nevin et al., 2019). Regardless of any added ingredients to the synthetic images (sky noise, resolution degradation, additional sources, etc.), the simulations provide foreknowledge of the true properties of each target: merger stage, mass ratio, gas fractions, initial morphologies, orbital parameters, etc. Consequently, one always has unbiased target classes upon which to evaluate the performance of the method. Furthermore, one can measure the biases affecting performance from both merger and image properties. Lastly, since the morphological features of galaxy interactions are induced primarily via gravitational effects, they should be largely insensitive to the particularities of the hydrodynamic model.

Two classes of methods which have gained significant traction in general astronomy and, in particular, galaxy astronomy are machine learning and deep learning (e.g. Teimoorinia & Ellison 2014; Hezaveh et al. 2017; Bluck et al. 2019; Ribli et al. 2019a; Jacobs et al. 2019; Snyder et al. 2019; Ntampaka et al. 2019; Hausen & Robertson 2019). Specifically, convolutional neural networks (CNNs) have been used to improve automated image-based galaxy morphology classifications with great success (Huertas-Company et al., 2015; Domínguez Sánchez et al., 2018). The level of intricacy in the features that can be identified by CNN and other machine learning models has made them an attractive tool for merger identification (e.g. Ackermann et al. 2018; Walmsley et al. 2019). Following the approaches adopted for quantitative morphologies by calibrating on hydrodynamical simulations, Pearson et al. (2019) train a CNN on synthetic SDSS images of galaxies from the EAGLE simulation (Schaye et al., 2015) and examine biases from redshift, star-formation rates, and apparent brightness on merger and non-merger classifications – though with poor

classification performance (65.5% in a binary classification). Nonetheless, one of the key elements of their synthetic SDSS images is that they were inserted into a handful of SDSS survey fields in an attempt to match observational biases in real images (realistic skies, resolution, and crowding by nearby sources). Indeed, Huertas-Company et al. (2019) used a similar but more rigorous approach with CNNs trained on the Nair & Abraham (2010) SDSS visual classification sample to perform Hubble type classifications of synthetic images of galaxies from the IllustrisTNG-100 simulations (Pillepich et al., 2018a; Nelson et al., 2018; Rodriguez-Gomez et al., 2019). Huertas-Company et al. (2019) found that injecting the TNG images into real fields following the statistical observational realism approach of Bottrell et al. (2017a) was crucial to obtaining consistent classification uncertainties when testing on SDSS and TNG images.

These previous studies touch upon core unanswered questions for training deep neural networks based on hydrodynamical simulations (and that are particularly relevant for characterizing merger stage). Namely, what kind of synthetic images should be used when training using simulations? How realistic do the images have to be in order to achieve high performance in identifying and characterizing mergers by stage in real images? Does a network that is trained on images which include contaminating effects (such as realistic skies, resolution degradation and additional sources in the image field-of-view [FOV]) perform better when handling new data which also contain these contaminants? In other words, what is gained by making synthetic images more realistic? Dust-inclusive radiative transfer can be used to generate photo-realistic images (e.g. Jonsson 2006; Jonsson et al. 2010; Baes et al. 2011; Camps & Baes 2015) but it is costly from computational and data storage perspectives. Is radiative transfer essential to merger classifications or can it be replaced with simpler images? These questions come at an important time when the state-of-the-art cosmological hydrodynamical simulations produce realistic and statistically representative populations of galaxies (e.g. IllustrisTNG; Pillepich et al. 2018a; Nelson et al. 2018). Crucially, mergers identified from these simulations' merger trees cover a range of mass ratios, orbital parameters, and initial galaxy properties that are comparable to the real Universe. Consequently, synthetic images generated along each merger sequence can be used to generate and calibrate deep network models to identify and characterize mergers in current and next-generation observational imaging surveys.

The goals of this chapter are to: (1) provide the methodology with which CNNs, trained and calibrated using hydrodynamical simulations, can be used to identify

mergers and predict merger stage in realistic images and (2) assess the importance of realism in the synthetic training images.[2] To realize these goals, I construct synthetic images with various levels of observational realism from a set of hydrodynamical binary merger simulations run with the FIRE-2 model (Moreno et al., 2019). Specifically, images are generated in two branches, starting with: (a) 2-D projections of the stellar particles and (b) photometry from dust-inclusive SKIRT radiative transfer. In each branch, images are are constructed with three levels of realism: (i) no observational effects (idealized); (ii) realistic skies and point-spread functions (semi-realistic); and (iii) statistical insertion into real survey images (fully realistic). These levels are designed to expose the roles of particular ingredients of realism in classification performance. Each image is assigned a target classification (isolated, pair or post-merger) corresponding to the definitions described in Section 3.2.1 and illustrated in Figure 3.1. Given that all training, validation, and test data are drawn from the same set of isolated/merger simulation runs, the training data are (by construction) highly generalizable to the test data in terms of the range of galaxy/merger properties covered. This experimental design allows me to isolate the role of realism in the performances of the networks.

This chapter is laid out as follows. The simulations, construction of the synthetic images, and neural network architecture are described in Section 3.2. My experiments and their results are presented in Section 3.3. My results are discussed in Section 3.4 and summarized in Section 3.5. I adopt a cosmology in which ($H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$). Additionally, the Bottrell et al. (2017a) observational realism suite, REALSIM, was released publicly as a companion to the publication associated with this chapter (see Section 3.2.2).

3.2 Methods

In this section, I describe the merger simulations (Sections 3.2.1 and 3.2.1), merger stage definitions and snapshot selection (Section 3.2.1), creation of the synthetic images (Section 3.2.2) and CNN architecture (Section 3.2.4).

[2]It should be noted that, based on the results of Huertas-Company et al. (2019) who use a similar approach for Hubble type classifications, the applications of my methods and results are not restricted to mergers.

3.2.1 Simulations

I use the suite of galaxy interaction simulations from Moreno et al. (2019) in this study. The suite is similar to a previous merger suite from those authors (Patton et al., 2013; Moreno et al., 2015) but with much higher resolution and a new physical model and hydrodynamic solver. I describe the salient features of the suite here but refer the reader to Moreno et al. (2019) and Hopkins et al. (2018) for full details of the suite and model, respectively.[3] I discuss the limitations with respect to the scope of the merger suite in detail in Section 3.4.2. Briefly, I emphasize that the suite does not offer sufficiently representative statistics and diversity in galaxy/merger properties to train networks that will be useful in applications to a real population of galaxies. Therefore, I do not apply my trained networks to real galaxies. However, for the objectives highlighted at the end of Section 3.1, the suite is appropriate.

FIRE-2 model

The simulations were run by Moreno et al. (2019) using the FIRE-2 physics model (Hopkins et al., 2018) and the "meshless finite-mass" (MFM) hydrodynamics solver, GIZMO (Hopkins, 2015, 2017).[4] The model includes treatment of radiative cooling and heating from free-free, photo-ionization and recombination, Compton, photoelectric, dust-collisional, cosmic ray, molecular, metal-line and fine-structure processes. It accounts for the UV background (Faucher-Giguère et al., 2009) and locally-driven heating and self-shielding. Gas which is locally self-gravitating, self-shielding, Jeans unstable and sufficiently dense (defined by critical gas density, $n_{\rm crit} = 1000$ cm^{-3}) can form stars stochastically in a sink-particle approach (see Appendix C of Hopkins et al. 2018). A stellar particle is treated as a single stellar population with a known age, $t_\star = t - t_{\rm form}$, and a metallicity and mass which are inherited from its progenitor gas particle. Masses, ages, metalliticities, luminosities, energies, mass-loss rates and stellar feedback event rates are tabulated (without tuning) using the STARBURST99 stellar population synthesis model (Leitherer et al., 1999) assuming a Kroupa (2001) initial mass function (IMF). Stellar feedback includes: (i) mass, metal, energy, and momentum injection from supernova type Ia & II; (ii) continuous stellar mass-loss through OB and Asymptotic Giant Branch (AGB) winds; (iii) photo-ionization and photo-electric heating; and (iv) radiation pressure. The model does not account

[3]Videos of the Moreno et al. (2019) galaxy merger simulations are available at research.pomona.edu/galaxymergers.

[4]For more information on the FIRE Project and FIRE-2, visit https://fire.northwestern.edu.

Table 3.1 Initial properties of the four galaxies in the Moreno et al. 2019 merger suite. The columns are the galaxy ID, total stellar mass, halo mass, gas fraction, stellar bulge-to-total mass fraction, gas disc scale length, and stellar disc scale length.

Galaxy ID	$M_\star/10^{10}M_\odot$	$M_\star/M_{\rm halo}$	$f_{\rm gas}$	$(B/T)_\star$	$R_{\rm d,gas}$/kpc	$R_{\rm d,stars}$/kpc
G1	0.206	0.0157	0.681	0.0185	4.73	1.42
G2	1.24	0.0361	0.392	0.0497	6.04	1.92
G3	2.97	0.0383	0.264	0.0773	5.32	1.61
G4	5.50	0.0228	0.192	0.103	5.26	1.57

for feedback generated via accretion of gas onto supermassive black holes (SMBHs). SMBH feedback is omitted because coupling between an AGN and the circumnuclear interstellar medium (ISM) is not yet well understood (though see Torrey et al. 2017 for an examination of the stability of feedback regulated star-formation in galactic nuclei). The MFM dark matter, gas, and stellar particle masses are $(m_{\rm dm}, m_{\rm gas}, m_{\rm star}) = (19, 1.4, 0.84) \times 10^4\ M_\odot$. The highest gas density and spatial resolution are 5.8×10^5 cm^{-3} and 1.1 pc, respectively. The typical snapshot resolution is 5 Myr. The gravitational softening lengths are 10 pc for dark matter and stellar components and 1 pc for the gaseous component.

Merger suite

Moreno et al. (2019) used FIRE-2 physics to generate a suite of non-cosmological binary galaxy interaction simulations covering a range of orbital parameters and mass ratios between four disc galaxies (G1, G2, G3, and G4, in order of increasing total stellar mass). The suite is complimented by secular runs (the controls used by Moreno et al. 2019) in which each galaxy is allowed to evolve in isolation. Individual galaxies are set up following the procedure described in Springel et al. (2005a) using the analytic framework provided by Mo et al. (1998). Stellar bulges and dark matter haloes are initialized analytically with Hernquist (1990) profiles. Halo masses are adopted for a given stellar mass following the abundance matching results from Moster et al. (2013). Stellar bulge-to-total fractions are assigned on the basis of median trends with total stellar mass using the Mendel et al. (2014) estimates of bulge, disc, and total stellar masses for galaxies in the SDSS. Similarly, gas fractions are assigned based on mean atomic and molecular gas mass fractions estimates along the main sequence (MS) of star-forming galaxies in the SFR-$M_\star$ plane from Saintonge et al. (2016). The properties of each galaxy are shown in Table 3.1.

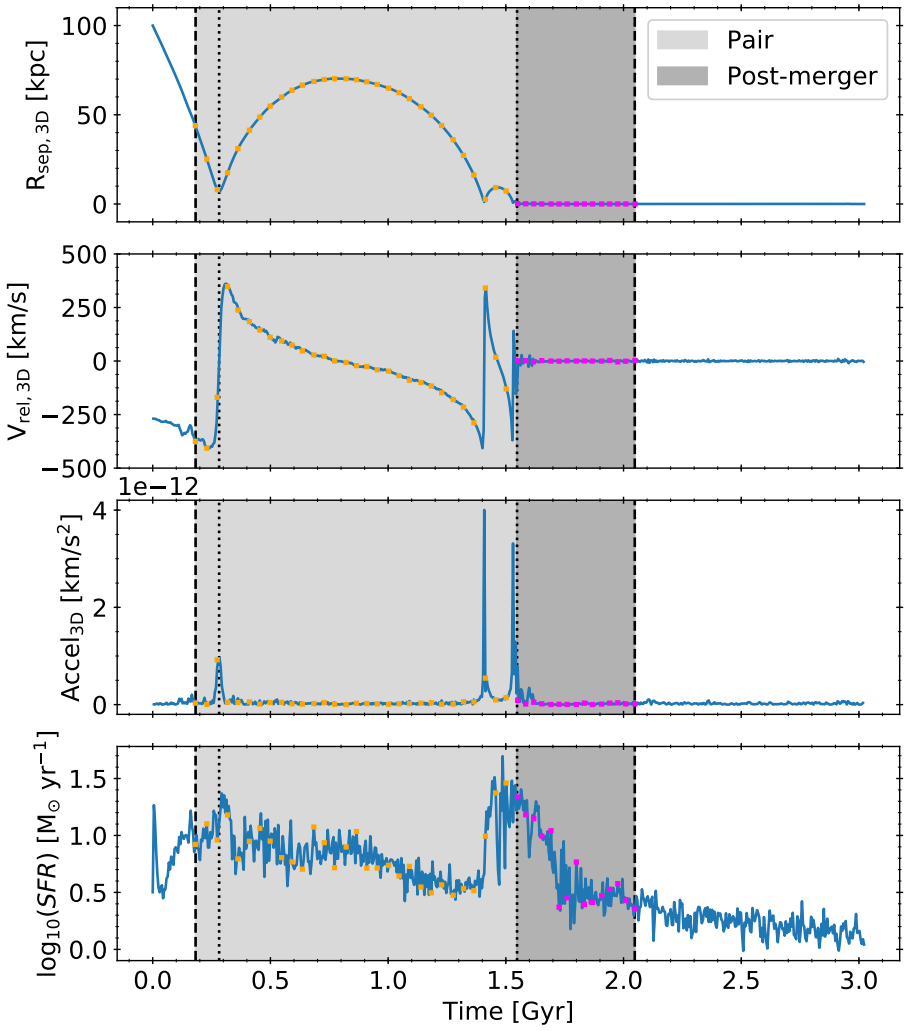

Figure 3.1 Radial separation, relative velocity, absolute acceleration, and total star formation rate (SFR) sequences for the G2G3 "e" orbit 1 merger (the fiducial run from Moreno et al. 2019) showing class definitions and snapshot selection. From left to right in each panel, the first thick dashed and dotted lines correspond to 100 Myr before first pericentric passage and the moment of first pericentric passage, respectively. The second dotted line corresponds to coalescence – which is taken to be the last time the central black holes of each galaxy are more than 500 pc apart. The second thick dashed line corresponds to 500 Myr after coalescence. Shading between lines correspond to the pair (light gray) and post-merger (darker gray) classes. In each panel, I show the snapshot selection for the pair (orange) and post-merger (magenta) classes. The SFRs are measured from the full simulation volume and so include contributions from both galaxies when they are separate.

The suite is divided into two components: (1) an orbit suite and (2) a mass ratio suite. The orbit suite comprises several interaction scenarios between the G2 and G3 galaxies (see Figure 3 of Moreno et al. 2019). It covers three unique spin-orbit orientations corresponding to the "e", "f" and "k" orbits from Robertson et al. (2006) (see Figure 1 of Moreno et al. 2015), three impact parameters, and three impact velocities (see Figure 3 of Moreno et al. 2019). The "e", "f" and "k" spin-orbit orientations correspond to approximately prograde, polar, and retrograde orbits, respectively. Permutation of the orbital parameters gives a total of $3 \times 3 \times 3 = 27$ unique mergers at a fixed mass ratio of $\mu \sim 2.5 : 1$. The mass suite adopts a single orbit (the fiducial "e" orbit in Moreno et al. 2019) in which each of the four galaxies interacts with every other galaxy and itself for a total of $4 \times 4 = 16$ interactions. The range in stellar mass ratios covered by the suite is $1 : 1$ to $1 : 16$. Combined, the orbit and mass components of the suite cover a broad range of encounter strengths and merging timescales.

Class definitions, merger selection, and snapshot sampling

For each merger simulation, I select snapshots from which to construct images based on a set of simple definitions of the pair and post-merger phases. For example, Figure 3.1 shows the radial separation (measured between the central SMBHs of each galaxy), relative velocity, absolute acceleration, and total star formation rate (SFR) sequences for the fiducial merger simulation from Moreno et al. (2019). The pair phase is defined to begin 100 Myr before first pericentric passage, $t_p - 100$ Myr, and to end just before coalescence, t_c – which, as in Moreno et al. (2019), I define as the last time the central black holes are more than 500 pc apart (light gray shading). I then define the post-merger phase as any time in $[t_c, t_c + 500$ Myr] (darker gray shading).

I use a temporal definition for the beginning of the pair phase primarily for convenience with the simulations. But additionally, a temporal definition may circumvent biases that can arise from selection based on projected separation or relative velocity. For example, a 300 km/s relative velocity cut (Ellison et al., 2008; Patton et al., 2013) may have missed snapshots around first and second pericentre for the merger in Figure 3.1 depending on line-of-sight.

Likewise, the definition of the post-merger phase is motivated by simplicity and the availability of temporal information in the simulations. Observability timescales

for post-merger features (shells, streams, etc.), as for the mergers overall, are sensitive to surface brightnesses, gas fractions, mass ratios, and initial orbital parameters (e.g. Lotz et al. 2008, 2010b,a; Ji et al. 2014; Nevin et al. 2019). Consequently, I select a clear-cut post-merger phase in which post-merger features are still expected to be prominent. Several mergers in the suite are either fly-bys or are not simulated to 500 Myr after coalescence. In order to preserve several mergers that coalesce but do not have 500 Myr of snapshot coverage after coalescence, I set a minimum post-coalescence criterion of 250 Myr. Consequently, the post-merger stage is defined as starting at coalescence and ending at $\max(t_{\rm last}, t_c + 500\ {\rm Myr})$ where $t_{\rm last} \geq t_c + 250$ Myr. The criteria that (a) the galaxy must coalesce and (b) the simulation has run for at least 250 Myr post-coalescence reduces the sample from 43 to 23 mergers. In particular, the interactions with the largest mass ratios and widest impact parameters are rejected on the basis of these criteria. I discuss the consequences of my class definitions on network performance in Section 3.4.4.

For each merger simulation, 30 (15) snapshots are selected which uniformly sample the pair (post-merger) phase as shown with the orange (magenta) squares in Figure 3.1. This approach provides a sampling cadence that is sparse enough that imaging in neighbouring snapshots are not overly correlated (see Appendix D) yet fine enough that a large number of images for each merger can be generated. Ten snapshots are selected from the isolated runs with uniform sampling cadence as with the pairs and post-mergers. Due to the significantly smaller number of snapshots corresponding to isolated galaxy runs, the isolated galaxy dataset is bolstered by increasing the number of orientations in which their synthetic images are generated (as described below).

3.2.2 Synthetic Observations

Synthetic images are made for isolated, pair and post-merger snapshots along four lines of sight corresponding to the arms of a tetrahedron whose vertex is coincident with the point of minimum potential. Consequently, images in the pair phase are always centred on the more massive galaxy.[5] Since the total number of pair (23×30) and post-merger (23×15) snapshots greatly outnumber isolated snapshots (4×10), producing only four camera angles for the isolated galaxies would leave the training

[5]For my purposes, this bias toward the more massive galaxy is not of any consequence. But for a dataset which must be large and general enough to handle real data, separate images should be constructed that are centred on both the primary and secondary – ideally along different lines-of-sight.

data highly imbalanced – with an order of magnitude more images for each of the pair and post-merger classes than for the isolated class.[6] As a first step in balancing the data, the number of camera angle orientations for snapshots from the isolated runs are increased by 11 inclinations and 11 position angles. Before any augmentation (see Section 3.2.3), there are $23 \times 30 \times 4 = 2\,760$ images with the pair class, $23 \times 15 \times 4 = 1\,380$ with the post-merger class, and $4 \times 10 \times (11 \times 11 + 4) = 5\,000$ with the isolated class.

Synthetic images are generated for each snapshot/orientation with various levels of realism. There are two distinct image types: (1) images originating from two-dimensional projections of stellar particles (stellar maps, Sm) and (2) from photometry generated using dust-inclusive radiative transfer (Ph). Images are produced with three different levels of realism for each type: (1) noiseless with high resolution (made by Maan H. Hani as described in the next two sections); (2) incorporating realistic (but analytically generated) noise and resolution degradation; and (3) insertion into real SDSS survey fields that may contain additional sources. I refer to these increasing levels of realism as "idealized", "semi-real" and "full-real", and they allow us to examine the importance of observational biases that are introduced level-by-level. The image types are described in detail in the sections that follow and are summarized in Table 3.2.

StellarMap (Sm)

The zeroth order stellar image that can be produced from a hydrodynamical simulation is a two-dimensional projection of the stellar particles along a given line-of-sight. This *idealized* image type has several important features: noiseless without resolution degradation; insensitivity to variations in mass-to-light ratio (M/L) from different stellar populations or dust absorption; and low computational and data management overhead. Initially, a fixed 50 kpc field of view (FOV) is adopted for each image

[6] Class imbalance – where the occurrence of one class in a dataset significantly outnumbers other classes – is a common problem in applications of deep learning to classification tasks including medical diagnosis (Grzymala-Busse et al., 2004; Mac Namee et al., 2002), fraud detection (Chan & Stolfo, 1998) and others (Radivojac et al., 2004; Cardie & Howe, 1997; Haixiang et al., 2017). An example is the natural imbalance in medical data between images of a particular diagnostic class (e.g. contains tumour) which might be 1000 times less frequent than images of another class (e.g. healthy). Unbalanced data can have significant detrimental effects on deep classifiers such as CNNs and a recent systematic study has investigated various methods which address class imbalance (Buda et al., 2017). Indeed, the method that appears to best address class imbalance is oversampling of data from the under-represented class (e.g. through augmentation) – which is the method I adopt in this chapter.

with spacial resolution of 0.097 kpc/pixel (512×512 pixels).[7] These high-resolution stellar mass projections were generated for this project by Maan H. Hani. All images (including other types) are subsequently mock-observed with the SDSS camera (0.396 arcsec/pixel) at a fixed redshift of $z = 0.046$ (the median redshift of galaxies in the DR14 MaNGA galaxy sample Bundy et al. 2015) where the scale is approximately 0.9 kpc/arcsec. Consequently, the Sm images are re-binned to a physical scale of 0.36 kpc/pixel (139×139 pixels or 55×55 arcsec^2). This still offers high resolution – particularly with respect to images that are further degraded by realistic (or real) SDSS PSFs.

Photometry (Ph)

Idealized SDSS *gri* photometric images were generated by Maan H. Hani using the Monte Carlo dust radiative transfer code, SKIRT (Baes et al., 2011; Camps & Baes, 2015). SKIRT predicts the light contribution from stellar particles and star-forming regions whilst modelling the effects of dust on the absorption, scattering, and re-emission of stellar light (note that radiation from the central engine is ignored). Stellar light from old stars (older than 10 Myr) is modelled using a Kroupa (2001) initial mass function and the associated STARBURST99 single-age SEDs (Leitherer et al., 1999). Emission from star-forming regions (stellar particles younger than 10 Myr) is represented by MAPPINGS-III SEDs (Groves et al., 2008) which include contributions from young stars and H II regions. The dust contribution is modelled assuming that the dust distribution traces the metal distribution where 30% of the metals are locked in dust particles. The multi-component dust mix of Zubko et al. (2004) is adopted which includes graphite grains, silicate grains, and polycyclic aromatic hydrocarbons (PAHs). Dust re-emission (and the associated self-absorption), which has a negligible contribution in the wavelength regime studied in this work, is ignored. The underlying radiation field is discretized by SKIRT using 10^5 photon packages per wavelength. SKIRT's output (spectral datacubes) are converged for $> 10^5$ photon packages per wavelength.

The same initial FOV and spatial resolution as for the Sm images is adopted in construction of the rest-frame SKIRT datacubes. Since broadband photometry is the final data product, a relatively coarse spectral resolution is adopted with 241 spectral elements which linearly and uniformly sample the rest-frame optical spectrum

[7] The higher the initial resolution, the greater the range of instrumental angular scales (arcsec/pixel) and redshifts (kpc/arcsec) that can be explored.

from the near-UV to near-infrared ($250 - 850$ Å). These datacubes are redshifted to $z = 0.046$ and convolved with the SDSS *gri* response functions to produce idealized photometry – accounting for stretch in the spectrum and $(1 + z)^{-5}$ reduction in specific intensities in each spectral element.[8] As with the SM images, the PH images are rebinned to the SDSS camera pixel scale. The PH images are: noiseless with high resolution; light-weighted and sensitive to variations in M/L for different stellar populations and dust; and very expensive from a computational and data management perspective when compared to SM (see Section 3.3.1). Furthermore, training networks with all three *gri* bands as input allows networks to develop sensitivity to colour.

StellarMap SemiReal (SmSR)

Ground-based imaging surveys are affected by sky surface-brightness limitations and blurring from the atmospheric PSF. These biases can be emulated using the statistics of sky brightnesses and PSF sizes measured in SDSS fields. Crucially, I match the *statistics* of sky noise levels and PSF resolution to the field properties for 1.12 million galaxies in the SDSS Legacy images (Abazajian et al., 2009) using the Simard et al. (2011) quantitative morphology catalog and ancillary data measured by the `PHOTO` pipeline (Lupton et al., 2001, 2002, 2012). I compute the means and standard deviations in the resulting sky noise and PSF resolution distribution functions. The results are tabulated in Table 3.3. I use these results to generate analytic Gaussian profiles from which sky noise and PSF resolution levels are sampled independently for each synthetic image.

The idealized SM images are not light-weighted and therefore do not offer straightforward conversion to calibrated AB flux units. To approximate the intensities of the stellar maps in realistic images, I scale each normalized stellar map by the total intensity in its corresponding idealized i-band photometry image. I choose to scale by the i-band light because it is less sensitive to variations in M/L from young stellar populations or starbursts compared to g or r. With the idealized SM images effectively "light-weighted", I sample the distribution function for the PSF and convolve. Before adding sky noise, I use the average SDSS photometric zero-point magnitude, airmass, extinction, and gain over all SDSS fields to convert the PSF-convolved images to

[8] I provide a code which performs all of these tasks (for a specified redshift) from the default rest-frame specific intensity datacubes from SKIRT (W $m^{-2}\mu m^{-1}$ $arcsec^{-2}$). The code produces output photometry in convenient AB mag/$arcsec^2$ units for each filter (Oke & Gunn, 1983) and can be found at the following url: https://github.com/cbottrell/RealSim/blob/master/SpecToSDSS_gri.py.

Table 3.2 Reference summary of image types used for training and testing of networks. Intensities in the StellarMap images are scaled to match the total surface brightnesses in the Photometry i-band images before adding realism effects.

Image Type	Shortform	Radiative Transfer	Bands	Gaussian Sky	Gaussian PSF	Real Sky	Real PSF
StellarMap	Sm	no	$i^\star$	no	no	no	no
StellarMap SemiReal	SmSR	no	$i^\star$	yes	yes	no	no
StellarMap FullReal	SmFR	no	$i^\star$	no	no	yes	yes
Photometry	Ph	yes	gri	no	no	no	no
Photometry SemiReal	PhSR	yes	gri	yes	yes	no	no
Photometry FullReal	PhFR	yes	gri	no	no	yes	yes

Table 3.3 SDSS sky and angular resolution measurements used to generate the background noise levels and convolution kernels for SemiReal images. Table quantities are computed from the ensemble of SDSS Field table values corresponding to the *full* Simard et al. (2011) galaxy sample. Columns: (1) SDSS bandpass; (2) Mean sky noise [AB mag/arcsec^2]; (3) Standard deviation in sky noise values [AB mag/arcsec^2]; (4) Mean PSF FWHM [arcsec]; (4) Standard deviation in PSF FWHM values [arcsec]. Individual SemiReal sky noise values and PSFs are drawn from normal distributions formed from these quantities.

SDSS band	$\langle\sigma_{\rm sky,Field}\rangle$	stdev($\sigma_{\rm sky,Field}$)	$\langle$ FWHM$_{\rm PSF}$ $\rangle$	stdev(FWHM$_{\rm PSF}$)
u	23.87	0.15	1.55	0.24
g	24.88	0.14	1.47	0.22
r	24.38	0.11	1.36	0.22
i	23.82	0.12	1.29	0.22
z	22.36	0.19	1.31	0.20

electron counts from which source Poisson noise can be added. I then convert back to calibrated flux units, sample the sky noise distribution and add Gaussian sky noise to the image.

Photometry SemiReal (PhSR)

The procedure for creating PhSR images in each band is the same as for creating SmSR images but without the normalization and "light-weighting" step. One feature of the current SemiReal procedure that can be remedied in the future is that the sky noise and PSF estimates are drawn independently in each band and so are not correlated as they should be. However, my results do not give me reason to suspect that this is a significant limitation of my methods.

StellarMap FullReal (SmFR)

Synthetic images with extensive observational realism are generated following the methods presented in Bottrell et al. (2017a,b). Similar to SemiReal, the FullReal procedure is designed to incorporate *statistical* observational realism (survey-realism) into the synthetic images so that real survey field statistics are matched between the simulated and observed galaxies. The main difference from the SemiReal procedure is that the synthetic images are added quasi-randomly to real survey fields in the FullReal procedure. In this approach, the insertion statistics are guided by a basis catalog of real galaxies (Simard et al., 2011). As such, the statistics of sky brightness, PSF resolution and crowding by nearby sources for real galaxies are reproduced in the synthetic images (along with any other field-dependent characteristics). The FullReal procedure is described in detail in Bottrell et al. (2017a). I provide a summary here of the procedure that is followed for every synthetic image to compliment the public release of the realism suite, RealSim[9]:

(i) A galaxy is randomly selected from the Simard et al. (2011) basis catalog. The SDSS *gri*-band fields in which that galaxy resides are extracted and converted to

[9] A public version of the RealSim observational realism suite is available the following url: https://github.com/cbottrell/RealSim for Python 3. It includes: the Simard et al. (2011) bulge+disc decomposition catalog from which it draws galaxy and field statistics; a Python 3 version of the SDSS `sqlcl.py` code which queries field information directly from the SDSS Data Archive Server; *ugriz* filter response functions from Doi et al. (2010); the Simard et al. (2011) SExtractor configuration files required for deblending of images when inserting into real SDSS fields; a Python notebook of example executions; a code for converting skirt output to SDSS images; and a sample of input images.

calibrated flux units using ancillary data queried from the SDSS DAS. A source mask is generated for the r-band field using SExtractor (Bertin & Arnouts, 1996) and deblending parameters optimized for SDSS in Simard et al. (2011) (specifically for the Patton et al. 2011 pair sample). A common injection site for each band (where applicable) is selected randomly with the restriction that the *centre* of the injected image cannot land on another object in the source mask.

(ii) The PSFs for each band corresponding to the injection site are reconstructed using the psField files and the standalone read_psf software. Each band of the synthetic image (in the SmFR case, the single "light-weighted" idealized stellar map) is converted to electrons using the ancillary data from Step (i) and convolved with the local SDSS PSF for that band. Source Poisson noise is then added.

(iii) A PSF-convolved and Poisson noise-added synthetic image (in each desired band) is finally converted back to calibrated flux units and inserted into the SDSS field at the injection site selected in Step (i). A cutout corresponding to the desired FOV (in this case, 50 kpc or approximately 55 arcsec at $z = 0.046$) is extracted around this location. This cutout now includes real sky, real PSF degradation and real additional sources in the FOV which track the statistics for galaxy image properties in the basis catalog.

Some particularities of this procedure for generating SmFR images are that (1) images are only generated in the i-band and (2) the r-band image is still used to generate the source mask as described in Step (i).

Photometry FullReal (PhFR)

Construction of PhFR images follows the same procedure as for the SmFR but for each of the SDSS *gri* bands. Because the PhFR images incorporate light-weighting from radiative transfer and the full rigour of statistical observational realism, the PhFR dataset is the benchmark for how a given network would be expected to perform on realistic data and is often discussed as such in Sections 3.3 and 3.4 (i.e. the PhFR images are the closest representation of observable galaxies in the suite of image types and are hence used as the ultimate test set of "real life" performance). As with the Ph and PhSR datasets, the PhFR has three channels of input corresponding to the three bands in which I produce photometry.

Figure 3.2 shows a recent post-merger for each of the 6 image types and demonstrates the impact of each level of realism. The upper panels show images originating from stellar maps and lower panels show images originating from radiative transfer. In the idealized Sm and Ph images (left hand column), morphological features post-merger are visually prominent including shells, streams, and tidal tails that have not yet decayed from the pair phase. In the Ph image, a dust lane obscures light emanating from the nucleus – giving it an asymmetric appearance with respect to the Sm image. Additionally, the Ph image has bright knots associated with the low M/L of young stellar populations whereas the Sm image is insensitive to these features. The SemiReal images in the middle panels show the results of adding realistic SDSS noise and resolution degradation to the images. Many of the features that made this object easily identifiable as a post-merger in the idealized images are drowned by the sky noise and PSF blurring. Features of the post-merger remain in the SemiReal images but are more subtle than in idealized images. The right panels show FullReal images for the post-merger. In addition to real skies and degradation by real PSFs, these images incorporate contamination by nearby sources. The lower right panel shows particularly striking chance projection with an interloping disc galaxy. Taken together, these images nicely encapsulate the rationale of this work.

3.2.3 Image normalization and augmentation

Normalizations and augmentation (oversampling) are applied to images in each database. I generate augmented images by applying zoom, rotation, and small translational transformations to the set of original images in order to (1) reduce class imbalance and (2) achieve rotational invariance in the models. The augmentations are performed with the using the ImageDataGenerator class from the Keras Python API (Chollet et al., 2015). All images of a particular class are augmented N times until the total number of images exceeds 10,000. Consequently, the final number of images in each class are $(N_{\rm Iso}, N_{\rm Pair}, N_{\rm Post}) = (10000, 11040, 11040)$ after augmentation for each image type.

The augmented images (starting in linear intensities) are normalized following a standardized algorithm which is applied to all image types:

- If the image contains sky noise (SemiReal and FullReal types), then the image is subtracted by its median intensity.

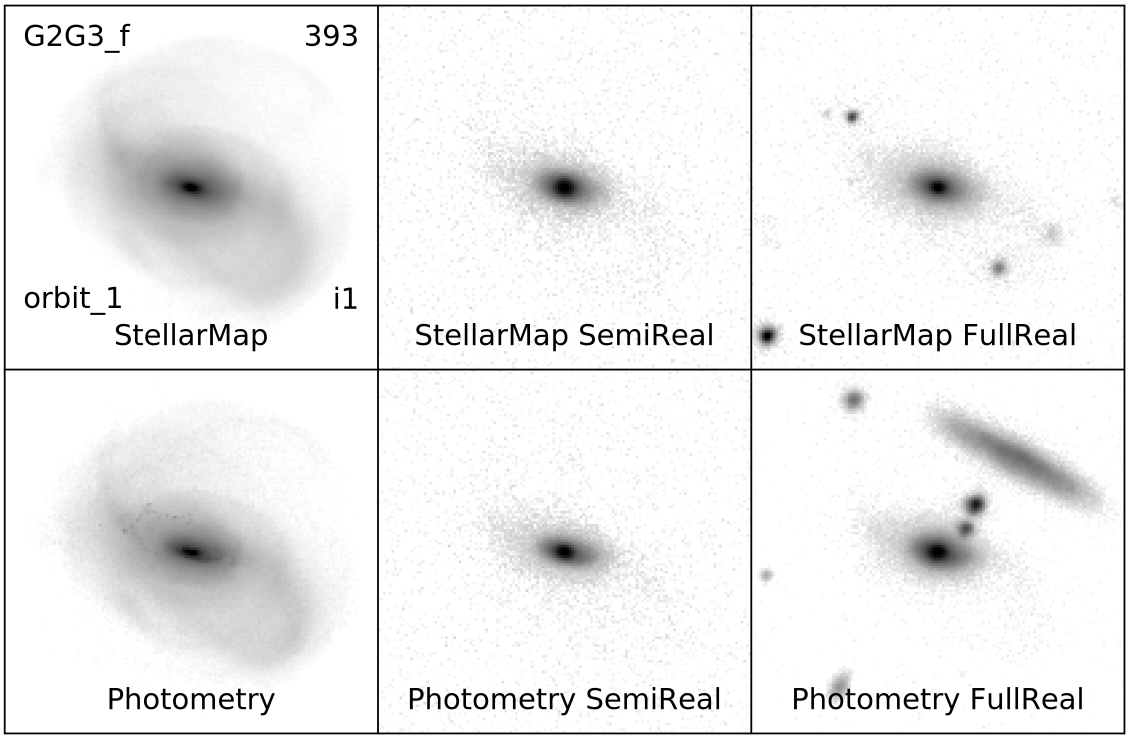

Figure 3.2 Visualization of a single post-merger galaxy realized with every image type. The post-merger image taken from orbit 1 of the G2G3 "f" orbit suite at snapshot 393 (post-coalescence). All images show $i-$band intensities. Note the potential for misclassification in the PhFR realization of this post-merger due to a chance projection with a field galaxy.

- Take the logarithm of the sky-subtracted image. All values less than -7 are converted to NaNs.
- The median of the full image, $a_{\rm min}$, and $99^{\rm th}$ percentile for the central 20×20 pixels, $a_{\rm max}$, are computed.
- All values below $a_{\rm min}$ (and NaNs) are set to $a_{\rm min}$ and values greater than $a_{\rm max}$ are set to $a_{\rm max}$.
- The clipped logarithmic image is subtracted by $a_{\rm min}$ and normalized by $a_{\rm max} - a_{\rm min}$.

The results are images with logarithmic-scale intensities in the range of 0 to 1 in which each image is scaled to maximize contrast for the central target in the image. This normalization procedure avoids the problem of reduced contrast in FullReal images which contain bright stars or other contaminating effects (such as would be produced by a more conventional standard scalar).

3.2.4 Neural network architecture

I use my synthetic images to train CNNs which classify galaxies as isolated, pairs, or post-mergers. CNNs are a class of deep learning model that are particularly useful for data which exhibit topological structure such as images (Fukushima, 1980; LeCun et al., 1989, 1995, 1998; Krizhevsky et al., 2012; Lecun et al., 2015). There is enormous flexibility in CNN architectures in terms of depth (number of layers), layer properties (kernel sizes, etc.) and layer structures (e.g. residual blocks, He et al. 2015). Given the successes of previous works using a particular (and relatively simple) CNN architecture for galaxy morphology classifications (Dieleman et al., 2015; Huertas-Company et al., 2015; Domínguez Sánchez et al., 2018, 2019; Huertas-Company et al., 2019), I adopt a similar (but not identical) CNN architecture for predicting galaxy merger stage. This architecture is summarized in Table 3.4. The output of the network for a given image is a class probability distribution function, $(P_{\rm Iso}, P_{\rm Pair}, P_{\rm Post})$. For my analyses and comparison with the known target classes, I adopt the class with the highest probability density.

A $(70, 15, 15)\%$ split is used for *training*, *validation*, and *test* images. The networks are optimized on the training images and corresponding known target classes. The overall performance of a network on the validation images, $N({\rm correct})/N_{\rm tot}$, is

Table 3.4 Convolutional neural network architecture for the full-colour model which accepts 3-channels of input comprising *gri* images. Convolution kernel sizes (Conv2D), max-pooling windows (MaxPool), dropout rates (Dropout), output shapes and total number of trainable parameters for each of the layers are indicated. The convolution layers use Rectified Linear Unit (ReLU) non-linear activation functions and have a (1,1) stride. The output of the 4th convolution layer is flattened to a one-dimensional feature array and passed to the fully connected (dense) component. Dense layers also use ReLU activation functions. The output layer uses a softmax activation function.

Layer (type)	Output shape	# Parameters
Input Layer	$(139, 139, 3)$	0
Conv2D-1 (6×6)	$(139, 139, 32)$	3 488
MaxPool-1 (2×2)	$(69, 69, 32)$	0
Dropout-1 (0.25)	$(69, 69, 32)$	0
Conv2D-2 (5×5)	$(69, 69, 64)$	51 264
MaxPool-2 (2×2)	$(34, 34, 64)$	0
Dropout-2 (0.25)	$(34, 34, 64)$	0
Conv2D-3 (2×2)	$(34, 34, 128)$	32 896
MaxPool-3 (2×2)	$(17, 17, 128)$	0
Dropout-3 (0.25)	$(17, 17, 17)$	0
Conv2D-4 (3×3)	$(17, 17, 128)$	147 584
Dropout-4 (0.25)	$(17, 17, 17)$	0
Flatten	36 992	0
Dense-1	(512)	18 940 416
DropFC-1 (0.25)	(512)	0
Dense-2	(128)	65 664
DropFC-2 (0.25)	(128)	0
Output Layer	(3)	387
Total # parameters		19 241 699

evaluated after each training epoch. While this step does not, strictly-speaking, affect optimization, it is used to determine an appropriate time to stop training and consequently prevent overfitting to the training images. In contrast, networks are never exposed to test images during training. For tests in which networks trained on a particular image type (e.g. PhFR) are tested on images of a different type (e.g. SmSR – which may *all* technically be considered distinct data), I find that there is no difference in network performance whether I test only on the corresponding test images or all images (including training and validation images). My results show the latter for such tests.

3.3 Experiments

Each synthetic image dataset is used to train neural networks which are then applied to test images of every type. This "handshake" of training/testing experiments includes cases where training and test data are of the same type. I generate 10 networks for each image type by splitting the data into training, validation and test images using 10 unique random states. This "bootstrapping" of the data allows me to characterize the random error associated with the selection of a particular training set and to statistically merge my test results.

3.3.1 Model and image handshake

The main experiment comprises a handshake between networks trained using the 6 unique image types (see Table 3.2) which are each tested on the 6 image types – resulting in 6×6 tests. Figure 3.3 shows a qualitative schematic of this experiment. The results for a single test can be characterized using a confusion matrix (e.g. Figure 3.4, description in Section 3.3.1). The full 6×6 matrix of confusion matrices can be found in the Appendix E where the reader can zoom in on every individual test. In this section, I focus on the matrices corresponding to specific tests from the main handshake which address my core questions.

Case study: training and testing with ideal Photometry

Figure 3.4 shows a single confusion matrix corresponding to the models trained on Ph and tested on Ph (as described in Section 3.2.2 and summarized in row 4 of Table 3.2). I use this as a simple case study to orient the reader to the general features of my

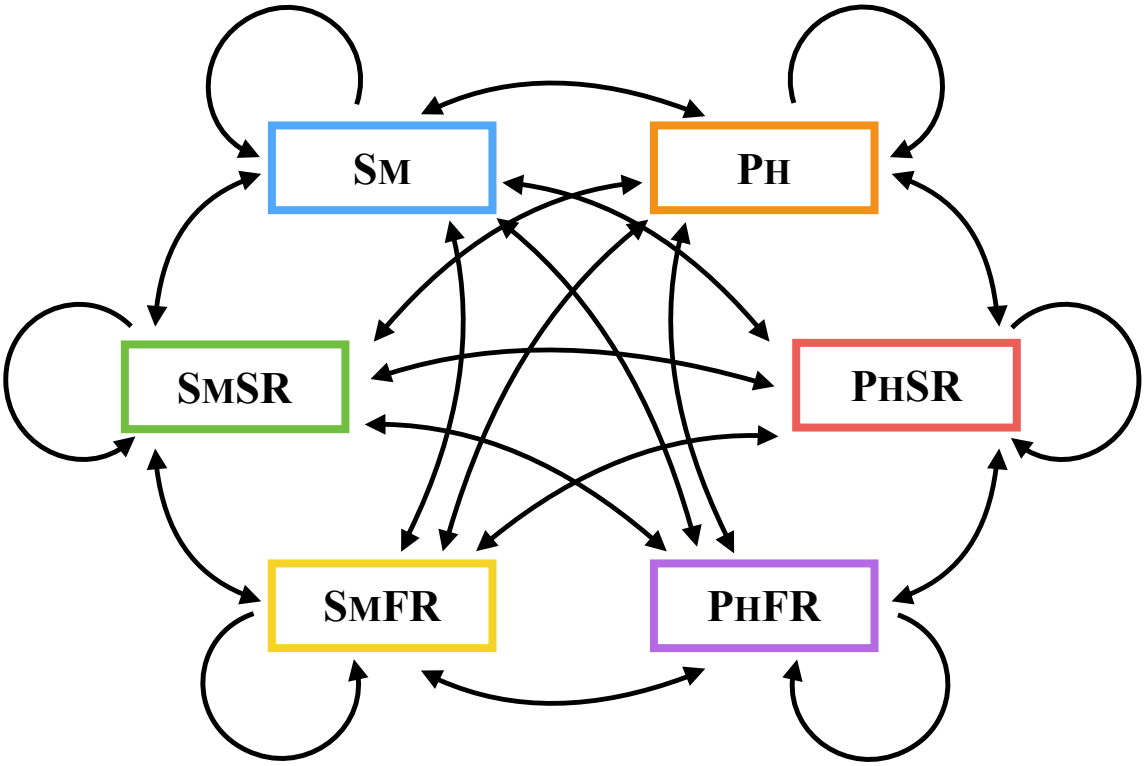

Figure 3.3 Schematic of the main handshake experiment. Networks are trained with images of each type. Each trained network is tested to images of every other type. Additionally, each network is tested on images of the same type but that the networks never see during training (outer looping arrows).

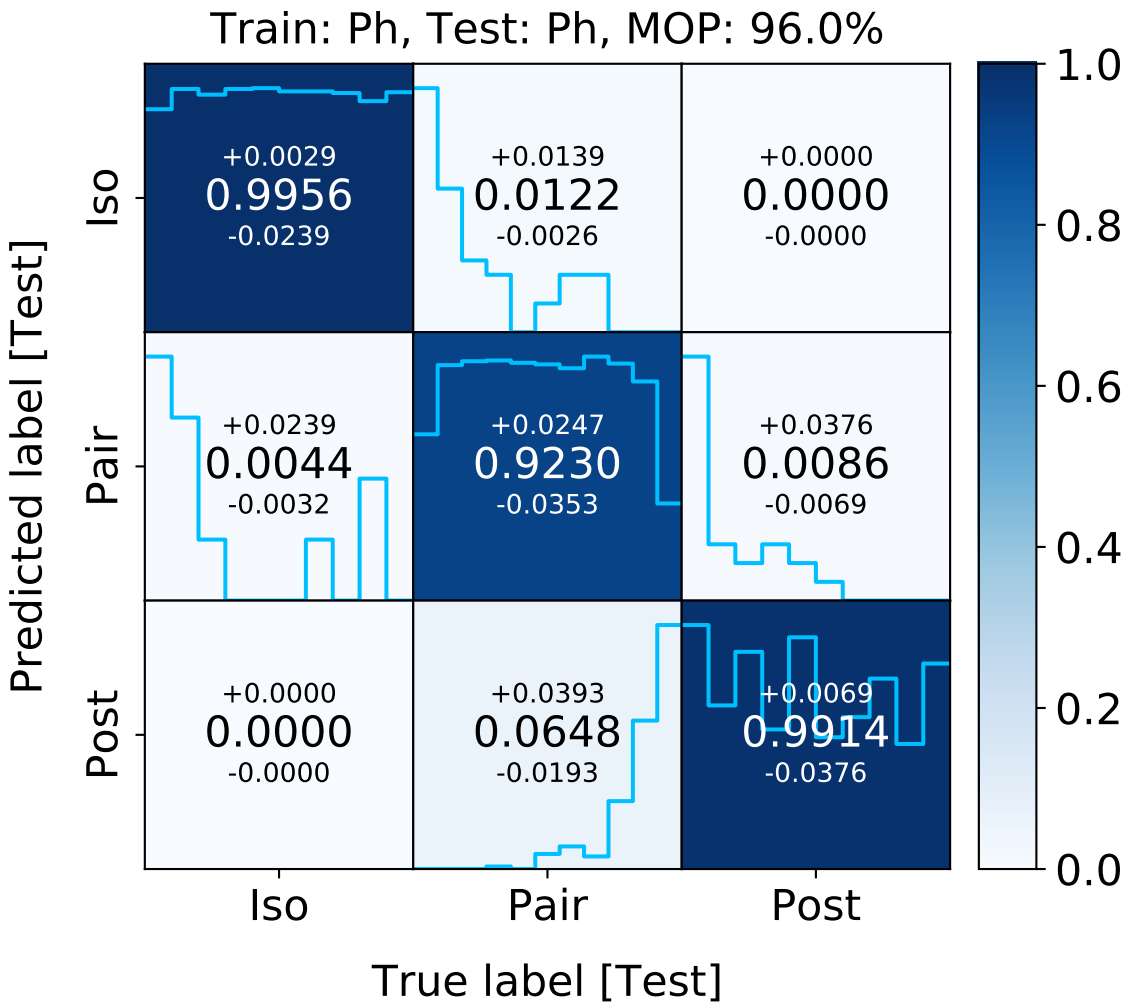

Figure 3.4 Confusion matrix for merger stage classifications using the Ph networks and corresponding test images. The number in each matrix element quantifies the median predicted fraction of images and 16^{th} and 84^{th} percentile offsets computed by bootstrapping the results of 10 unique realizations of training, validation, and test data. Inset in each matrix element is a light blue bar plot showing the median empirical distribution function of relative merger class timescales (see Equation 3.1) for the images which fall in that matrix element in each of the 10 bootstraps. These timing histograms in each element are normalized by their respective maximum values for visibility in cases where the total number of images in a particular matrix element is small. Networks trained on Ph training data have a median overall performance (MOP) of 96.0% when applied to Ph test data.

analysis for subsequent experiments. Each column of the confusion matrix shows the normalized distribution of predicted labels (y-axis) for all test images with a particular truth label (x-axis). A perfect classification network would therefore produce an identity matrix where all of the power is in the diagonal elements. Off-diagonal elements correspond to misclassifications denoted by a combination of predicted and true labels. Values shown in each matrix element are the median and 16th and 84th percentile range computed from the 10 bootstrap realizations of training, validation, and test data.

Figure 3.4 demonstrates that exceptional performance is achievable under ideal but generally unrealistic conditions: (1) noiseless images; (2) high spatial resolution; (3) no contamination by projection effects or other objects in the field of view; and lastly (4) a training set that is (by construction) generalizable to the test images. The ideal Ph networks have a median overall performance (MOP) of 96.0%. Nonetheless, 7.7% of pairs are misclassified by the networks – with 6.5% being misclassified as post-mergers and 1.2% misclassified as isolated galaxies. Similarly, a small number of isolated and post-merger images are misclassified as pairs.[10]

My confusion matrix includes an additional dimension which allows deeper investigation of network systematics and misclassifications. Histograms are inset in each matrix element which show the distributions of *relative merger class timescales*:

$$t_{\rm rel,iso} = \frac{t - t_1}{t_{10} - t_1}, \quad t_{\rm rel,pair} = \frac{t - t_p}{t_c - t_p}, \quad t_{\rm rel,post} = \frac{t - t_c}{t_{\rm last} - t_c} \tag{3.1}$$

Here, t is the simulation time-stamp associated with a particular snapshot in any simulation run. For isolated runs, t_1 and t_{10} are the timestamps for first and tenth of the 10 snapshots selected. For merger simulations, t_p is the time of first pericentric passage, t_c is the coalescence time, and $t_{\rm last}$ is the timestamp of the last snapshot selected for a given run. $t_{\rm last}$ is therefore a number between $t_c + [250, 500]$ Myr as per my class definitions and merger selection criteria. These normalizations allow me to place the timestamp of each snapshot (from each simulation) on a *relative* timeline corresponding to its target class. An image from the isolated or post-merger classes has a $t_{\rm rel,iso}$ or $t_{\rm rel,post}$, respectively, that is between 0 and 1 by definition – which I divide into 10 bins each. An image from the pair phase has $t_{\rm rel,pair}$ between -0.1

[10]However, a subtle but notable feature of the networks is that the 55 arcsec fields of view (50 kpc at z=0.046) of the images often do not contain both galaxies for pairs. Nonetheless, networks distinguish pairs from isolated and post-merger images with great accuracy in these cases.

and 1 because (a) I start the pair phase 100 Myr before first pericentre and (b) the shortest $t_c - t_p$ is roughly 1 Gyr. Accordingly, each pair timing histogram has 11 bins and starts at 100 Myr before first pericentre and ends at coalescence. With these definitions, a uniform timing distribution in any matrix element would indicate that there is no temporal preference for the images assigned to that element. For visibility, the timing histograms in each element are normalized by their maximum values rather than the total number of images with the corresponding truth label.

Despite the good overall performance of this network and small fraction of misclassifications, the timing histograms reveal temporal preferences for misclassifying certain classes. True pairs that are misclassified as isolated are predominantly in the very early pair phase – with the largest fraction in the pre- first pericentre bin (middle-top panel of Figure 3.4). In contrast, true pairs that are misclassified as post-mergers are preferentially near coalescence (middle-bottom panel of Figure 3.4). Consequently, the distribution of $t_{\rm rel,pair}$ for the correctly classified pairs is truncated in the first and final bins. The choppy $t_{\rm rel,post}$ distribution for the correctly classified post-mergers is due to the chance temporal resonance of snapshots selected for each merger (i.e. the snapshots naturally fall within certain bins of $t_{\rm rel,post}$ more than other bins). Coarser binning reveals an essentially uniform timing distribution. Lastly, post-mergers that are misclassified as pairs show a strong preference towards snapshots shortly after coalescence (right-middle panel of Figure 3.4).

The timing histograms in Figure 3.4 show that images which correspond to snapshots at the temporal interface between two neighbouring classes are the most challenging for the network to accurately classify. The subtlety is that these misclassifications are not completely spurious but rather follow intuitive temporal distribution functions – which is actually a validation that the network is behaving as it should. For example, the most common mis-classification for pairs early in their interaction is "isolated". Likewise, the most common misclassification of late stage pairs, shortly before coalescence, is "post-merger". Conversely, it is rare for early pairs to be misclassified as post-mergers, or for late stage pairs to be classified as isolated. The misclassifications arise because the features of images on either side of a particular class boundary are genuinely similar. Indeed, the timing distributions for correctly and incorrectly classified pair and post-merger targets are qualitatively similar in my other tests (except where additional systematics due to strongly contrasting network/data types dominate).

However, the timing distribution of isolated galaxies that are misclassified as pairs

is less intuitive. First, the timing distribution of *correctly* classified isolated galaxies is largely uniform. This result is important. Temporarily discounting secular changes to morphology (such as the emergence of bars and spiral arms), the main changes to these galaxies are their star-formation rates (SFRs) – which decay exponentially with time. The uniform timing distribution for correctly classified isolated galaxies indicates that *most* isolated galaxies are being correctly classified despite significant changes in SFR. However, the timing distribution of isolated galaxies that are misclassified as pairs suggests that early (and incidentally high-SFR) snapshots from the isolated runs are favoured.

Figure 3.5 shows 16 randomly selected images of isolated galaxies that are correctly classified as isolated (left panel) and incorrectly classified as pairs (right panel). The comparison reveals that, occasionally, misclassified isolated galaxies are not easily visually distinguished from the correctly classified isolated targets (e.g. first row, second column of the right panel). However, the right panel of Figure 3.5 shows that the majority of misclassified isolated galaxies have not yet dynamically relaxed from the initial conditions of the simulations and consequently have non- steady-state morphologies. Many have unusually bright spiral arms or rings of star-formation which may confuse a network which would desirably exploit morphological features such as tidal tails and shells to identify pairs. In addition, galaxies in the early stages of the merger simulations are similarly unrelaxed and should increase confusion between the pair and isolated classes. Again, there is a subtle importance to these results. The mischaracterization of these dynamically unrelaxed isolated galaxies as pairs *confirms* that the network is exploiting desirable morphological features to make pair classifications. One may also visually note that (unlike morphology) high central surface brightnesses (such as those induced by a starburst) do not necessarily translate to a high pair probability. In Section 3.3.2, I perform additional tests outside the main handshake which allow me to examine the temporal misclassification of isolated galaxies more deeply.

Is radiative transfer necessary?

It is particularly useful to know whether Sm images are an adequate replacement for Ph. Radiative transfer makes photometric images computationally expensive to produce in large quantities and the data products from radiative transfer can be very large depending on the desired spectral resolution. For reference, in my dataset a

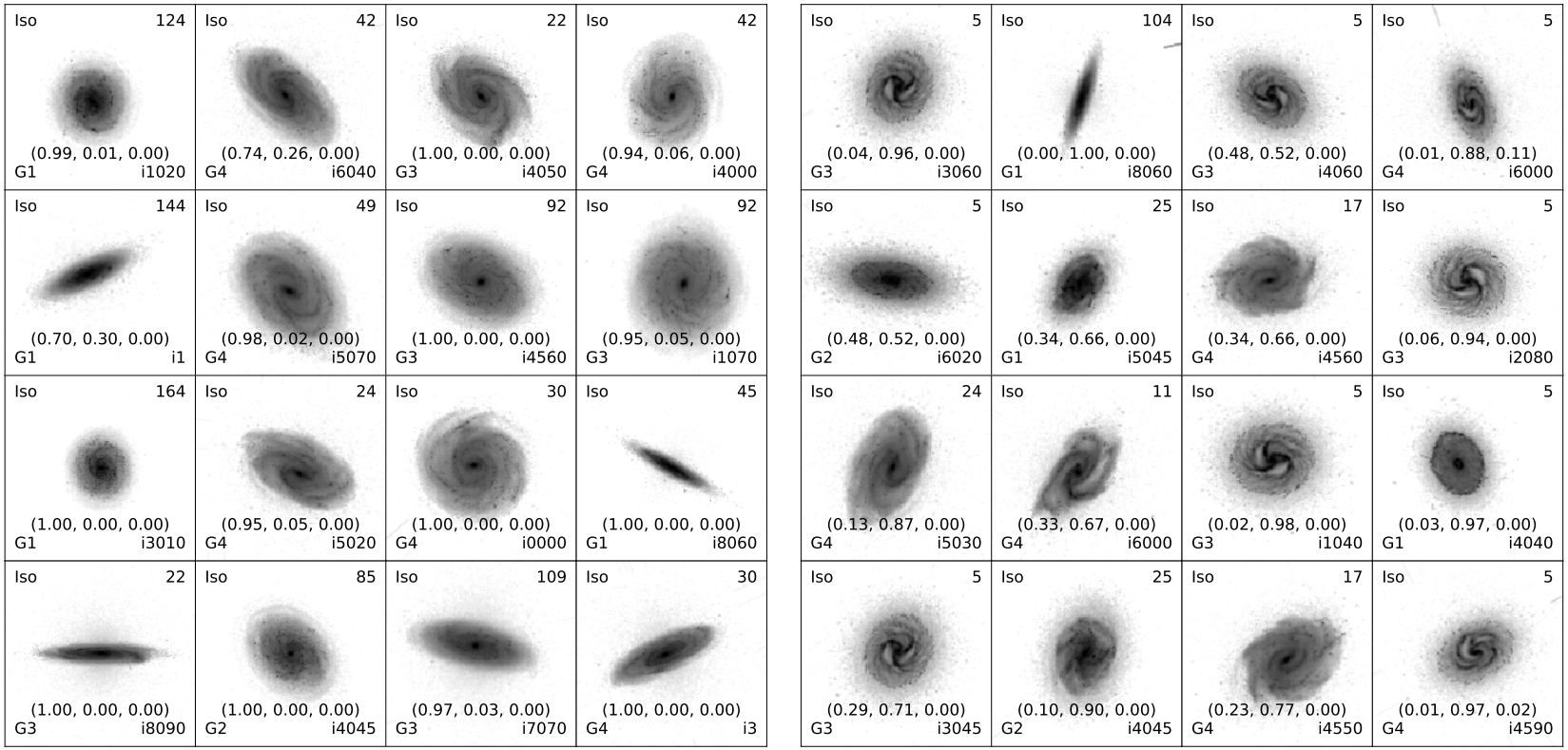

Figure 3.5 Randomly selected $r-$band idealized PH images of isolated galaxies: (left) correctly classified as isolated; (right) incorrectly classified as pairs by the networks from Figure 3.4. The unique identifier for each image is composed of the four labels in the corners of each image. The upper left shows the simulation type (all Iso in this case), the upper right is the snapshot number, the lower left is the galaxy ID, and the bottom right is the camera angle ID. The normalized probability distribution for each image is given as the tuple of ($P_{\rm Iso}$, $P_{\rm Pair}$, $P_{\rm Post}$). These single-band images show that there are notable differences in the surface-brightness distributions between correctly classified isolated galaxies and those misclassified as pairs. Many of the misclassified isolated galaxies have not yet relaxed from the initial conditions of the Iso simulation runs and exhibit non- steady-state morphologies. Streaks at the edges of some images are artifacts of border-handling when images are rotated/shifted during augmentation.

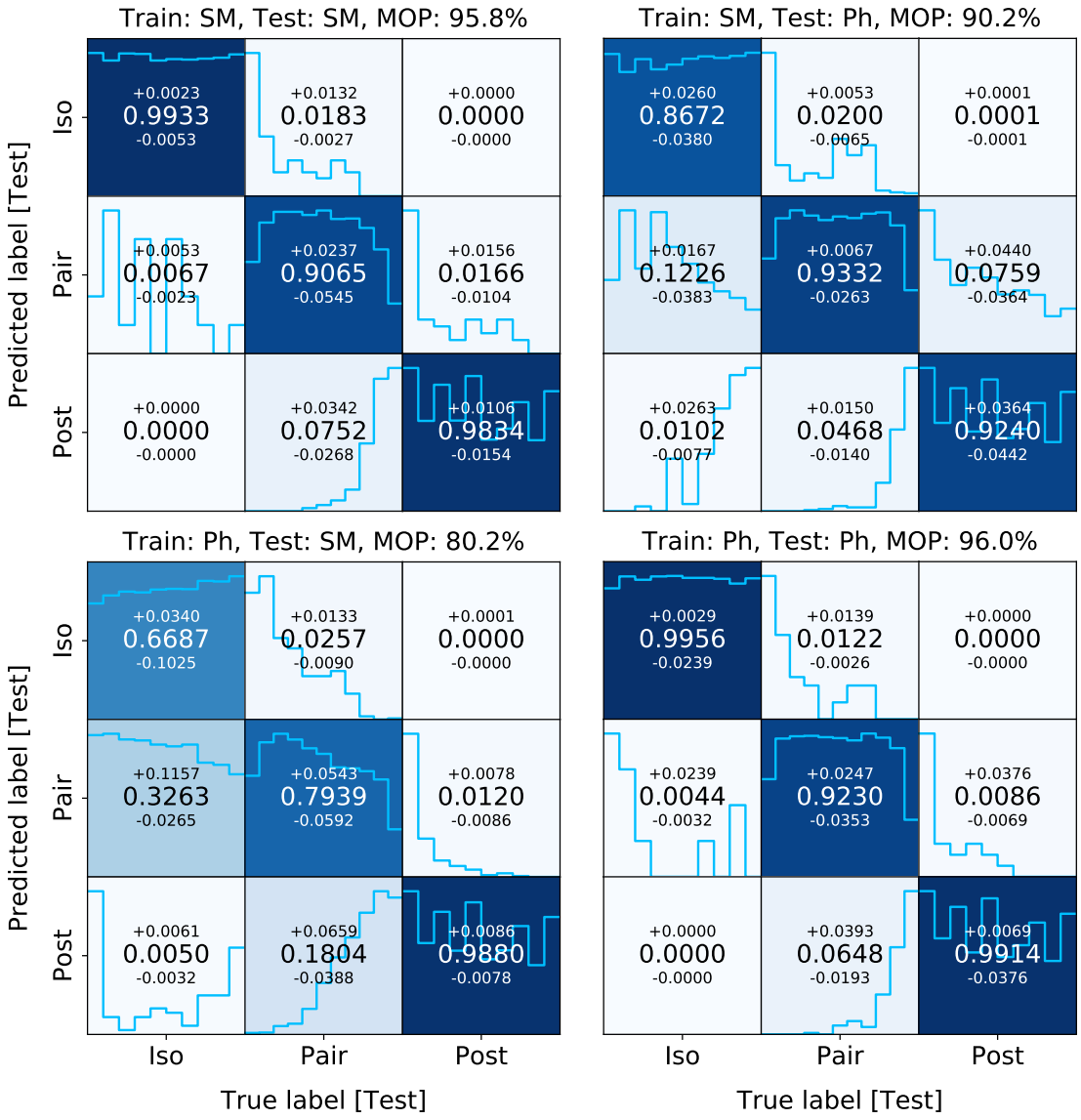

Figure 3.6 The importance of radiative transfer. Confusion matrices for merger stage classifications with Sm and Ph images/models. Models for matrices in the top row are trained on Sm images and matrices in the bottom row are trained on Ph. Models for matrices in the left column are tested on Sm images. Models for matrices in the left column are tested on Ph images. The columns of each matrix show the normalized distribution of predicted labels for each true label. The median and 16^{th} and 84^{th} percentile offsets are computed by bootstrapping the results of 10 unique realizations of training, validation and test data.

single datacube with (512×512) spatial elements, a modest 241 spectral elements and 32-bit floating precision is 252.7 MB. Constructing a sufficiently large training set for application to cosmological volumes ($\sim 10^6$ images) with these standards is a data management expense on the order of hundreds of Terabytes for the raw data products alone. Therefore, if a neural network performs equally well when trained on Sm images (1.05 MB/image for the same spatial resolution) as with images produced using radiative transfer (i.e. generation of Ph), it would not only save a significant step in the image generation pipeline, but also be a massive computational saving.

Figure 3.6 shows the confusion matrices corresponding to the test of the importance (or not) of including radiative transfer and the generation of multi-band photometric images. The lower right panel shows the same high-performance result as in Figure 3.4 where the idealized Ph-trained networks are applied to Ph test data. The upper left panel shows that nearly equally high performance is achieved by the Sm networks on Sm test data. The off-diagonal panels show the results of testing of these networks on data from the other type. The lower left panel shows that networks trained on Ph and tested on Sm images have significantly lower performance (median performance of 80.2%) than when either Sm or Ph networks are tested on data of their own respective type. In contrast, the upper right panel shows that networks that are trained on Sm images and tested on Ph still have excellent performance (median performance of 90.2%). These results are intuitive when one reflects on the differences between a single-band photometric image and a map of stellar mass. With respect to the Sm images, Ph images include higher-order information from which the network can draw (such as locally varying mass-to-light ratios due to the ages and metallicities of stellar populations and dust). If these higher-order features correlate with the target classifications then the Ph network may suffer from an unconventional form of overfitting with respect to the corresponding Sm images – because these higher-order features are absent in the Sm images.

In contrast, the morphological disturbances that are exploited by the Sm network when training on Sm images will *always* be present in the Ph images. They will simply underlay any higher-order Ph features. Consequently, the Sm model tests with higher performance on Ph images than vice-versa because the Sm network is *guaranteed* to focus on lower-order features and thus is generalizable to Ph. Later, in Section 3.4.2, I argue that the disparity between Ph and Sm networks/data arises due to the limitations of the training data and predict it would disappear with a galaxy population that is more diverse in stellar populations, colours and gas fractions.

The results of this section demonstrate that radiative transfer provides a network with more exploitable features than are available in Sm images. These include higher-order features of the surface brightness profiles and the colour information that can be made accessible by producing images in multiple bandpasses. However, I also show that idealized Sm networks are, nonetheless, highly effective at handling idealized Ph images. This means that, at least in the idealized case, one can avoid the (potentially enormous) computational and data management expenses of radiative transfer for large datasets by using Sm-based images for a modest trade-off in performance.

Is observational realism necessary?

We are ultimately only interested in networks that will perform well on realistic images. In the last two sections, I have shown that the networks trained on idealized Sm and Ph images perform very well on their respective selves and reasonably well on each other. In this section, I address the question of whether networks trained with idealized images can accurately classify images with realistic noise, resolution degradation and contamination by nearby objects in the images' fields of view. The benchmark for assessing how well any of the networks will perform on real data is the Photometry FullReal dataset – the best representation of real data amongst my datasets. The tests in this section are designed to show whether it is sufficient to construct idealized StellarMap or Photometry synthetic images as training data for networks that can be applied to real data. If satisfactory performance can be achieved on realistic test images with networks trained on idealized Sm and/or Ph data, then adding additional layers of realism to training images would be unnecessary.

Figure 3.7 shows the results of applying the Ph (left), PhSR (centre) and PhFR (right) trained networks to PhFR test data. In this sub-section, I will focus on the left and right panels – which demonstrate the importance of realism, returning to the central panel of Figure 3.7 in the following sub-section. The left panel shows that Ph networks have very poor performance when tested on realistic images. Similarly poor results are obtained using the idealized Sm networks (see corresponding panel in Figure E.1 in Appendix E). Both idealized Ph and Sm networks systematically classify targets in the FullReal images as pairs. The histograms inset in the elements of each matrix show that there are no particular temporal preferences for post-merger or isolated targets that are misclassified as pairs by these networks.

On the other hand, the right panel of Figure 3.7 shows that networks that are

trained using PhFR images perform very well on other PhFR survey-realistic images – despite the full statistical rigour of noise, resolution and contamination effects. Indeed, these results suggest that it is *only* because this network was exposed to these biases in training that it is capable of handling other realistic images. I investigate this hypothesis more closely in the next section.

Is the *level* of realism important?

In the previous sub-section, I showed that networks that are trained on either Ph or Sm idealized images perform poorly when tested on realistic images, whereas the PhFR network performed very well. To interpret these results, I now focus on the following question: what ingredients of the FullReal images are key to the success of the FullReal network? Are realistic skies and resolution sufficient criteria? Or are realistic additional sources necessary too? I answer these questions using the networks trained on SemiReal images which are tested on FullReal images. Recall that the sky noise and spatial resolution in the SemiReal images are statistically the same as for the FullReal images by construction (see Section 3.2.2). Consequently, the only difference between these two datasets is that the FullReal images can contain other objects in the field of view which may confuse a network that is not used to seeing additional sources.

The central panel of Figure 3.7 shows the results of applying PhSR networks to PhFR test sets. Again, the results are qualitatively similar when I apply the SmSR networks to the PhFR or SmFR images (see corresponding panels in Figure E.1 of Appendix E). These tests reveal that the SemiReal networks (whether originally deriving from Sm or Ph images) systematically classify targets as pairs in the FullReal images – as was the case for the idealized Ph and Sm networks. The consequently poor overall performance, particularly when compared to the successes seen when training and testing using FullReal images (right panel of Figure 3.7), demonstrate that the *level* of realism is crucial to network performance with realistic images. Specifically, without exposure to projection effects and additional sources of contamination during training, the SemiReal networks preferentially associate secondary sources in the images as companions to the target. In contrast, networks trained on images that include contaminating effects beyond sky and resolution degradation must learn ways to separate false positives and true positives with respect to the pair class.

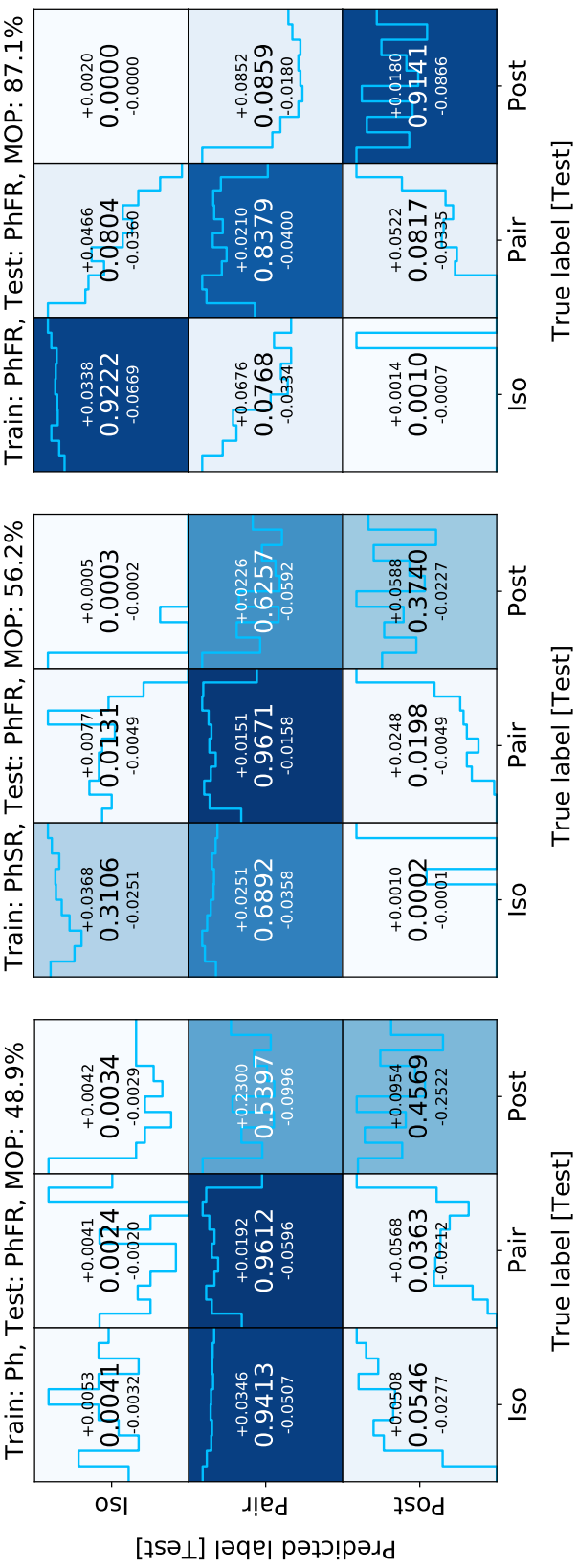

Figure 3.7 The importance of realism. PH (left), PHSR (centre) and PHFR (right) networks are applied to the PHFR test data. Both the idealized PH and PHSR networks systematically misclassify non-pair targets as pairs. In contrast, networks that are exposed to training data with full realism (real noise, resolution *and* crowding) do not appear to be affected by this systematic and an overall median performance of 87.1% is achieved with the PHFR networks on PHFR test data.

Figure 3.8 examines some important details in the relationship between the level of realism in training data and network performance on realistic test images. The upper panel of Figure 3.8 shows that SemiReal networks perform very well on SemiReal test data – while we know from the middle panel of Figure 3.7 that this performance in uncontaminated FOVs does not translate to the FullReal images. However, the reverse of that test (training on FullReal images and testing on SemiReal images) would show whether the contaminants in the FullReal training data negatively affect network performance on images which do not contain contaminants. The lower panel of Figure 3.8 shows the results of applying FullReal networks to the PhSR test images. This test confirms that the networks trained on FullReal images have no trouble testing on images which have similar skies and resolution but which do not contain contaminants. Indeed, statistically fewer isolated (-4%) and post-merger (-2%) targets in the SemiReal test images are mis-classified as pairs by the PhFR network compared to the right panel of Figure 3.7. It is important to note that, until now, we have not seen a network that is trained on one image type and performs equally well (or better) on another image type. Given that the network architecture is the same for every model (with the exception of the number of input channels in particular cases), Figure 3.8 is also a crucial validation that the networks are not predestined to overfit to their own training data due to some property of the model architecture.

The results of this section show that training images with realistic noise and resolution are important but (on their own) insufficient criteria for achieving high merger classification performance in realistic images. Networks will only learn to discriminate between true and false pairs if they have been exposed to realistic fields of view in their training data (FullReal images). As such, training data must include realistic noise and resolution *as well as* contamination by additional sources.

Is realism more important than radiative transfer?

In the previous section, I showed that the highest performance on realistic test data is obtained when training with PhFR images (see Figure 3.7). In addition, in Section 3.3.1 and Figure 3.6 I showed that networks trained on Sm images performed well when tested on Ph data. So I return to the guiding question of Section 3.3.1: *Can we get away with using* StellarMap*-based images in lieu of radiative transfer?* In other words, is the realism more important than whether the realistic image are

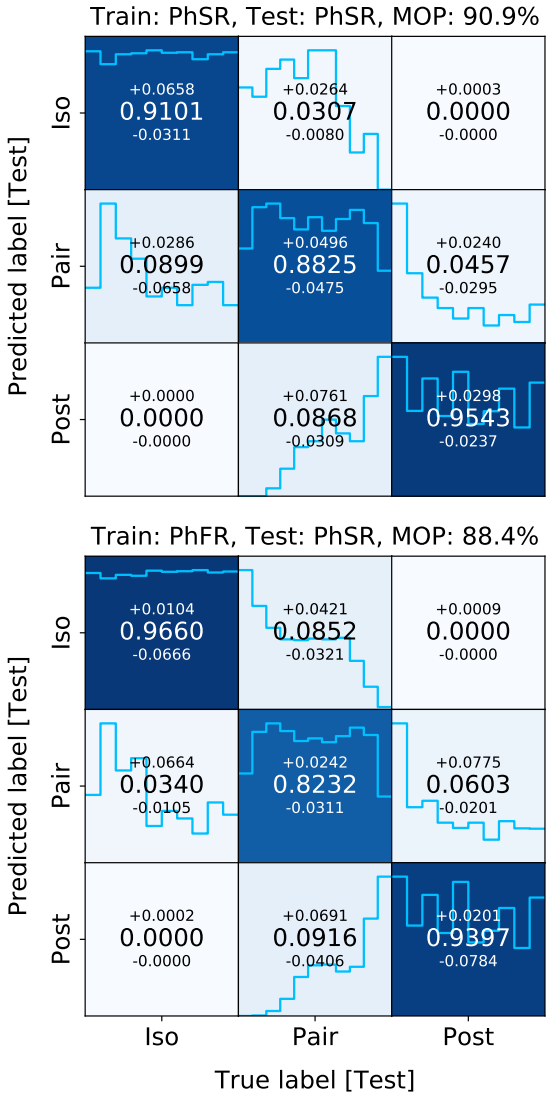

Figure 3.8 The importance of the *level* of realism. The upper panel shows that PhSR networks (trained on images with realistic skies and resolution but no field objects) handle other PhSR images with very good performance. In contrast, the middle panel from Figure 3.7 showed that SemiReal networks perform very poorly when applied to fully realistic images and systematically classify targets as pairs – regardless of their true class. The lower panel here shows that the reverse of this test (training on FullReal images and testing on SemiReal images) produces even better results than when the PhFR networks to PhFR test data. This test shows that networks that are trained on images which include contaminated FOVs have no trouble handling images which have similar noise and resolution but do not contain contaminants.

originally derived from a StellarMap or from Photometry?

Figure 3.9 offers a compressed view of every test in the main handshake shown in Figure E.1 of Appendix E. Figure 3.9 shows the median overall performances of each network applied to each set of test data. The overall performance is computed as the number of images in the diagonal elements of a confusion matrix relative to the total number of images. Each panel shows the results of networks trained using each type of training data (labels along x-axis) and tested on a particular type of test data (indicated in the tan box). Coloured bars show the median and sample standard deviation test performance of each network type for the 10 networks trained using different random samplings of the training images. The dashed black line denotes a random performance of 1/3 for a three-class model. Since PhFR data are closest to what would be observed with a real instrument, the lower right panel is the focus of this section. As I showed in Figure 3.7, the Sm and Ph networks do only slightly better than random when applied to the PhFR test images because of the lack of realism. Similarly, I showed in Section 3.3.1 that the SemiReal networks do only marginally better than models with no realism because they are not exposed to projection effects in training. In contrast, *both* FullReal networks (SmFR and PhFR) perform well on PhFR images. The SmFR and PhFR networks have median overall performances of 79.6% and 87.1%, respectively, when applied to PhFR images.

In contrast, networks trained on either idealized or SemiReal images never exceed 60% performance when handling the more realistic PhFR images. This is true whether the training images are derived from photometry or stellar maps. These results show that the level of realism is more important than radiative transfer and that one can achieve strong performance with Sm-based images as long as they are fully realistic. Although there is a big difference between a network that can achieve $\sim 90\%$ performance compared to one that achieves $\sim 80\%$, the difference in performance may be an acceptable trade-off for being able to side-step radiative transfer and its associated computational and data management expenses – particularly on the scale of the current state-of-the-art cosmological simulations.

3.3.2 Single-channel experiments

I supplement the main handshake experiment with a series of additional single-band tests with Ph-based networks. These tests are designed to determine the importance of colour and bandpass to network performance. In particular, we are interested in

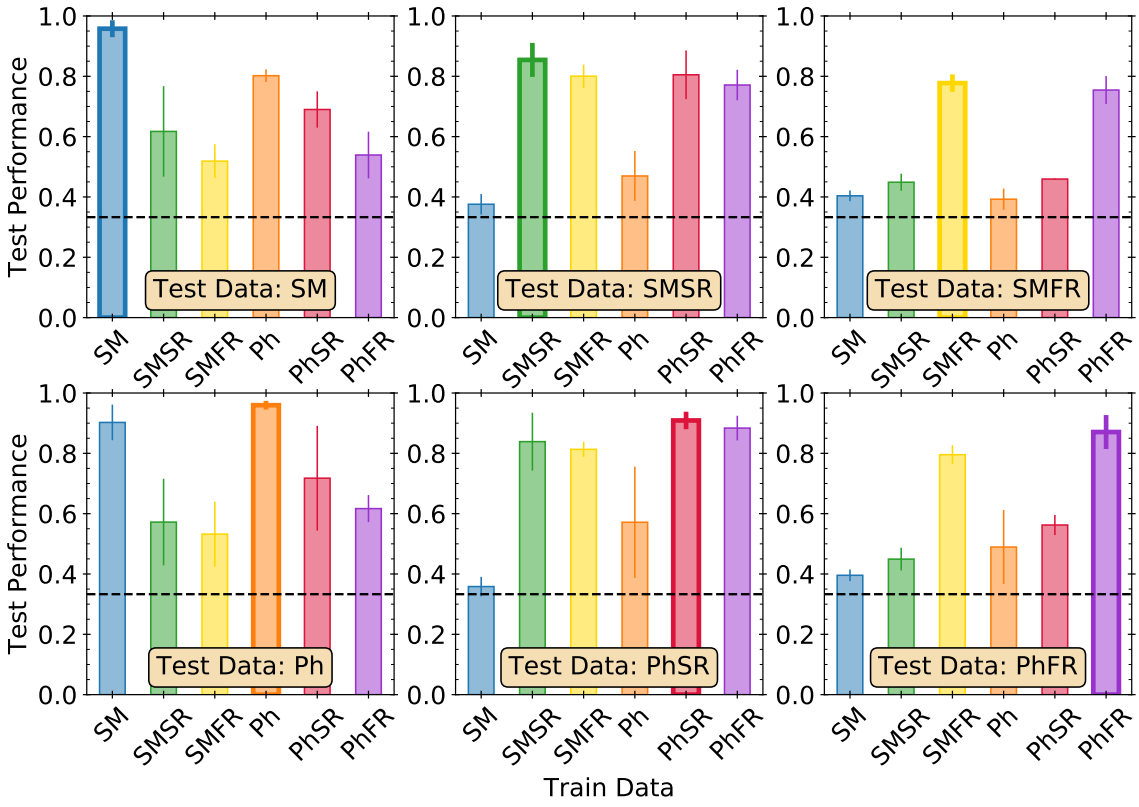

Figure 3.9 Overall median test performances in the network/data handshake. Each panel shows the median test results of applying a specified network to test images of every type. The ordering of test data types are indicated along the x-axes. For reference, the overall performance of a single test is computed from the number of test images in the diagonal elements of a confusion matrix as a fraction of the total number of images. Median overall performance is computed over the 10 bootstraps of training/validation/test sets. Similarly, errorbars show the 5^{th} to 95^{th} percentile range in overall performance. Bolded borders are placed on bars corresponding to cases in which the test data is of the same image type as the training data. The dashed line denotes a uniformly random classification performance – which in the case of three possible classes is 1/3. The only model that performs comparably with the PhFR model on the PhFR test data is the SmFR. This result demonstrates that matching the realism is more important than whether or not the training data derives from images generated with proper radiative transfer.

whether the timing preference for misclassified isolated galaxies seen in several Ph-based tests (e.g. Figure 3.4 and the right panel of Figure 3.7) and discussed in Section 3.3.1 persists for networks that are colourblind. If the timing preference persists and similar performance is achieved when a network is trained using a single band, then colour can be ruled out as a major factor in distinguishing pairs from isolated galaxies by the network.

More generally, we are also interested in knowing the degree to which overall network performance is sensitive to colour. There are important advantages of a network that can achieve high performance without exploiting colour and focuses primarily on morphological features. For example, star-formation correlates strongly with colour. Since interactions between gas rich galaxies are proven triggers of central star formation (Bushouse, 1987; Noguchi, 1991; Carlberg et al., 1994; Mihos & Hernquist, 1994, 1996; Barton et al., 2000; Springel, 2000; Smith et al., 2007; Cox et al., 2008; Ellison et al., 2008; Patton et al., 2011; Hopkins et al., 2013; Patton et al., 2013; Moreno et al., 2015; Sparre & Springel, 2016; Thorp et al., 2019), a colour-sensitive network may learn to exploit central star-formation to characterize interaction stage through its correlation with colour. However, the negative consequence of identifying/characterizing interactions based on triggered star-formation is that any study which then examines the relationship between interaction stage and star-formation is automatically biased. Therefore, it is of great value to know that the networks are able to make merger-stage classifications without exploiting colour information.

The single-channel experiments are divided into two handshakes. New Ph, PhSR and PhFR networks are each trained and tested using only the corresponding r-band and i-band images to produce a single 3×3 handshake for each band. As with the main handshake, I statistically combine the individual test results of 10 bootstraps of training/validation/test images for my final results. The results of every test are shown in Appendix C. I discuss selected results in the subsections that follow.

How important is colour to network performance?

Previous tests already give us reason to believe that colour is not an *essential* ingredient of satisfactory network success. For example, Figure 3.9 showed that even the SmFR networks (which do not use radiative transfer) achieve a reasonable median overall performance of 79.6% on PhFR test images using only single-channel input. Similarly, the upper right panel of Figure 3.6 showed that the colourblind, idealized

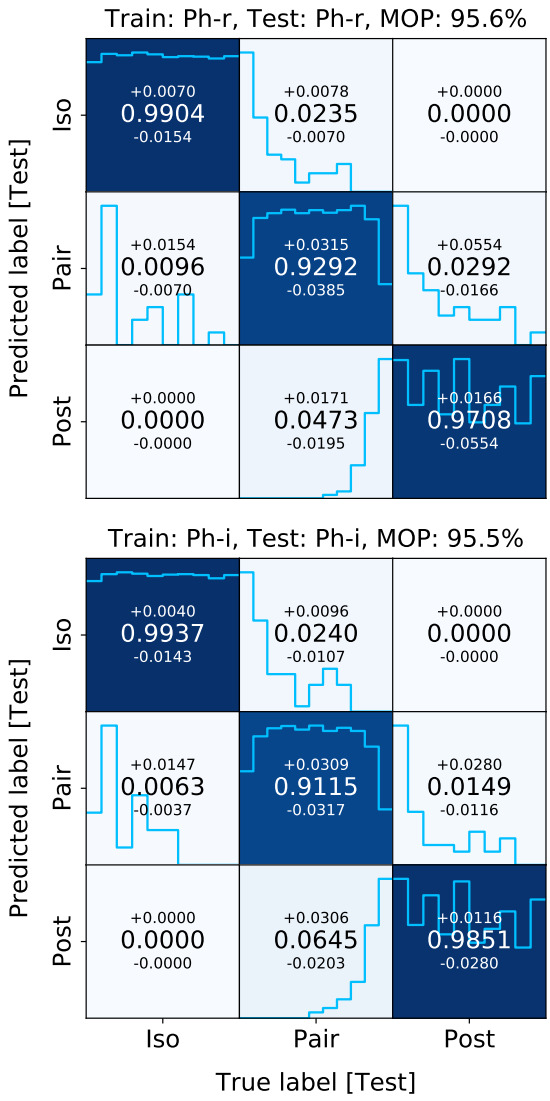

Figure 3.10 The importance of colour to network performance with idealized images. Network results shown in the upper and lower panels were trained and tested on single-channel r- and i-band idealized PH images, respectively. The networks achieve median overall performances of 95.6% (r-band) and 95.5% ($i-$band) compared to 96.0% when training and testing using idealized images from all three gri bands. The losses in performance when reducing to single-band photometry are minor.

Sm network achieved 90.2% median overall performance when tested on idealized Ph test images – only a 5.8% drop with respect to the results of the full-colour Ph network.

Figure 3.10 shows the confusion matrices for the idealized r-band (upper panel) and i-band (lower panel) Ph networks which were tested on the respective r-band and i-band Ph test images. For the idealized networks/data, the single-channel Ph networks still achieve exceptional overall performances. The change in median overall performance, ΔMOP, for each of the single-channel networks are minor with respect to the 96.0% performance of the three-channel, full-colour Ph networks from Figure 3.4: $\Delta\text{MOP}(r, i) = (-0.4, -0.5)\%$.

However, there is a potential problem with using the idealized images as a "representative" scenario for examining the importance of colour: the possibility of interplay between realism and colour. Between the three realism levels (idealized, SemiReal, FullReal), networks trained using idealized images are least likely to exploit colour information because the low surface brightness morphological features are most easily exploitable. In contrast, low surface brightness morphological features are often hidden in the sky noise or blurred by the PSF in the more realistic SemiReal and FullReal images. Consequently, a colour-sensitive network that is trained using these more realistic images is more likely to use colour to classify galaxies if the correlation between colour and merger-stage is strong *and* morphological information is limited.

To evaluate the importance of colour in the more realistic image data, I compare the single-channel PhFR test results (see Appendix C) with the three-channel, colour-sensitive PhFR network results (the right panel of Figure 3.7, MOP= 87.1%). The PhFR networks trained using individual bands have mild losses in MOP with respect to the three-channel networks: $\Delta\text{MOP}(r, i) = (-1.1, -2.3)\%$. While these are greater losses than in the idealized case, these results still demonstrate that colour is not an essential ingredient to network success and that the network is primarily targeting morphological features. However, without a training set which includes both red and blue discs, it is uncertain whether it is not simply this information that the network is exploiting to achieve the mildly higher performance in the full-colour data.

Lastly, I compare the timing histograms of both correctly and incorrectly classified galaxies for the single-channel networks (Figure 3.10) and three-channel networks (Figure 3.4). The results are qualitatively similar in every case – *including the timing preference for isolated galaxies which are misclassified as pairs.* The fact that the

timing preference persists in the single-channel networks disqualifies colour as being a driver of preferential misclassification of these isolated galaxies as pairs – in particular, those corresponding to early, high SFR, and morphologically unstable snapshots from the isolated simulation runs. A network that does not have access to colour cannot exploit its relationship with star-formation. Combined with the visual analysis of correctly and incorrectly classified isolated galaxies in Section 3.3.1 and Figure 3.5, this result demonstrates that the networks are focusing on morphological features. However, it should still be noted that certain morphological properties can be significantly enhanced by star-formation (for example, compactness). In my discussion of the limitations of the merger simulation suite, I argue that this problem can be solved using cosmological training sets which include greater variety of isolated galaxies and merger properties (gas fractions, initial morphologies, etc.).

Does the bandpass make a difference?

Table 3.3 shows that the typical sky surface brightness uncertainty, $\langle \sigma_{\mathrm{sky,Field}} \rangle$, in the SDSS r-band is 1.5 mag/arcsec^2 fainter than in the i-band. By comparing the performances in each of these bands individually, I determine the sensitivity of network performance to a modest change in imaging depth and the differences in the intrinsic brightnesses of targets in each band. Figure 3.10 shows that the difference in median overall performance in the i-band with respect to the r-band is only -0.1% in the idealized PH images, as expected. The idealized images contain no noise – so the only change between the r-band and the i-band are the brightnesses of stellar populations in each bandpass. The difference in performance broadens for networks trained on the more realistic single-channel PHFR images. The single-channel PHFR networks achieve median overall performances of 86.0% (r-band) and 84.8% ($i-$band) with a minor difference in MOP of -1.2% in the i-band with respect to the r-band. While this is a small change in performance, one must recall that source surface brightness in a bandpass diminishes rapidly with source redshift – by a factor of $(1 + z)^{-5}$. Consequently, a difference of 1.5 mag/arcsec may be a much greater hinderance to network performance for images of galaxies at the higher-z end of a realistic SDSS redshift distribution.

3.4 Discussion

3.4.1 The importance of realism

In Section 3.3.1 and Figures 3.7 and 3.8, I showed that adding full realism to synthetic training images (including realistic skies, resolution, and crowding by nearby sources) is a necessary condition of strong network performance in identifying and characterizing galaxy interactions in realistic images. Indeed, I have shown that every ingredient of full observational realism is essential and that omitting any ingredient leads to systematic misclassifications in testing. Without exposure to contamination by nearby sources in the training images, networks systematically preferred to assign images to the pair class. In particular, the systematic misclassification of fully realistic images as pairs persists even when the training images have both realistic skies and resolution but lack crowding effects.

Similarly, the tests in which I train on idealized Ph or Sm images and test on PhSR or SmSR images (orange and blue bars in the two middle panels of Figure 3.9) reveal that realistic skies and resolution are also vital (also see Figure E.1 of Appendix E). The sky and resolution effects bury and wash away the low surface-brightness morphological features that made idealized Ph or Sm networks successful on images of their own types (e.g. tidal tails, bridges and shells). Consequently, networks that are trained using idealized images perform very poorly on test images that contain realistic skies and resolution – even without crowding effects.[11] Combined with my results showing the independent biases that arise from excluding crowding effects in training images, the importance of each component of a FullReal image is clear.

Indeed, in Section 3.3.1, I demonstrated that the level of realism is even more important than whether the synthetic images originate from radiative transfer or from StellarMap images. The bottom right panel of Figure 3.9 shows that the only networks with performances that approach those of the PhFR networks on the PhFR test images are the SmFR networks – which exploit neither colour nor any higher-order features available from radiative transfer. Ultimately, we want to be able to construct training images from the current state-of-the-art cosmological hy-

[11] Note that this simultaneously demonstrates that the realism *should* be survey-specific. However, the results of Domínguez Sánchez et al. (2019) for morphological classifications demonstrate that it may be possible to use a *transfer learning* approach – in which CNNs or other deep learning models optimized for one survey can be adapted to another using a small sample of images from the target survey.

drodynamical simulations which will be used to train networks that can then identify and characterize real galaxy interactions by merger-stage. Only these simulations can provide the necessary scope, diversity and accuracy in galaxy properties (e.g. morphologies, masses, merger characteristics, orbital properties, gas fractions, etc.) to form training sets which are sufficiently generalizable to real galaxies and interactions. Consequently, training sets generated from these simulations will necessarily comprise very large numbers of synthetic images. However, in Section 3.3.1, we noted that constructing training sets on these scales using radiative transfer corresponds to potentially enormous computational and data management expenses. Therefore, the results that (1) radiative transfer is secondary to realism and (2) one can avoid radiative transfer using SmFR images for a modest compromise in performance are of great importance to the primary science application of the methods I present.

Lastly, with more diverse data (for example, data that are not so tightly temporally correlated) realism may become an even more important factor in generating large samples of images and training networks which effectively handle crowding effects. For the goals of this chapter, it was sufficient to insert each projection into a single field of view and increase the size of the datasets by augmentation. However, one could produce a much larger sample of images and train a network that may be even less sensitive to crowding effects by inserting *each* projection of a synthetic image into, for example, $N = 10$ unique fields. This would expose the network to a greater diversity of crowding effects for each target. As long as the training set is sufficiently large and each target gets an equal amount of unique insertions, the network will not overfit to a particular target. This would simultaneously ensure that no overfitting can ever occur due to particular configurations of targets and projected objects in the image FOV.

3.4.2 Limitations of the suite

As I have previously stated, the *specific* networks I trained in this study would have limited application to real data. This is primarily because the simulations are not cosmological. Despite the fact that the FIRE merger suite covers a broad range of mass ratios and orbital properties, the range is small in comparison to the parameter spaces of galaxy and merger properties encompassed by the observable Universe or by the current state-of-the-art cosmological hydrodynamical simulations (Patton et al., 2020; Blumenthal et al., 2020; Hani et al., 2020). Consequently, I know that my

training data are not generalizable to real data. Only cosmological simulations can provide the necessary scope to construct training data that are *both* similar and diverse enough to train networks that can be applied to real data.

However, generating networks that can be applied to real data is not a goal of this work. The goals are (1) to provide the methodology with which convolutional neural networks, trained and calibrated using hydrodynamical simulations, can be used to identify mergers and predict merger stage in *realistic* images and (2) to assess the importance of realism in the synthetic training images. The experiments I used to accomplish these goals did not require cosmological simulations. Indeed, my experiments correspond to scenarios in which I know that the training data are fully generalizable to the test data, as desired, because both training and test data are drawn from the same merger suite – just different parts of it. Ultimately, whether I use training data from a cosmological simulation or from the FIRE merger suite would make no qualitative difference to my results regarding the importance of realism. However, it will be important to assess whether the reduction in intrinsic simulation resolution in the cosmological simulations has an additional effect on performance. I aim to address this question in a follow-up study.

Another limitation of the suite used in the current work is that all of the galaxies used therein are relatively gas rich disks, and hence not representative of the full morphological complexity seen in the real Universe. This morphology bias would undoubtedly be an issue if I were to apply my networks to real data, but is not a limitation for the goals of this work. However, observations and numerical simulations alike show that mergers between gas-rich discs induce central star-formation in galaxies during the pair phase and in the merger remnant. Therefore, the relationship between merger-stage and star-formation may be exploited by networks that are sensitive to properties related to central star-formation. While I have demonstrated that eliminating colour sensitivity makes an insignificant difference to network performance in Section 3.3.2, I do not rule out a morphological connection with high central star-formation rates – such as with CAS Concentration index (Bershady et al., 2000; Conselice, 2003) or Gini coefficient (Abraham et al., 2003; Lotz et al., 2004). However, the increases in central surface-brightnesses from recent star-formation are associated with the low M/L of young stellar populations formed in the bursts. So while the PHOTOMETRY-based images and networks are more liable to exploit such connections between recent central star-formation and morphology, the STELLARMAP-based images will be largely insensitive to the morphological effects of recent star-formation

because they are *completely* insensitive to M/L ratio. Figure 3.9 showed that the SmFR networks performs nearly as well as the PhFR network. Indeed, a median overall performance of 95.8% is achieved with networks trained and tested on the idealized Sm images compared to the 96.0% achieved by the networks trained and tested on idealized Ph images. Therefore, while a connection between central morphology (as induced by central star-formation) and merger stage may exist in the Photometry-based images, it is not essential to network performance. Additionally, the exploitation of such a connection would be expected to be further suppressed in a more homogeneous galaxy sample (such as from a cosmological simulation) with mergers between galaxies that are red, blue, gas-rich, gas-poor and everything between.

3.4.3 Overfitting

As explained in the previous section, I know that the networks are limited to the set of merger scenarios encompassed by the FIRE merger suite with respect to galaxy and merger properties that would be present in a representative volume of the Universe. However, for evaluating the importance of realism, this limitation is immaterial because all I needed was a reasonably sized training set which includes typical merger features and test sets to which the training data are known to be generalizable. In contrast, a bias that would not be desirable is one that might arise from the construction of the synthetic images – such as camera angles or orientation. For example, in Figure 3.5, the correctly classified isolated galaxy image in the 3rd row, 2nd column of the left panel is the same galaxy and snapshot as the one that is incorrectly classified as a pair in the 3rd row, 1st column of the right panel. The only difference between these images is a slight change in zoom and rotation. A high sensitivity of predicted class to orientation is a common characteristic of overfitting – where a network learns to exploit properties of the training data that are not generalizable to test data.

While CNNs with max-pooling layers are architecturally invariant to translation, they are not rotationally invariant by default and require large and diversified training data to achieve *learned* rotational invariance (see Chapter 9, Figure 9.9 of Goodfellow et al. 2016 for an intuitive example). I apply rotational, translational, and zoom augmentation to all datasets in an effort to (1) increase the data size and (2) achieve rotationally invariant networks. Given that every image in the augmented training data (including all possible orientations) contributes equally to network optimization,

I find it unlikely that the networks are classifying based on orientation. However, another example from Figure 3.5 is the correctly classified inclined disc in the 3rd row, 4th column of the left panel and its incorrectly classified counterpart in the 1st row, 2nd column of the right panel. Both images correspond to the same galaxy and inclination – only the incorrectly classified one is from a much later snapshot and is rotated. Despite the visual similarity between these targets, the network confidently classifies these images as isolated and pair, respectively. Although Figure 3.4 shows that such misclassifications are rare, this high sensitivity between isolated and pair classifications, without obvious visible justification, may arise from my class definitions.

3.4.4 Class Definitions

By using hydrodynamical simulations to train networks, I attempt to eliminate as much subjectivity as possible for merger stage classifications. The advantage of this strategy is that, based on a set of simple quantitative definitions for each class, one is always optimizing network performance on the absolute truth. However, the definitions themselves are one remaining source of subjectivity that cannot be avoided in supervised learning. The beginning of the post-merger class requires a definition of coalescence which also defines the end of the pair class. The beginning of the pair phase also requires a definition. I defined the pair phase as beginning 100 Myr before first pericentric passage. Was the choice to use this temporal criterion appropriate? What were the consequences?

Figure 3.4 and the right panel of Figure 3.7 show that pairs that are misclassified as isolated are preferentially *early* pairs. The clear consequence of my definition is that galaxies in the early pair phase are indistinguishable from isolated galaxies because no galaxies in these early pairs have experienced visible disturbances resulting from gravitational interaction with their companions. Subsequently, this definition is also a likely culprit for the seemingly spurious misclassifications of a few isolated galaxies shown in Figure 3.5 that were discussed in the last section. However, for the purposes of this chapter, this definition happened to be beneficial (see Section 3.3.1). The fact that the networks had difficulty distinguishing early pairs (by my definition of the pair phase) from isolated galaxies was evidence that the networks were behaving intuitively. Meanwhile, since the majority of images from the pair class do not resemble isolated galaxies (all those except for the early pairs), the networks

still accurately classified most pairs in the test images.

Ultimately, I propose that reduced continuity between the isolated and pair images through an alternative definition of the pair phase would lead to better network performance and fewer misclassifications in these classes (for example, starting the pair phase at first pericentre). However, testing the sensitivity of performance to alternative definitions for each class is beyond the scope of this chapter. There is a large parameter space to be explored. The time or spatial separation at which a galaxy's properties start to be affected by an interaction and persist after coalescence are sensitive to the masses, morphologies, mass ratio and orbital properties at hand (e.g. Lotz et al. 2008, 2010b,a; Ji et al. 2014; Nevin et al. 2019). Nonetheless, I highlight that a few key advantages of calibrating networks using simulations are that, for a given set of class definitions, one can (1) train networks which make completely reproducible predictions and (2) evaluate the biases associated with these definitions. So, while *my* class definitions resulted in some confusion between early pairs and the isolated class, these definitions can be easily changed and optimized to improve performance.

3.5 Summary

Convolutional neural networks are becoming a popular tool for identifying galaxy mergers in large surveys. In this chapter, I use galaxy merger simulations to train CNNs which identify mergers and predict merger stage. I assess the importance of producing realistic images from simulations to the performance of CNNs, in order to guide future applications of this method.

I train and calibrate a set of convolutional neural networks using synthetic images generated from a suite of hydrodynamical binary merger simulations (Moreno et al., 2019) run with the FIRE-2 physical model (Hopkins et al., 2018). Training networks on simulations offers the significant benefit of foreknowledge of interaction stage and, therefore, optimization targets that are not biased by factors such as image quality or personal subjectivity. I examine the importance of adding realistic ingredients to the synthetic images. To do so, networks are trained using two types of galaxy images, stellar maps and dust-inclusive radiatively transferred images, each with three levels of observational realism: (1) no observational effects (idealized images), (2) realistic sky and point spread function (semi-realistic images) and (3) insertion into a real sky image (fully realistic images) (see Section 3.2.2 and summary in Table 3.2). Each

image dataset covers the same set projections and simulation snapshots and is divided into isolated, pair, and post-merger classes. In the main handshake experiment (see Section 3.3.1 and Figure 3.3), I test each network on data of every other type. Each network is also tested on data of the same type upon which it was trained but that the network never sees during training. The PhFR data – in which the synthetic images are injected into real survey fields – are the most realistic representation of real observations. Therefore, the PhFR test data are used to evaluate how well networks trained on images of a particular type would handle real data (see Section 3.4.2 for an important discussion on the limitations of *this suite* for applications to real data). The results of the main handshake experiment are:

1. [Section 3.3.1] **Networks trained on idealized images (Sm and Ph) classify images of the same type with** 96.0% **accuracy** (Figure 3.4 and the upper left panel of Figure 3.6). Misclassifications behave predictably. Early pairs are difficult to distinguish from isolated galaxies. Recent post-mergers are difficult to distinguish from pairs nearing coalescence. Isolated galaxies and post-mergers are *never* confused for one-another.

2. [Section 3.3.1] Sm **images can be used in place of more computationally expensive (but more realistic) images produced with radiative transfer at a modest cost in performance.** Networks trained on idealized Sm images classify idealized Ph images with 90.2% accuracy (upper right panel of Figure 3.6).

3. [Section 3.3.1] Ph **and** PhSR **networks – of which neither are exposed to training images which include contamination by nearby sources – systematically classify** PhFR **images as belonging to the pair class.** Networks trained on idealized Ph or semi-realistic PhSR images (realistic skies and resolution) both perform very poorly (48.9% and 56.2% accuracies, respectively) on the PhFR images (left and centre panels of Figure 3.7).

4. [Section 3.3.1] **As long as networks are exposed to all ingredients of realism in training (skies, resolution and crowding) they can learn to**

efficiently handle these effects in test images. While PH and PHSR networks fail to handle realistic images, networks trained on PHFR images classify PHFR test images with 87.1% accuracy (right panel of Figure 3.7). Additionally, (a) there is no clear systematic preference toward classifying images as pairs and (b) PHFR networks are even more accurate on PHSR test images (88.4%) than PHFR test images (see Figure 3.8).

5. [Section 3.3.1] **Realism is more important than whether the images originate from radiative transfer or from maps of stellar mass.** Networks trained on SMFR images classify PHFR test images with 79.6% accuracy (lower right panel Figure 3.9). Indeed, these are the only networks other than the PHFR networks that achieve reasonable performance on the PHFR test images.

I perform a secondary handshake experiment aimed at characterizing the roles of colour and depth to network performance (see Section 3.3.2). Single-channel networks are trained on the r and i band images, individually, taken from the PH, PHSR and PHFR datasets. These tests eliminate the possibility for networks to exploit colour information and allow one to compare results for networks trained on images in bands of varying photometric depths. The main results of these tests are:

1. [Section 3.3.2] **Networks trained without colour incur very mild penalties to performance with respect to colour-sensitive networks**. The performances of the single-channel r- and i-band PH (PHFR) networks are 95.6% (86.0%) and 95.5% (84.8%), respectively, compared to 96.0% (87.1%) with the full-colour networks (see Figure 3.10). These results demonstrate that, while the colour-sensitive networks *can* exploit colour information, colour is not a *necessary* ingredient for high network performance.

2. [Section 3.3.2] **The difference in average photometric depth between the r- and i-bands ($\sim$ 1.5 mag/arcsec^2) yields a small difference in the performances of networks trained on each band individually** (see Figure 3.10). However, in this study, I do not match the redshift distribution of SDSS galaxies and instead insert galaxies at the median redshift of galaxies in the DR14 MaNGA galaxy sample. Therefore, these differences might be

expected to be larger for training and test data which include galaxies that are more distant or have lower intrinsic brightnesses.

The pertinent applications of this work are: (1) to train networks using realistic synthetic images from cosmological simulations and (2) to use a model trained on cosmological simulations to identify and characterize interactions in the real Universe. The most important feature of cosmological simulations in this respect is that mergers *and* isolated galaxies that are selected from a statistically representative simulation will cover a larger range of morphologies, masses, gas fractions, etc. This diversity will be a necessary component of a training set that can be expected to perform well on real test data.

Chapter Acknowledgements

I acknowledge the support of a National Sciences and Engineering Research Council of Canada (NSERC) Graduate Scholarship. The individuals who contributed to this project and are listed in the corresponding publication (Bottrell et al., 2019a). The contributors are listed in order of their relative contributions specific to the project. Maan H. Hani (MHH) and Hossen Teimoorinia (HT) contributed equally to this research. MHH was responsible for the highly significant data processing component of running the SKIRT radiative transfer code on the simulations and generating the stellar mass projections. HT committed four months to leading and directing my studies in machine learning and deep learning. HT also provided invaluable technical advice for this project. Sara L. Ellison provided valuable scientific advice. Luc Simard's image processing algorithms provided the knowledge-base from which my own algorithms were developed. Jorge Moreno (JM) and Paul Torrey (PT) produced the FIRE-2 merger simulations used in this study. The data used in this chapter were, in part, generated and hosted using facilities supported by the Scientific Computing Core at the Centre for Computational Astrophysics, a division of the Simons Foundation. The opportunity to use these facilities was enabled through collaboration with Christopher C. Hayward – who also provided scientific advice. Mallory D. Thorp compiled the galaxy catalog for MaNGA DR14 that was used derive median redshift of the synthetic observations. The data processing and analysis conducted for this project were enabled, in part, by the computational resources provided by Compute Canada (www.computecanada.ca). The numerical simulations in this chapter were run on

the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. The opportunity to use these facilities was enabled through JM and PT's collaboration with Lars Hernquist. I also thank the anonymous referee for the corresponding publication to this chapter whose suggestions helped to improve the quality of the work.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Chapter 4

Toward improved merger classification with kinematics

Preamble

The previous chapter showed that galaxy mergers can be distinguished from non-merging galaxies and characterized by stage (pair and post-merger) using photometry. In particular, Chapter 3 highlights the promise of combining hydrodynamical simulations, synthetic observations, and deep learning in executing these tasks. Chapter 3 also emphasizes the requirements for calibrating models that will be successful in classifying real galaxy mergers: the synthetic training data must be (a) *highly comparable* to real data in terms of observational limitations and effects and (b) statistically generalizable in terms of galaxy and merger properties encompassed by the simulations. Furthermore, the accuracy achieved by the models in Chapter 3 on the *fully realistic* test images was always $< 90\%$ – even for the relatively small set of disc galaxy merger simulations used therein. Therefore, while the results of Chapter 3 can be largely considered encouraging, generating pure and complete samples of real galaxy galaxy mergers will require: (1) greater accuracy and (2) statistically representative training data. Satisfying criterion (2) may be accomplished by generating much larger sets of synthetic observations using simulations which model the evolution of many thousands of galaxies. Satisfying criterion (1) may require complimentary data to aid in the classifications – namely, spatially resolved galaxy kinematics. The work presented in this chapter is intended for a future publication.

4.1 Introduction

Galaxy mergers have a fundamental and critical role within the Lambda cold dark matter (ΛCDM) concordance cosmogony (White & Rees, 1978; White & Frenk, 1991). In this paradigm, large structures form through continuous and diverse merging events between smaller structures (e.g. Lacey & Cole 1993). However, the role of mergers in galaxy formation and evolution is not limited to the *ex-situ* assembly of mass.

Theoretical models and numerical simulations show that the strong gravitational and tidal forces in mergers can be responsible for morphological and dynamical transformation of galaxies (e.g. Toomre & Toomre 1972; Toomre 1977; Negroponte & White 1983; Frenk et al. 1985; Hernquist 1992; Steinmetz & Navarro 2002; Naab & Burkert 2003; Jesseit et al. 2007; Hopkins et al. 2009; Naab et al. 2014). These theoretical results are supported by the disrupted morphologies (e.g. Patton et al. 2005; Hernández-Toledo et al. 2005, 2006, 2007; De Propris et al. 2007; Ellison et al. 2010; Casteels et al. 2013; Patton et al. 2016) and spatially-resolved kinematics (e.g. Barrera-Ballesteros et al. 2015; Hung et al. 2015; Bloom et al. 2017, 2018; Pilyugin et al. 2020; Feng et al. 2020) observed in interacting galaxies and recent merger remnants (*post-mergers*) compared to non-merging galaxies. Transformation of galaxy morphologies and kinematics through interactions and mergers is also broadly consistent with the observational framework relating morphology and kinematics to galaxy environment (e.g. Oemler 1974; Davis & Geller 1976; Dressler 1980; Bamford et al. 2009; Cappellari et al. 2011). In addition to morphological and kinematic transformation, interactions are connected to a number of internal processes. Compared to non-interacting galaxies, observed merging galaxies and post-mergers exhibit enhanced central star formation rates (e.g. Barton et al. 2000; Lin et al. 2007; Ellison et al. 2008; Scudder et al. 2012; Patton et al. 2013; Ellison et al. 2013; Thorp et al. 2019; Pan et al. 2019), suppressed central gas metallicities (e.g. Kewley et al. 2006; Ellison et al. 2008; Michel-Dansac et al. 2008; Rupke et al. 2010b; Ellison et al. 2013; Thorp et al. 2019), and increased incidence of AGN (e.g. Alonso et al. 2007; Ellison et al. 2011; Silverman et al. 2011; Satyapal et al. 2014; Goulding et al. 2018; Ellison et al. 2019).

However, the *relative* role of galaxy mergers in these morphological, dynamical, and internal processes and their sensitivities to merger stage and initial conditions are currently uncertain. For example, what is the relative contribution of mergers to triggering star formation and AGN compared to the contribution from accretion of

cold gas along cosmic filaments (e.g. Dekel & Birnboim 2006; Birnboim et al. 2007; Dekel et al. 2009; Bournaud et al. 2011)? What is their relative role in the build-up of stellar mass in galaxies? Is the role of mergers in these processes static or does it evolve with cosmic time? Ultimately, robustly characterizing the role of mergers in galaxy evolution requires (1) statistically representative and unbiased censuses of observed galaxy mergers and (2) the ability to link galaxy mergers to their stages and initial conditions (stellar mass ratio and orbital parameters of the merging galaxies, their gas fractions, and morphologies). Both of these criteria present significant challenges from an observational perspective.

Models that can quantitatively classify and characterize galaxy mergers automatically present an attractive solution to these challenges because they overcome the potential for ambiguity in visual classifications and can rapidly process large samples of galaxies. Furthermore, automatic merger classification models can be calibrated based on synthetic observations of galaxies and galaxy mergers from hydrodynamical simulations (e.g. Lotz et al. 2008, 2010a,b; Pearson et al. 2019; Nevin et al. 2019; Bottrell et al. 2019a). The advantage of this approach is that the simulations provide *a priori* knowledge of whether galaxies are undergoing interactions as well as the stage and initial conditions of the interactions. Consequently, the biases and limitations in estimating these parameters for observed galaxies do not become embedded in the calibration of the models. Indeed, the observational features which visually distinguish merging galaxies from non-merging galaxies (such as stellar tails, tidal arms, bridges, and shells, and kinematic asymmetries) are motivated by numerical simulations (e.g. Toomre & Toomre 1972; Hernquist & Quinn 1989; Barnes & Hernquist 1992; Mihos et al. 1998; Helmi & White 1999; Colina et al. 2005; Johnston et al. 2008; Blumenthal et al. 2020). Consequently, an automatic merger classification model that is calibrated on realistic synthetic observations from simulations will be optimized to distinguish mergers from non-mergers based on the same set of features that are expected to appear in observations.

Nonetheless, an automatic merger classification model that can generate pure and complete merger samples remains elusive. Despite their variety and advantages, automatic classifiers are not immune to statistical laws. In particular, the statistical prior from the low incidence of observed mergers presents one of the greatest challenges to automatic merger classification. Figure 4.1 illustrates the role of statistical priors in a hypothetical binary classification model. The hypothetical model is designed give to a positive result (+) for Group A and a negative result (−) for Group B where

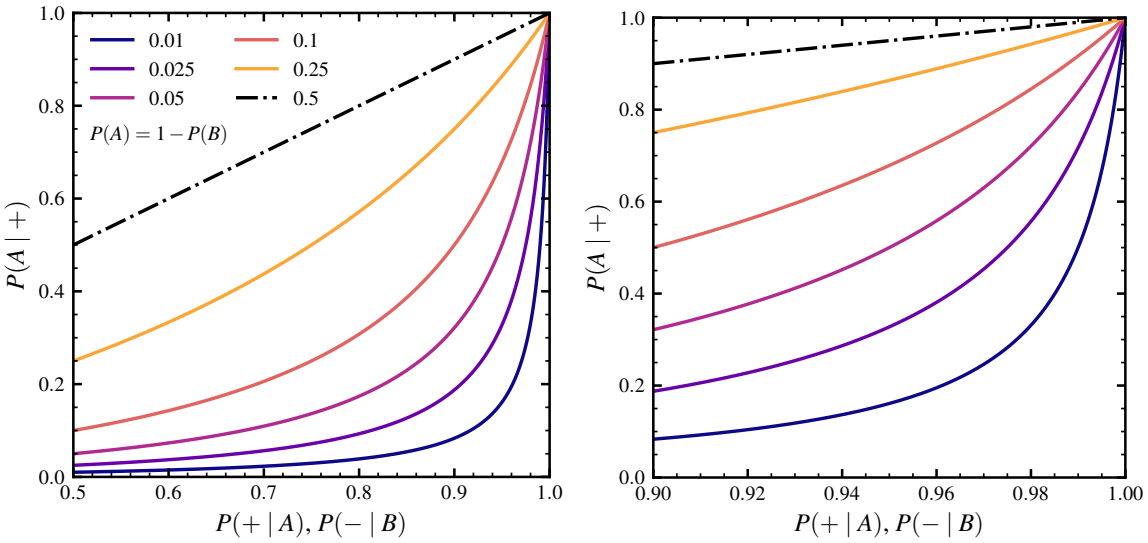

Figure 4.1 An illustration of the role of priors in Bayesian statistical inference with a hypothetical binary classification model. The left panel shows the relationship between the posterior probability, $P(A|+)$, and a hypothetical model's sensitivity, $P(+|A) > 0.5$, for a range of of priors, P(A), plotted as dash-dot and coloured curves. The specificity of this hypothetical model is equal to its sensitivity, $P(B|-) = P(A|+)$, and the two groups are complementary, $P(A) = 1 - P(B)$. The right panel zooms-in on $P(+|A) > 0.9$ for better visibility.

Groups A and B are complementary, $P(A) = 1 - P(B)$. For perspective, Group A could be mergers while Group B comprises non-mergers. Equivalently, Group A could be people who have a disease while Group B comprises people who do not. $P(+|A)$ is then the probability of a positive classification for a member of Group A (true positive rate, or *sensitivity*). $P(-|B)$ is the probability of a negative classification for a member of Group B (true negative rate, or *specificity*). Consequently, the false positive rate (fall-out) and false negative rate (miss rate) of the binary classifier are $P(+|B) = 1 - P(-|B)$ and $P(-|A) = 1 - P(+|A)$.

The hypothetical model in Figure 4.1 has equal sensitivity and specificity, $P(+|A) = P(-|B)$. The left panel shows the *posterior* probability, $P(A|+)$, that a sample is truly a member of Group A given that it received a positive (+) result following Bayes' Theorem shown in Eq. 4.1 (Bayes, 1763). $P(A|+)$ is plotted as a function of sensitivity for a range of prior incidence probabilities, $P(A)$ – plotted as dot-dashed and coloured curves. The right panel zooms-in on $P(+|A) > 0.9$. Figure 4.1 shows that a decrease in the underlying incidence of Group A with respect to the overall population (Group A and B) strongly effects the posterior probability. For example, even with 99% sensitivity and specificity but prior incidence of 1% (navy curve), the corresponding posterior probability that a sample is actually a member of Group A given that it has a positive (+) result from the model is only 50%.

$$P(A|+) = \frac{P(+|A)P(A)}{P(+)} = \frac{P(+|A)\ P(A)}{P(+|A)\ P(A) + P(+|B)\ P(B)} \tag{4.1}$$

The selection functions for galaxy mergers (which could be taken the place of Group A in Figure 4.1) varies greatly – both in terms of merger definition and selection methods. For reference, typical merger fraction estimates in the nearby Universe ($z \lesssim 0.5$) are in the range of $\sim 1-10\%$ (e.g. Lin et al. 2004; Kartaltepe et al. 2007; Lin et al. 2008; Patton & Atfield 2008; de Ravel et al. 2009; Lotz et al. 2011; López-Sanjuan et al. 2012). For a merger classification model that is 99% sensitive and specific, merger fractions in this range would yield corresponding posterior probabilities of $50 - 92\%$. Figure 4.1 also shows that these purities fall rapidly with decreasing sensitivity. At $P(+|A) = P(-|B) = 0.9$, galaxies that are classified as positive (+) by the model only have $8 - 50\%$ probabilities of actually being a merger. In other words, the sample of "mergers" classified by the model would realistically comprise $8-50\%$ true mergers and, respectively, $92-50\%$ contaminating non-merger galaxies.

Presently, automatic merger classification and characterization models are predominantly based on optical morphologies or other information derived from imaging (e.g. recently Pawlik et al. 2016; Ackermann et al. 2018; Snyder et al. 2019; Walmsley et al. 2019; Pearson et al. 2019; Nevin et al. 2019; Bottrell et al. 2019a). However, classification performance may be improved using other information from which a model can draw – independently, or in tandem with imaging. In particular, mergers and recent post-mergers exhibit complex kinematics compared to non-merging galaxies (e.g. Mihos et al. 1998; Colina et al. 2005; Shapiro et al. 2008; Hung et al. 2015; Bloom et al. 2018; Feng et al. 2020) which are sensitive to interaction stage and other merger characteristics (Hung et al. 2016). Importantly, kinematics and morphology can tell different stories about the state of a galaxy. Merging galaxies and post-mergers can appear morphologically ordinary in images that are too shallow to reveal their low surface-brightness tidal signatures (Lotz et al., 2008, 2010b; Ji et al., 2014; Bottrell et al., 2019b). Such galaxies (particularly post-mergers) may still show elevated disturbances in their kinematics (Hung et al., 2016). Similarly, non-merging galaxies with high star formation rates can be morphologically asymmetric or "clumpy" but kinematically regular (Glazebrook, 2013). Galaxy kinematics are therefore highly complementary to imaging from the perspective of merger characterization and may help to overcome the challenges of building accurate models.

Fortunately, modern highly-multiplexed IFS surveys have generated a revolutionary expansion in the volume and quality of spatially-resolved galaxy kinematic data in the local Universe (e.g. the Calar Alto Legacy Integral Field Area (CALIFA): 600 galaxies Sánchez et al. 2012, the Sydney-AAO Multi-object IFS survey (SAMI): 3,600 galaxies Croom et al. 2012, the Mapping Nearby Galaxies at Apache Point Observatory survey (MaNGA): 10,000 galaxies Bundy et al. 2015). Indeed, the growth and quality of the data are expected to continue with forthcoming instruments (e.g. Hector: 15,000 galaxies Bryant et al. 2018). Consequently, galaxy kinematics have finally become both attractive and *viable* for building pure and reasonably complete merger samples that can subsequently be further characterized. Synthetic kinematic observations of galaxies and mergers from hydrodynamical simulations are likely to be important in realizing this goal – given that they enable proper calibration of automatic merger classification models with foreknowledge of the true target classification. However, Bottrell et al. (2019a) showed that the level of *realism* in synthetic observations is crucial to the calibration of models which can then be reasonably applied to real observations – particularly for deep learning classifiers which are sensitive

to high-order features in the data. That study (Chapter 3 of this thesis) focused on imaging but it is intuitive that the results should hold for kinematics. Briefly, if the characteristics of the synthetic calibration data are not highly comparable to real data to which the model is to be applied, the model can be expected to produce spurious results and perform poorly.

Currently, no tool exists for generating synthetic galaxy kinematic data that is highly comparable to the data from modern IFS instruments. In this chapter, I satisfy the need for such a tool with RealSim-IFS[1] – an expansion of the RealSim[2] suite for generating realistic synthetic IFS observations of galaxies from hydrodynamical simulations. In particular, RealSim-IFS makes use of the rigorous documentation for the technical specifications, data reduction, and data analysis pipelines of IFS instruments (e.g. CALIFA: Mármol-Queraltó et al. 2011, SAMI: Bryant et al. 2015; Sharp et al. 2015, MaNGA: Law et al. 2016; Westfall et al. 2019) in an attempt to achieve synthetic data characteristics that match the characteristics of real fibre-based galaxy IFS data. The suite is demonstrated in an application to galaxies from the IllustrisTNG cosmological magnetohydrodynamical simulations (Pillepich et al., 2018b,a; Nelson et al., 2018) using merger histories and new ancillary data synthesized by Hani et al. (2020). These demonstrations are geared towards creating synthetic MaNGA survey data products. Testing is also carried out with real fibre data from the MaNGA survey to examine the continuity that can be achieved between the RealSim-IFS pipeline and the MaNGA data reduction pipeline (Law et al., 2016).

This chapter is laid out as follows. Section 4.2 describes the simulations, their data products, and the real MaNGA observations that are used in testing and in generating survey-realistic synthetic MaNGA kinematics of IllustrisTNG galaxies. In particular, the characteristics of the data products outlined in Section 4.2 provide the basis for shaping the methods – which are summarized in Section 4.3. Section 4.4 presents the results of various tests and visually illustrates the usefulness of realistic synthetic galaxy kinematics for classifying mergers. This chapter is summarized in Section 4.5. The calculations in this chapter adopt a Planck-based cosmology with $H_0 = 67.4$ km s^{-1} Mpc^{-1}, $\Omega_{\rm m} = 0.315$, and $\Omega_\Lambda = 0.685$ (Planck Collaboration et al., 2018).

[1] https://github.com/cbottrell/RealSim-IFS/
[2] https://github.com/cbottrell/RealSim

4.2 Data

This section describes the data used for the various tests and demonstrations with RealSim-IFS shown later in Section 4.4. First, one of the primary goals of this chapter is to describe RealSim-IFS and demonstrate its application to generating realistic synthetic kinematic observations of galaxies from hydrodynamical simulations. The IllustrisTNG cosmological magneto-hydrodynamical simulation suite is used in these demonstrations (Weinberger et al., 2018; Pillepich et al., 2018a,b; Springel et al., 2018; Naiman et al., 2018; Marinacci et al., 2018; Nelson et al., 2018, 2019a,b). The IllustrisTNG simulations are summarized in Section 4.2.1. The construction of line-of-sight velocity distribution (LOSVD) cubes for galaxies from these simulations by Maan H. Hani is described in Section 4.2.2. Finally, certain tasks within RealSim-IFS can be used on real fibre-based IFS survey data (for example, the spatial reconstruction of fibre intensities onto a Cartesian grid) and validated against the final data products from these surveys. These validation tests are carried out using data from the MaNGA survey (Bundy et al., 2015; Drory et al., 2015; Law et al., 2015, 2016; Westfall et al., 2019) – which is summarized in Section 4.2.3.

4.2.1 IllustrisTNG

Demonstrations producing survey-realistic synthetic observations using RealSim-IFS are performed using publicly available data from the IllustrisTNG simulations (Weinberger et al., 2018; Pillepich et al., 2018a,b; Springel et al., 2018; Naiman et al., 2018; Marinacci et al., 2018; Nelson et al., 2018, 2019a,b). IllustrisTNG comprises a suite of large-volume cosmological magneto-hydrodynamical simulations of galaxy formation and evolution. The simulations trace the evolution of dark matter, gas, stars, and supermassive black holes from redshift $z = 127$ to $z = 0$ using the AREPO moving-mesh hydrodynamic code (Springel, 2010). The simulations are run in three cubic volumes (51.7^3, 110.7^3, 302.6^3) Mpc^3 with descending levels of resolution. All use the same physical model described by Weinberger et al. (2018) and Pillepich et al. (2018a) – which builds on the original Illustris model (Vogelsberger et al., 2013; Torrey et al., 2014; Vogelsberger et al., 2014a; Genel et al., 2014).

The highest resolution run for the 110.7^3 Mpc^3 volume (TNG100-1) is adopted for testing with RealSim-IFS. The TNG100-1 volume comprises a statistically representative sample of galaxies with realistic and diverse morphologies and kinematics (e.g. Rodriguez-Gomez et al. 2019; Tacchella et al. 2019; Huertas-Company et al. 2019;

Du et al. 2020). Given that one of the primary applications of RealSim-IFS is to produce realistic synthetic kinematic data for galaxy and galaxy merger classification, TNG100-1 is a good candidate because it offers (1) a diverse and representative galaxy population whose properties should be generalizable to real galaxies and (2) an effective baryonic spatial resolution that is similar to the median physical resolution for observed targets in the MaNGA and SAMI surveys (~ 1 kpc)[3]. The merger trees of TNG100-1 galaxies are derived using the Sublink algorithm (Rodriguez-Gomez et al., 2015) – which links a given galaxy (in a given snapshot of the simulation) to its progenitor(s) in the previous (higher redshift) snapshot and to its descendent in the next snapshot. I also make use of a wealth of ancillary data on the mass ratios, separations, time since the most recent merger, and other merger-related properties derived by Patton et al. (2020) and Hani et al. (2020).

4.2.2 LOSVD cubes

RealSim-IFS can be applied to a number of data products from the simulations to obtain synthetic kinematics. The *most* realistic approach is to use radiative transfer (e.g. Jonsson 2006; Jonsson et al. 2010; Baes et al. 2011; Steinacker et al. 2013; Camps & Baes 2015). In radiative transfer, each stellar particle is assigned a SED based on its mass, age, metallicity, and velocity. These SEDs can then be propagated through gas and dust (which themselves imprint their velocities) toward an arbitrarily placed camera to produce idealized synthetic *spectral* datacubes in which each spaxel contains a spectrum (with no atmospheric or instrumental effects). This data format is the most similar to real IFS data. However, this fidelity comes with a significantly increased workload. To then derive kinematics, the LOSVD of the underlaying particles in the spectrum in each spaxel must be derived from LOSVD-convolved emission line and stellar template fitting (e.g. ppxf, Cappellari 2017). This process is exceptionally labour- and data-intensive – going well beyond the scope of testing RealSim-IFS.

A less intensive approach than radiative transfer is to produce idealized LOSVD cubes directly from the particle data – such that each spaxel contains the LOSVD

[3]There are 6,492 galaxies with stellar masses $M_\star \geq 10^{10} M_\odot$ in the $z = 0$ snapshot from TNG100-1. There are 303,110 galaxies above this threshold in TNG100-1 snapshots ranging from $z = 0$ to $z = 1$. While TNG50 offers greatly improved spatial resolution, using TNG50 reduces the galaxy sample size by around a factor of 10. In contrast, TNG300 offers a significantly larger galaxy sample than TNG100-1 (by a factor of 20). However, this increase in sample size in TNG300 comes at a considerable cost in both mass resolution (factor of eight decrease compared to TNG100-1) and spatial resolution (factor of two decrease).

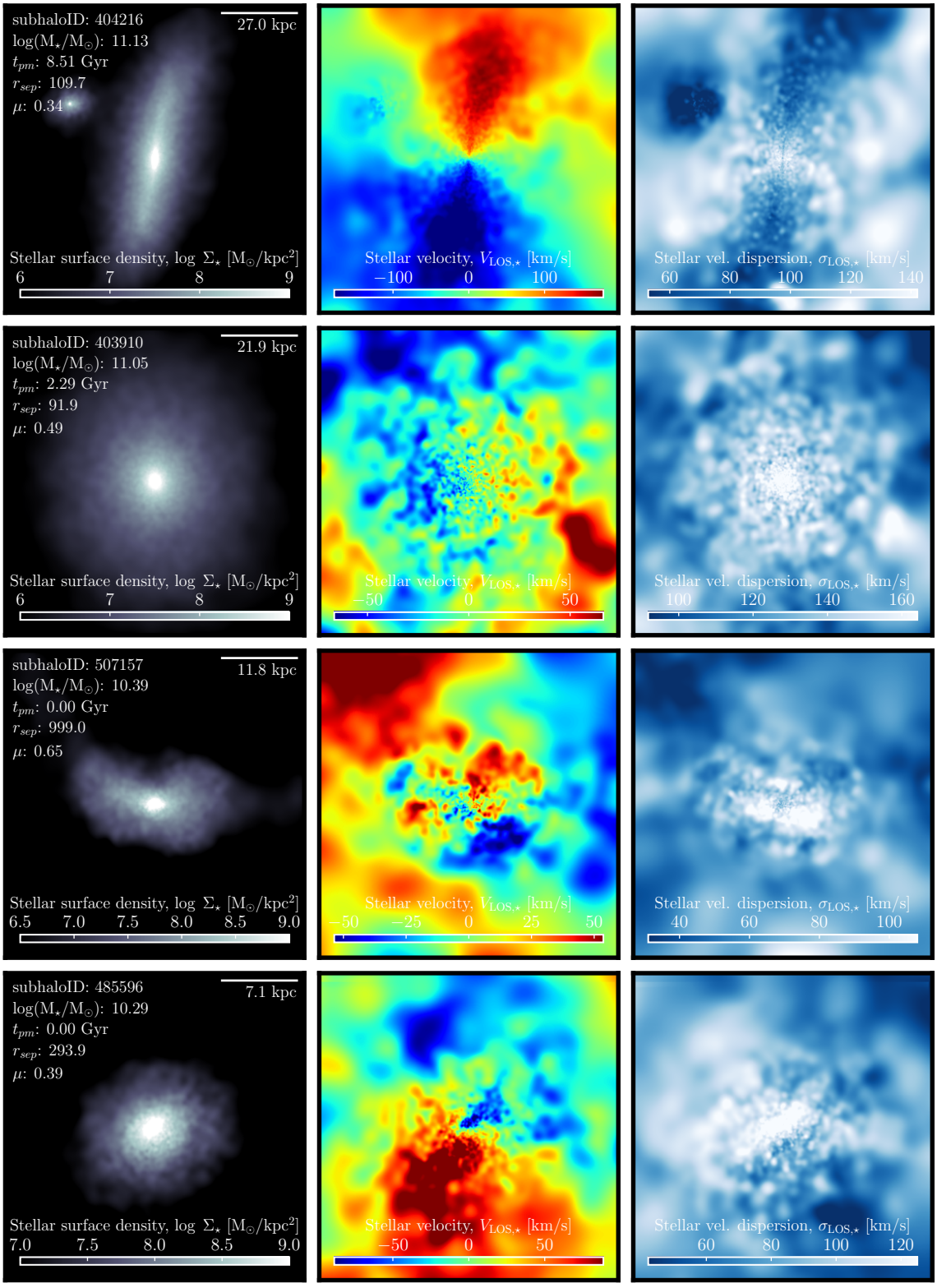

Figure 4.2 Moments derived directly from idealized stellar LOSVD cubes for four galaxies from TNG100-1. From left to right, panels show the stellar surface density, velocity, and velocity dispersion, respectively. See Section 4.2.2 for further description.

of the gas or stellar particles it subtends. This avoids radiative transfer and spectral fitting, but is similar in structure and nature to spectral datacubes. In particular, the incorporation of the spatial response for an IFS instrument to idealized synthetic spectra and LOSVDs is equivalent. Consequently, for testing RealSim-IFS, idealized stellar LOSVD cubes were constructed by Maan H. Hani for galaxies from snapshot 99 ($z = 0$) of the TNG100-1 simulations with stellar masses $\log(\mathrm{M}_\star/\mathrm{M}_\odot) \geq 10$. This stellar mass cut, which corresponds to a stellar particle number of ~ 7000 in TNG100-1, is made so that stellar kinematic structures (such as discs and spheroids) should be at least reasonably resolved.

The stellar LOSVD cube for each galaxy has dimensions (300, 512, 512). The velocity axis ranges from $[-700,\ 700]$ km s^{-1} in bins of 4.67 km s^{-1} with respect to an origin defined at the gravitational potential minimum of the galaxy. The FOV of the spatial axes are adaptively set to 10 times the physical half stellar mass radius of the galaxy, $R_{1/2}$ (the radius containing 50% of the stellar mass of the galaxy). The intensities in each velocity element have units of stellar mass, such that the summation over the velocity axis gives the total stellar mass within a spaxel. Lastly, each stellar particle in the simulation does not represent a single star but rather a full unresolved stellar population of mass $\mathrm{M}_\star \approx 1.4 \times 10^6\ \mathrm{M}_\odot$. Therefore, to avoid unrealistically discretized distributions of stellar mass in the cubes, a cubic spline smoothing kernel (Monaghan, 1992) is applied spatially to the data with an adaptive characteristic radius equal to the the distance to the 32$^{\mathrm{nd}}$ nearest neighbouring particle. LOSVD cubes are generated in four camera orientations along the arms of a tetrahedron that is randomly oriented with respect to the galaxy.

Figure 4.2 shows four examples of idealized LOSVD moment maps (amplitude, velocity, velocity dispersion) derived from these cubes by assuming single-component Gaussian LOSVDs[4]. The zeroth moment of the LOSVD is simply the stellar mass, and has been converted to stellar surface density, $\Sigma_\star$, using the physical scale of each spaxel in the first panel of each row. The first two rows show galaxies with relatively regular morphologies and kinematics. The first row shows a nearly edge-on spiral galaxy with kinematics dominated by a rotating stellar disc. A small companion is seen next to it (less than a tenth of the stellar mass of the primary galaxy). The second row shows a galaxy with more dispersion-dominated kinematics and elliptical-

[4]While the LOSVD in each spaxel may generally include contributions from multiple kinematic components which should be modelled simultaneously, a single-component Gaussian model is adopted for simplicity in testing.

like morphology.

The third and fourth rows of Figure 4.2 show recent post-mergers, denoted by the time since the most recent merger, $t_{\rm pm}$, for which the merger stellar mass ratio[5] was at least $\mu = 0.1$. $t_{\rm pm} = 0$ Gyr indicates that the merger event happened between the last snapshot and the current snapshot, a span of 162 Myr in TNG100-1. The third row shows a galaxy with strong disturbances visible in both the morphology and kinematics – particularly in the stellar velocity map. The bottom row shows an important case in which the galaxy's morphology is regular (at least in this orientation), but has highly irregular kinematics. These examples highlight the viability of using kinematics to improve the classification of galaxy mergers and their remnants (independently or combined with photometric morphologies). The left panels also show $r_{\rm sep}$ for each galaxy – defined by Patton et al. (2020) as a dimensionless, relative distance to the nearest companion that is at least a tenth of the host's stellar mass:

$$r_{\rm sep} = \frac{r}{R_{1/2}^{\rm host} + R_{1/2}^{\rm comp}} \tag{4.2}$$

where r is the 3-dimensional radial separation and $R_{1/2}^{\rm host}$ and $R_{1/2}^{\rm comp}$ are the half- stellar mass radii of the main target and its companion, respectively. All galaxies in Figure 4.2 have high $r_{\rm sep}$ ($r_{\rm sep} = 999$ indicates that $r_{\rm sep} > 300$ in this figure). Later, $R_{1/2}$ is taken to be the effective radius of each galaxy when deciding on appropriate MaNGA IFU designs.

4.2.3 MaNGA

Some tests with RealSim-IFS are carried using MaNGA IFS data from Data Release 15 (DR15) the SDSS (Aguado et al., 2019). The MaNGA survey (Bundy et al., 2015) aims to obtain observations for $\sim 10,000$ nearby galaxies with $\log(\mathrm{M}_\star/\mathrm{M}_\odot) \gtrsim 9$ from the SDSS main galaxy sample (Strauss et al., 2002). The survey uses a suite of custom-built fibre-based integral field units (IFUs, Drory et al. 2015) fed into the Baryon Oscillation Spectroscopic Survey (BOSS) spectrographs (Smee et al., 2013) installed on the SDSS 2.5 m telescope at Apache Point Observatory (Gunn et al., 2006). Each IFU comprises a number of 2 arscecond fibres bundled together in a hexagonal pattern.

[5]The merger stellar mass ratio, μ, is always defined as the ratio of the *less* massive companion in the merger to the more massive companion. Under this definition, $\mu \leq 1$ for all mergers.

The IFUs cover a range of fibre numbers and corresponding observational footprints. The IFUs used for galaxy observations are bundles of 19, 37, 61, 91, and 127 individual fibres with respective FOV diameters of 12.5, 17.5, 22.5, 27.5, and 32.5 arcseconds. The IFU design for a given galaxy is selected to cover out to 1.5 effective radii for 63% of the full MaNGA sample (called Primary and Colour-Enhanced samples) and 2.5 effective radii for 37% (Secondary sample) (Wake et al., 2017). A more detailed description of the target selection follows in Section 4.4.2 from which synthetic observation statistics are drawn (such as redshift and atmospheric seeing).

Each MaNGA observation comprises between 2 and 7 sets of three dithered exposures with a given IFU to (1) fill the gaps between fibres in the observational footprint, (2) improve signal, and (3) enable outlier rejection (Law et al., 2015). Raw fibre spectra as well as intermediate and final data products from the MaNGA Data Reduction Pipeline (DRP, Law et al. 2016) were released for 4,824 galaxies in DR15: including calibrated and sky-subtracted fibre spectra and spatially reconstructed data cubes. The testing performed in Section 4.4.1 make use of both the individual calibrated fibre spectra and the final data cubes.

4.3 Methods

This section describes the core functionalities of RealSim-IFS for reproducing the instrumental responses and corresponding covariances of fibre-based IFS instruments in data from hydrodynamical simulations. These core functionalities are: (1) generating fibre patterns and dithering strategies; (2) integrating the intensities in spatially discretized datacube spaxels within arbitrary fibres; and (3) the distribution of (potentially) irregularly sampled fibre intensity measurements onto regularly sampled Cartesian grids. Redshift and atmospheric seeing selection should generally precede these steps. An example redshift and atmospheric seeing selection strategy is outlined later in Section 4.4.2.

4.3.1 Instrumental designs and observing strategies

The designs of the observing units (IFUs) for modern fibre-based IFS instruments vary significantly. These designs are generally optimized to a set of science objectives and the boundaries set by (1) the precision to which new components can be fabricated, (2) existing components (e.g. telescopes and spectrographs), and (3) cost.

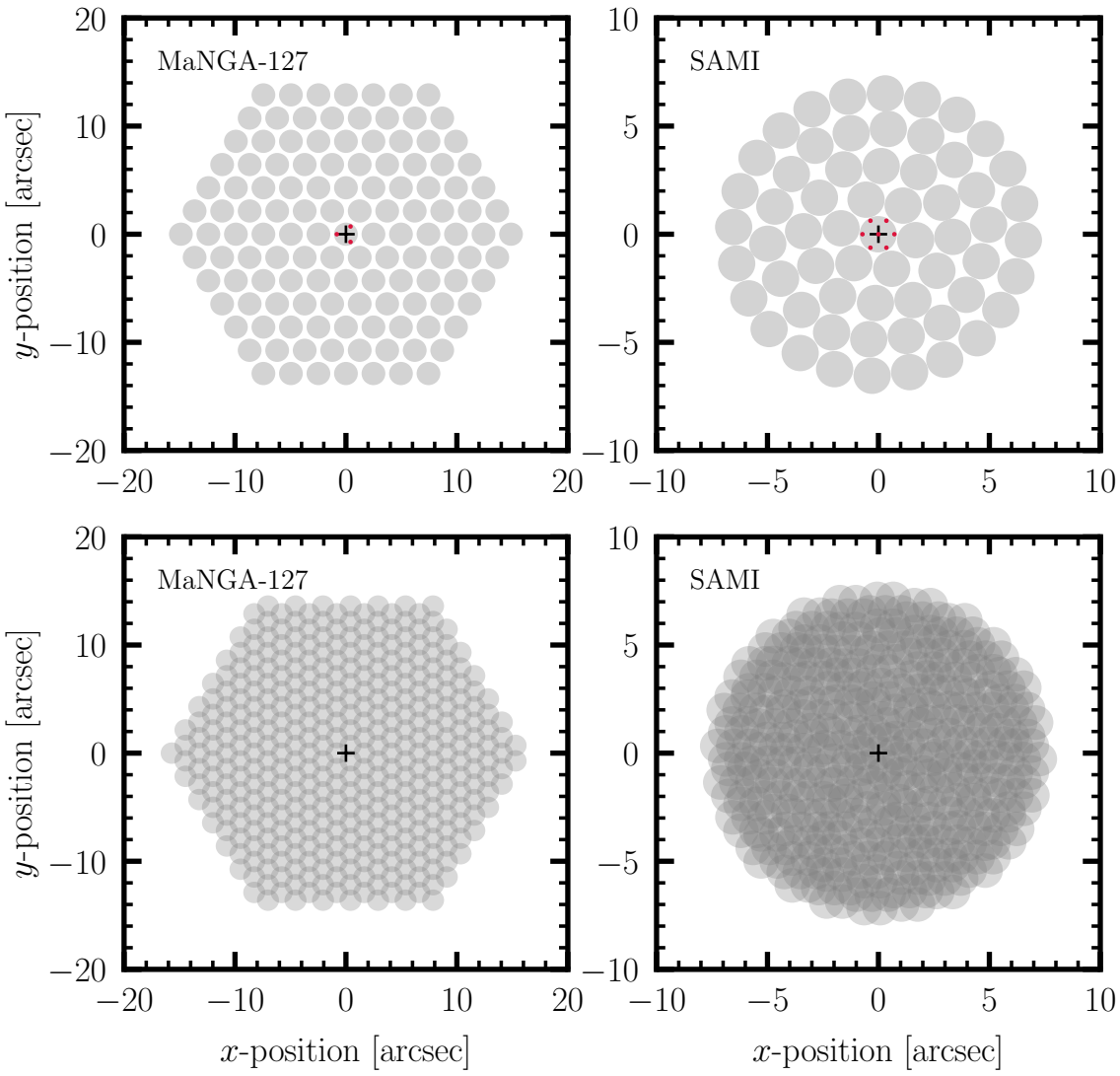

Figure 4.3 MaNGA (upper left) and SAMI (upper right) IFU designs and respective observing strategies (lower panels) emulated by RealSim-IFS. Individual fibre cores are marked by light-grey shaded circles. Black crosses mark the centre of the FOV (typically galaxy barycenter). Upper panels show the fibre pattern for a single exposure made by an IFU from each instrument. The IFU centroid positions of a complete set of dithered exposures are indicated by red markers in the upper panels – three for MaNGA, seven for SAMI. The dithered exposures are shown in the lower panels in which increased shading shows increased fibre overlap.

Furthermore, the observing strategies for these instruments also vary greatly. Consequently, RealSim-IFS must be sufficiently flexible that existing *and* anticipated fibre-based IFU designs and observing strategies can be emulated.

These requirements are accommodated by allowing a user to provide an arbitrary array of on-sky fibre positions and corresponding core diameters (in arcseconds). In general, these fibres are free to overlap – as it is assumed that consecutive observations of the same piece of sky are possible. It is up to the user to ensure that their synthetic IFU design and observing strategies are feasible. However, since the packing and positioning of fibres in real IFUs are physically limited by the diameters of fibre cores and their surrounding buffer and cladding, it is useful to adopt a single IFU design and perform repeated measurements with slight offsets (*dithering*). The purpose of dithering is to improve (1) the fill-factor of the FOV that is subtended by fibres and (2) the signal-to-noise in regions where fibres from consecutive observations overlap. Consequently, in addition to an arbitrarily long list of potentially overlapping fibres, a user can provide an array of fibre positions corresponding to a particular IFU design and the x and y offsets for each dithered exposure. RealSim-IFS also includes tailored modules which are hard-coded to emulated the current state-of-the-art fibre-based IFU designs and observing strategies of MaNGA and SAMI using information gathered from the corresponding technical papers (Drory et al., 2015; Law et al., 2015; Bryant et al., 2015; Sharp et al., 2015)[6]. These tailored modules allow a user to automatically set up SAMI and MaNGA IFU designs and strategies (including *all* current MaNGA IFU sizes).

Figure 4.3 shows MaNGA and SAMI IFU designs and observing strategies as set-up by RealSim-IFS. These designs are "idealized" in the sense that they do not incorporate random or systematic errors involved in the manufacture of the IFUs, for example. The black crosses mark the centre of the FOV (all panels) and the centre of the IFUs (upper panels). The upper left panel shows the MaNGA 127-fibre IFU design. Shaded light-grey circles show the individual fibre cores of the IFU: 2 arcseconds (120 μm) in diameter. Each MaNGA fibre core is surrounded by a buffer and cladding that that is 15 μm in thickness – bringing the total diameter of each fibre to 2.5 arcseconds (150 μm). The diameter of the full MaNGA 127-fibre IFU is around 32.5 arsceconds. The right panel shows the SAMI 61-fibre IFU design. SAMI fibre

[6]The fibre positions in all MaNGA IFU designs and dithering offsets were derived analytically using the information from the survey strategy paper (Law et al., 2015). The SAMI fibre bundle positions were generously provided by Nic Scott on behalf of the SAMI Survey team.

cores are 1.6 arcseconds (105 μm) in diameter with a buffer and cladding thinned to 0.076 arcseconds (5 μm) for denser packing. The diameter of the full SAMI 61-fibre IFU is around 14.7 arcseconds.

The MaNGA and SAMI surveys employ unique dithering strategies that are optimized for their respective IFU designs. The small red markers in the upper panels of Figure 4.3 show the centres of each dithered IFU exposure for a given MaNGA and SAMI target. For MaNGA (upper and lower left panels), the IFU in each of three dithered exposures is centred at the vertex of an equilateral triangle (red markers) whose centroid is the centre of the FOV (black cross, typically the target galaxy's barycentre). MaNGA observations consist of several repeats of this three-point dither pattern to increase signal-to-noise and enable outlier rejection. The SAMI survey (right panels) uses an optimized 7-point dither pattern comprising the vertices and centroid of a hexagon for which the radial distances between the centroid and vertices are 0.72 arcseconds (45% of the fibre core diameter, Sharp et al. 2015). Setting up complete on-sky fibre coordinate arrays (arcseconds) for these MaNGA and SAMI IFU designs (and all other IFU designs from MaNGA) and their dither strategies are one-line operations with RealSim-IFS.

With the on-sky positions and diameters of each fibre for a chosen observing strategy, fibre measurements of the synthetic datacubes can be made. In the next section, I describe the method used by RealSim-IFS to spatially integrate synthetic datacube intensities within fibres of arbitrary positions and apertures.

4.3.2 Fibre observations of spatially discretized data

The datacubes produced from the simulations are spatially discretized on a regular Cartesian grid. The non-spatial channels of these cubes can be, for example, wavelength or frequency bins from radiative transfer, or the velocity bins in a line-of-sight velocity distribution. In either case, to measure the intensities registered within an arbitrary fibre, an algorithm for spatially integrating datacube spaxels within the fibre aperture is needed.

Figure 4.4 illustrates the simple algorithm used by RealSim-IFS to integrate spatially discretized input datacube spaxels in arbitrary fibre apertures. The algorithm first isolates a rectangular set of spaxels around a given fibre to restrict weight calculations to a relevant subset of original spaxels. These spaxels are then further refined spatially by a factor which guarantees that there are at least 100 spatial ele-

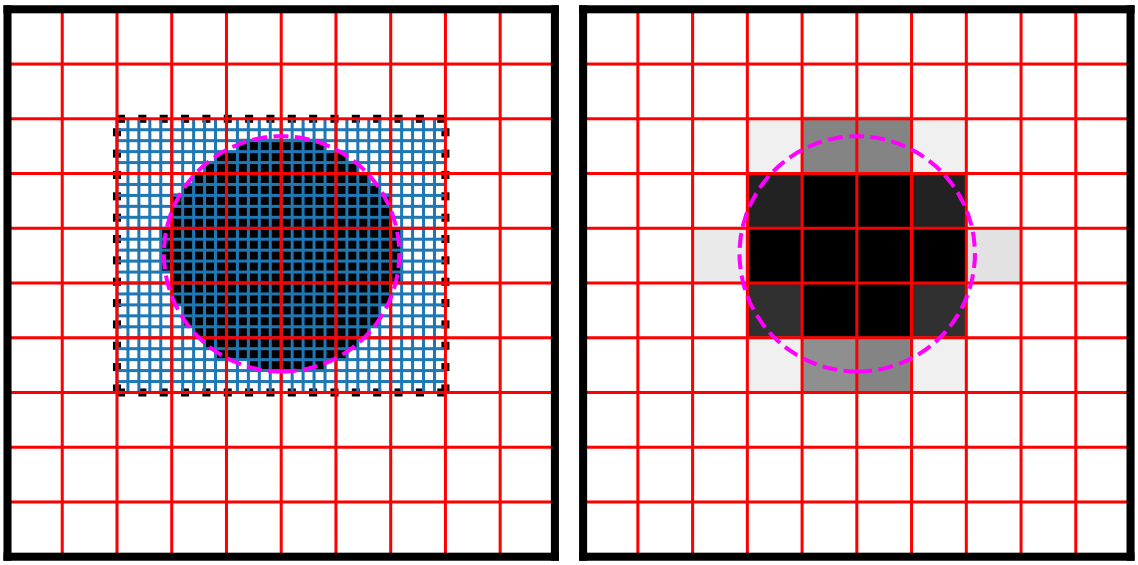

Figure 4.4 A simple algorithm is used to weight and integrate the contributions from spatially discretized datacube spaxels (red squares) within an arbitrary fibre (magenta). First, as illustrated in the left panel, a rectangular grid (black, dotted) is defined such that all spaxels which are completely or partially enclosed by the fibre aperture are within. The grid is then refined into sub-spaxels (blue squares) such that at least 100 sub-spaxels (or original spaxels) subtend the fibre diameter. For visibility, each original spaxel in the figure is refined simply by a factor of five. The fraction of sub-spaxels in each original spaxel whose centres are within the fibre aperture is taken as the weight of that original spaxel in the spatial integration. The right panel shows the resulting weight-map.

ments along the diameter of the fibre (for visibility, the original spaxels in Figure 4.4 are refined simply by a factor of five). The fraction of sub-spaxels of a given original spaxel whose centres are within the fibre aperture is taken as the weight of that original spaxel in the spatial integration. The resulting weight-map, $w[i, j]$ (right panel), is then applied in a weighted sum to each wavelength, frequency, or velocity channel, $f[i, j]$, to produce the intensity registered by the fibre in that channel, F:

$$F = \sum_{i=1}^{N_y} \sum_{j=1}^{N_x} w[i, j]\ f[i, j] \tag{4.3}$$

The outcome after applying the fibre's weight-map to each channel is then a one-dimensional fibre array of length $N_{\rm ch}$, the number of channels. After obtaining $N_{\rm fib}$ fibre measurements of the data, the result is a two-dimensional $N_{\rm fib} \times N_{\rm ch}$ matrix in which each row corresponds to a single fibre's one-dimensional spectrum or LOSVD. If the non-spatial axis is wavelength, then the data resulting from this step are analogous to the row-stacked spectra (RSS) data products from observational IFS surveys (e.g. see Section 4.4.1). Conveniently, the algorithm used here is an inversion of the "drizzle" algorithm (Fruchter & Hook, 2002) used by various fibre-based IFS instruments to distribute spatially irregular intensity measurements onto regular Cartesian grids (e.g. CALIFA: Sánchez et al. 2012, SAMI: Sharp et al. 2015). Elements of this same algorithm are incorporated into the drizzling modules of RealSim-IFS that are specific to instruments which use this data reduction strategy.

4.3.3 Spatial reconstruction of fibre measurements

With the stacked measurements for each fibre, the final step for RealSim-IFS is to spatially redistribute these measurements onto a regular Cartesian grid of output spaxels. Mirroring the variety in IFU designs and dithering strategies, there are several algorithms that can be employed to spatially reconstruct the flux from non-Cartesian sampled fibre measurements.

First, there is the simple inverse of the algorithm employed by RealSim-IFS to make fibre measurements as detailed in Section 4.3.2. This method is adopted by the CALIFA and SAMI surveys. The SAMI pipeline uses a substantial modification to the standard drizzling approach in which the intensities from a given fibre are redistributed within spaxels in an aperture that is half the size of the actual fibre core. This strategy, afforded by SAMI's large number of dithered exposures, enables im-

proved preservation of spatial resolution (Sharp et al., 2015). Another is the modified (flux-conserving) Shepard approach adopted by the MaNGA data reduction pipeline (Law et al., 2016). RealSim-IFS provides both the drizzling and modified Shepard algorithms as options for spatial reconstruction of the fibre measurements and each is described below.

Drizzling algorithm

The intensity distributed to a given output spaxel (for a given channel: wavelength, frequency, or velocity) by a given fibre is determined through a spatial weighting scheme as illustrated in Figure 4.4. Spaxels completely contained within the k^{th} fibre's aperture are assigned a non-normalized weight, $w[k,i,j] = 1$. The non-normalized weights on spaxels that partially overlap with the aperture are computed from their fractional overlapping area with the aperture. The weight assigned to each $[i,j]$ spaxel from each $[k]$ fibre is then:

$$W[k,i,j] = \alpha[k]\frac{w[k,i,j]}{\sum_{k=1}^{N_{\text{fib}}} w[k,i,j]} \quad \text{where} \quad \alpha^{-1}[k] = \sum_{i=1}^{N_y}\sum_{j=1}^{N_x} w[k,i,j] \tag{4.4}$$

Each fibre core on the IFU (which may or may not be of the same radius, r_{core}) has an associated flux scaling factor, α, equal to the area, in output spaxels, subtended by the fibre core, $\alpha^{-1}[k] = \pi r^2_{\text{core}}[k]$. This combination of fibre normalization and flux scaling guarantee that calibrated flux is conserved within the limits of the spatial sampling. With these normalizations, the calibrated flux in a given output channel, $I[i,j]$, is:

$$I[i,j] = \sum_{k=1}^{N_{\text{fib}}} W[k,i,j]\ F[k] \tag{4.5}$$

Alternatively, to conserve intensity only, such that the sum of the flux in an output channel is equal to the sum of the flux in that channel from each fibre, the fibre normalization term can be dropped in Eq. 4.4. RealSim-IFS provides pre-normalization and pre-scaling weights in addition to the output cubes to make it straight-forward to conserve total integrated intensity rather than calibrated flux. In either case, the weight maps are valuable for computing variances and covariances of output spaxels.

The SAMI survey uses a slight modification of the drizzling algorithm in which the

"drop radius" of each fibre's flux on the output grid is 50% of the original fibre size (see Sharp et al. 2015). This adaptation is easily implemented in REALSIM-IFS. First, the size of the fibre aperture in the output grid (optionally computed by REALSIM-IFS's built-in coordinate-transformation module) can be set manually to the desired drop radius. Second, provided all fibre cores have the same diameter, flux conservation can be achieved by multiplying the output cube by the factor, $\zeta^2 = (r_{\rm drop}/r_{\rm core})^2$, which derives from the relative areas of the drop aperture an true core aperture. If fibres are not all of the same diameter, ζ is specific to each fibre and must be integrated into the weights. In either case, ζ must always be incorporated into the variance maps.

Modified Shepard algorithm

The modified Shepard algorithm is used in the MaNGA data reduction pipeline (DRP, specifically, see Law et al. 2015, Section 9.1). The non-normalized flux contribution from each fibre to each regularly spaced output spaxel is mapped with a two-dimensional Gaussian distribution:

$$w[k,i,j] = q[k,i,j] \exp\left(\frac{-r^2[k,i,j]}{2\sigma^2}\right) \tag{4.6}$$

where $r[k,i,j]$ is the radial distance from the $k^{\rm th}$ fibre core centroid to the centroid of output spaxel $[i,j]$ and σ is an adopted standard deviation for the Gaussian distribution (taken to be 0.7 arcseconds or 1.4 output grid spaxels in the MaNGA DRP). Because the Gaussian distribution is non-zero to infinite radii, a characteristic truncation radius, $r_{\rm lim}$, can be defined beyond which $w[k,i,j] = 0$. This truncation is represented by the $q[k,i,j]$ term – which is a binary mask equal to 1 where $r[k,i,j] \leq r_{\rm lim}$ and 0 beyond. In MaNGA, $r_{\rm lim}$ is taken to be 1.6 arcseconds or 3.2 output spaxels. The effect of $r_{\rm lim}$ on the correlation coefficients between spaxels is nicely illustrated by (Westfall et al., 2019) (see their Figure 7).

These weights are necessarily normalized to conserve flux. This normalization follows Eq. 4.4 using the initial non-normalized Gaussian weights, $w[k,i,j]$. However, in this case, the non-normalized Gaussian weights do not sum to the fibre area in output spaxels. Therefore, the α term in Eq. 4.4 can be computed simply as $\alpha^{-1}[k] = \pi r^2_{\rm core}[k]$ where $r_{\rm core}[k]$ is the radius, in output spaxels, of the $k^{\rm th}$ fibre core. However, as with any algorithm for redistributing fibre fluxes onto Cartesian grids, perfect flux conservation is only guaranteed with very high (and extended) spatial sampling of the original data by the fibres.

RealSim-IFS also includes a number of other functions aimed at setting up synthetic observations (e.g. redshift selection and atmospheric seeing selection). These are demonstrated in Section 4.4.2.

4.4 Test results and demonstrations

This section comprises several demonstrations with RealSim-IFS using the MaNGA and TNG100 data outlined in Section 4.2. First, I demonstrate the precision of RealSim-IFS in emulating the reconstruction component of the MaNGA DRP using real MaNGA data (Section 4.4.1) and highlight some of the higher-order limitations of RealSim-IFS. Then, using similar targeting selection to MaNGA, I show the early stages of a synthetic MaNGA survey of TNG100 galaxies and provide visual comparisons of the output and input kinematic structures of mergers and non-mergers (Section 4.4.2).

4.4.1 Precise emulation of MaNGA data reduction

The precision with which RealSim-IFS can reproduce data reduction of real fibre measurements can be tested explicitly using real observations. IFS surveys typically generate two main data products: (1) row-stacked spectra (RSS) files and (2) output cubes. The RSS files contain the calibrated spectrum from each fibre core and a wealth of ancillary data (including astrometric registration of the fibre cores). The RSS files are therefore an important intermediate data product between the raw spectra and the output cubes. Figure 4.5 shows an infographic for a MaNGA RSS file.

The RSS files provide the information necessary to spatially reconstruct the IFU fibre measurements onto a regular Cartesian grid. In the MaNGA DRP (at least), the output cubes (linearly spaced in wavelength) are the direct results of the spatial reconstruction of the spectra in the RSS files. Consequently, one straight-forward test is to examine whether RealSim-IFS reproduces this spatial reconstruction component using real fibre data from the RSS files. I carry out this test with MaNGA Plate-IFU 7443-12703 (the same target as from Figure 4.5). MaNGA Plate-IFU 7443-12703 was chosen for testing completely arbitrarily from the completed 127-fibre IFU MaNGA targets and is conveniently a visually stunning interacting pair that belongs to the (Fu et al., 2018) MaNGA galaxy pair and binary AGN samples.

Figure 4.6 compares the spatially reconstructed cube generated by RealSim-

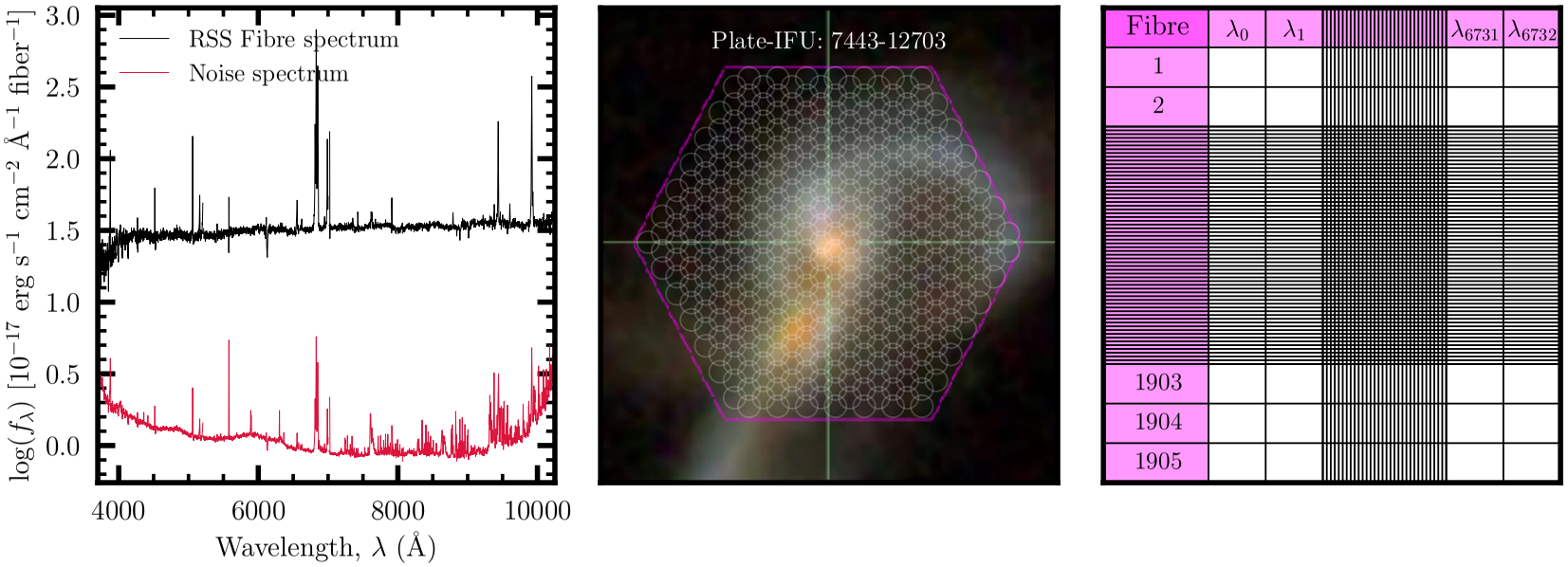

Figure 4.5 Row stacked spectra. The central panel shows the *gri* image of MaNGA Plate-IFU target 7443-12703 overlaid with three dithered 127-fibre IFUs. Five repeated observations with this three-point dither pattern were used for a total of $5 \times 3 \times 127 = 1905$ fibre spectra. The left panel shows one such spectrum (from a fibre near the centre of the IFU) and its corresponding noise spectrum generated by the MaNGA DRP. The right panel shows one of the data structures contained within the RSS file: the calibrated spectrum from each fibre core.

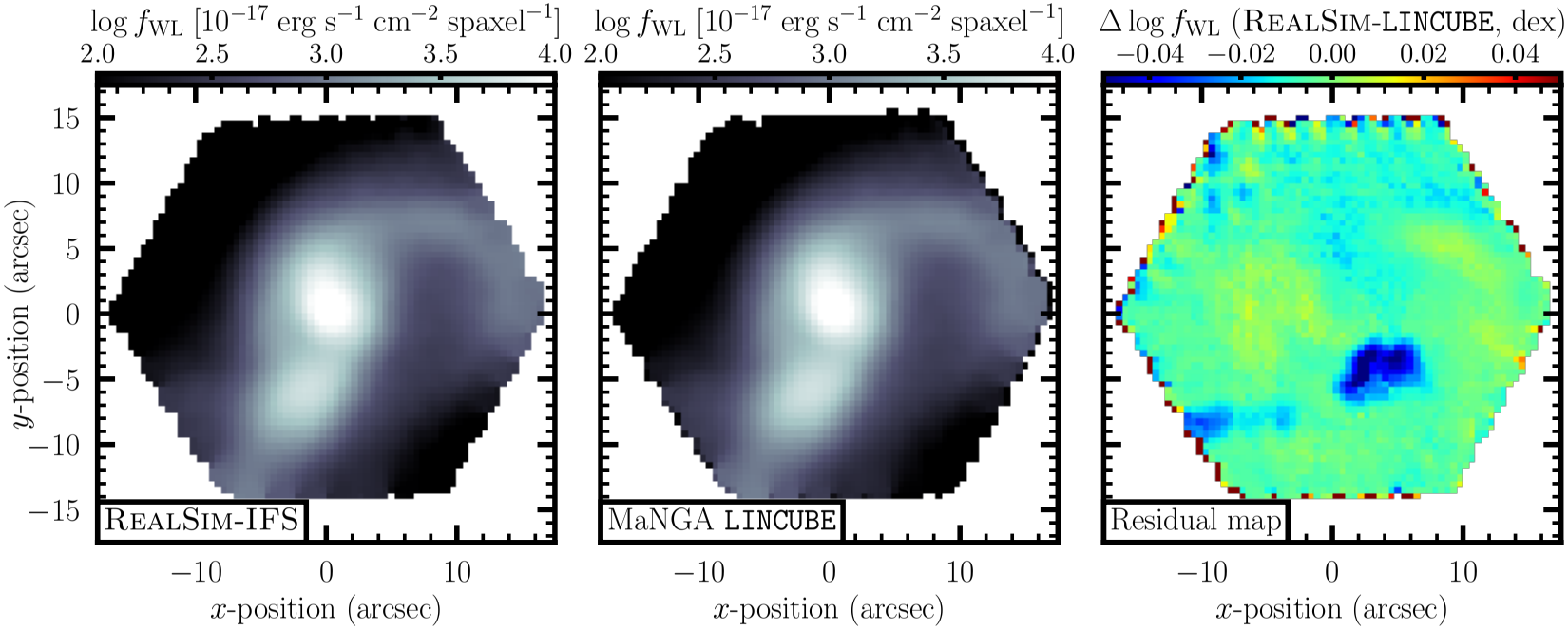

Figure 4.6 Comparison of RealSim-IFS and MaNGA spatial reconstructions of a MaNGA RSS file. $f_{\rm WL}$ is the integrated white light flux in each spaxel computed following 4.7. The left and middle panels show the RealSim-IFS and MaNGA DRP reconstructions, respectively. The right panel shows the residual map (RealSim-IFS−LINCUBE). The offsets are largely well-behaved apart from some faint, low-level structure in the offsets throughout the map which reach maxima on the order of 10% and a large blue cavity.

IFS from the MaNGA RSS file with the corresponding MaNGA datacube (`LINCUBE`) for that Plate-IFU target. The left panel of Figure 4.6 shows the RealSim-IFS reconstruction of flux integrated overall wavelengths in each spaxel (integrated white light) from the RSS files:

$$f_{\mathrm{WL}} = \sum_{\lambda_1 \to \lambda_N} \Delta\lambda \; f_\lambda \tag{4.7}$$

where $\Delta\lambda$ is the width of each wavelength channel. In MaNGA, $\Delta\lambda = 1$ Å in the linear cubes. The middle panel shows the same quantity from the corresponding MaNGA cube. While the visual similarity is impressive, the quantitative comparison in the right panel reveals some higher-order errors.

The right panel of Figure 4.6 shows the residual in $\log f_{\mathrm{WL}}$, where $\log f_{\mathrm{WL}}$ from the MaNGA `LINCUBE` is subtracted from the $\log f_{\mathrm{WL}}$ from RealSim-IFS. The extremes of the colour map are -0.05 and 0.05 dex ($\pm 12\%$ offset). Most of the residual map shows near-zero offsets (green) within 0.02 dex ($\pm 5\%$). However, there is some visible structure in these low-level offsets shown as faint arcs of cyan and yellow. Furthermore, a large blue blotch highlights a local cavity in RealSim-IFS reconstructed flux compared to the MaNGA cube.

First, low-level structured offsets with equal parts positive and negative are often a good indicator of astrometric registration issues. This brings us to the first important caveat when using RealSim-IFS in its current state: spatial reconstruction is not, by default, performed uniquely in each channel. In other words, RealSim-IFS assumes that there is no change in the effective positions of fibres with channel. This is a limitation that is specific to processing *real data* due to the astrometric fibre position corrections and non-negligible effect of differential atmospheric refraction (DAR, e.g. Filippenko 1982) in ground-based observations when objects are not at the zenith.

DAR is generated by the increased refractive index of the atmosphere for shorter wavelengths than longer wavelengths – an effect which is exacerbated at larger airmasses (i.e. the greater the angle between the target and the zenith, the greater the effects of DAR). If the airmass is known, then DAR can be reasonably accounted for in the effective astrometric positions of the fibres in each wavelength channel. Figure 4.7 shows the effects of DAR (right panel) and other systematic and random errors on effective fibre positions. The left panel shows the idealized, analytically generated fibre positions for a single 127-fibre IFU as light-grey shaded circles (produced by RealSim-IFS). Overlaid as black unfilled cirlces are the median positions

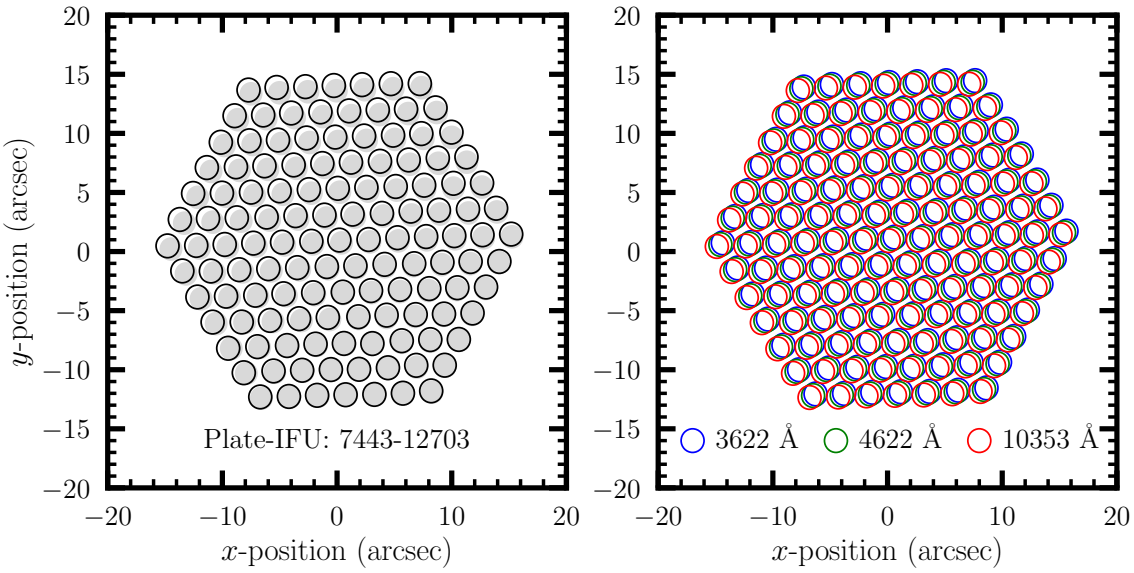

Figure 4.7 The effect of manufacture precision, differential atmospheric refraction (DAR), and other astrometric corrections on the wavelength-dependent departure from an idealized IFU fibre pattern. The left panel shows the astrometric offsets between the idealized IFUs (light-grey filled circles) and the median effective fibre positions over all wavelength channels for a single MaNGA 7443-12703 exposure. The right panel shows the effect of DAR on the effective positions of fibres as a function of wavelength. RealSim-IFS assumes that fibres do not change in effective position with channel. A wrapper-code can be written to run RealSim-IFS channel-by-channel with unique fibre positions in each channel. However, the data reduction pipelines of observational IFS campaigns generally correct for these effects with precision of ~ 0.1 arcseconds – a precision error that is dwarfed by the $2-2.5$ arcsecond effective spatial resolution of SAMI and MaNGA after spatial reconstruction. Therefore, synthetic cubes generated with RealSim-IFS are still highly comparable to real cubes (as illustrated in Figure 4.6) despite RealSim-IFS's default imprecision in the astrometric treatment of real data.

of real fibres estimated over all channels for a single exposure from MaNGA Plate-IFU 7443-12703. The small offsets between the idealized and measured fibre positions are predominantly due to (1) the random and systematic errors in IFU manufacture, (2) DAR, and (3) the error in astrometric corrections of (1) and (2). The right panel illustrates the specific the role of DAR in the effective fibre positions of this IFU exposure as a function of wavelength. Note the shift in the blue circles and (to a lesser extent) green circles with respect to the red. Taken together, these effects result in small astrometric uncertainties on the order of 0.1 arcseconds.

The MaNGA and SAMI DRPs perform the spatial reconstruction of fibre fluxes independently in each channel using the effective, astrometrically corrected fibre positions corresponding to each channel. Because RealSim-IFS assumes, by default, no change in fibre position with wavelength, the median fibre positions over all wavelength channels were adopted for the comparison in Figure 4.6. The astrometric registration errors that arise from this bias in the RealSim-IFS reconstructed cube for MaNGA Plate-IFU 7443-12703 are likely culprits for the asymmetric flux offsets in the residual map. Fortunately, these astrometric effects are accounted and corrected to precisions on the order of ~ 0.1 arcseconds by the SAMI and MaNGA DRPs, for example (Sharp et al., 2015; Law et al., 2016). Furthermore, this small precision error (though exacerbated in a direct comparison of reconstructed cubes as in the residual map of Figure 4.6) is dwarfed by the $2 - 2.5$ arcsecond effective spatial resolution of these IFS surveys. Consequently, as long as RealSim-IFS is used for its intended purpose of generating survey-realistic *synthetic* observations, one can expect data products that are highly comparable to the real IFS data in which astrometry effects are appropriately accounted and corrected.

Nevertheless, a simple wrapper for RealSim-IFS can be used to reconstruct the flux independently in each channel by using the corresponding unique effective fibre positions[7]. This is demonstrated in the comparison of three individually reconstructed channels of the 7443-12703 target in Figure 4.8. In this case, I have also used the inverse variance maps from the RSS files to mask data with arbitrarily high variance (`IVAR`$= 0$). The low-level structure in the flux offsets from Figure 4.6 disappear entirely in Figure 4.8 owing to the channel-by-channel spatial reconstruction. Furthermore, the sets of spaxels which have defined fluxes (non-NaN) are also exactly the same in each channel. The larger blue blotches from Figure 4.6 also disappear in the residual maps of Figure 4.8 due to the inverse variance masking. The masking

[7]Such a wrapper is provided in the RealSim-IFS example notebooks.

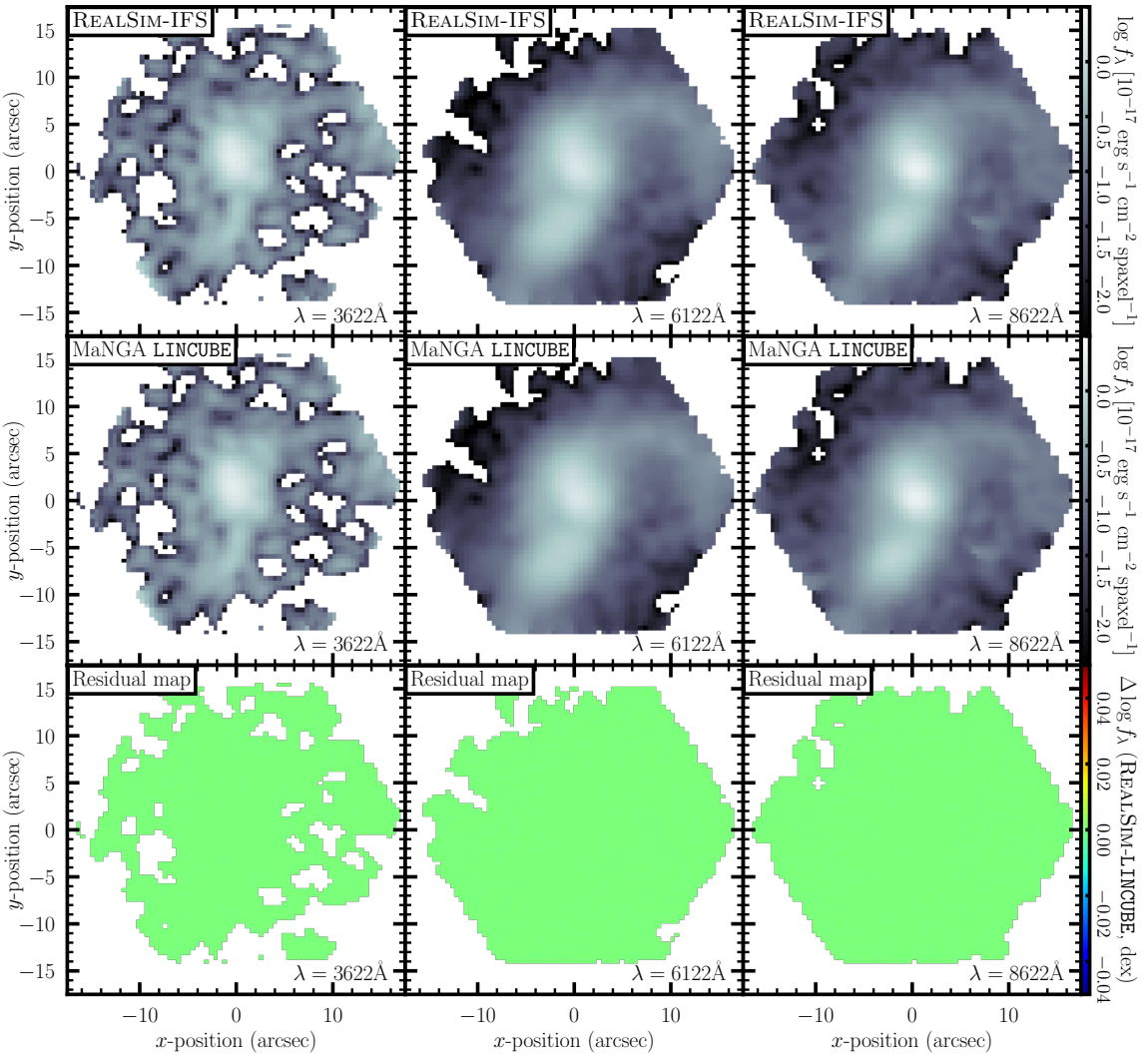

Figure 4.8 Similar to Figure 4.6 but showing the reconstruction results for the MaNGA Plate-IFU 7443-12703 cube using the unique RSS astrometry info in each channel (which include DAR and other astrometry effects) and masking where the inverse variance reported in the RSS files is zero. Each row corresponds to a single wavelength channel of the RealSim-IFS and `LINCUBE` output cubes and their corresponding residual maps (RealSim-IFS−`LINCUBE`). The maps are now perfectly reproduced due to the incorporation of the corrected effective fibre positions and bad-data masking in each channel. The non-NaN spaxel masks are also identical in the RealSim-IFS and MaNGA DRP reconstructed cubes.

of data with exceptionally high variance eliminates the signatures of artifacts such as cosmic rays, dead fibre elements in the IFUs, and dead CCD pixels in the spectrographs. Quantitatively, RealSim-IFS reconstructs the specific fluxes for MaNGA Plate-IFU 7443-12703 with an exceptional median precision of $\Delta \log f_\lambda \lesssim 10^{-8}$ dex in each channel.

In summary, RealSim-IFS can be used to precisely reproduce the data reduction pipelines of modern IFS surveys. Because of this, a high level of realism in synthetic flux and kinematic observations produced with RealSim-IFS can be expected – including the distribution of intensity, variances, and spatial covariances. The successes of RealSim-IFS in the tests performed in this section provide a strong foundation for using RealSim-IFS for its intended purpose: creation of survey-realistic *synthetic* IFS observations.

4.4.2 Synthetic MaNGA kinematic observations

In this section, I demonstrate the application of RealSim-IFS to generating synthetic IFS observations of galaxies from the IllustrisTNG simulations with MaNGA realism. The first steps in this process concern the emulation of selection strategies to roughly match the survey statistics of MaNGA (e.g. atmospheric seeing conditions, redshift selection, and the assignment of particular IFU sizes corresponding to each target).

Redshift selection

MaNGA galaxies are selected from the SDSS main galaxy sample (Strauss et al., 2002) following a strategy outlined by Wake et al. (2017). The MaNGA survey is divided into Primary (47%), Secondary (37%), and Colour-Enhanced supplement (16%) samples. Figure 4.9 shows the MaNGA survey galaxy selection and samples in the plane of redshift and absolute $i-$band magnitude. Galaxies in the Primary sample (orange) are selected such that they are covered out to 1.5 effective radii, $R_{\rm eff}$, by the observational footprint of one of the five MaNGA science IFU designs, here denoted: N19, N37, N61, N91, N127. The diameters of observational footprints of these IFUs are: 12.5, 17.5, 22.5, 27.5, and 32.5 arcseconds, respectively, after dithering. The Colour-Enhanced sample (black) is a supplement to the Primary sample that includes galaxies from poorly sampled regions of colour-magnitude plane including low-luminosity red galaxies, high-luminosity blue galaxies, and green valley

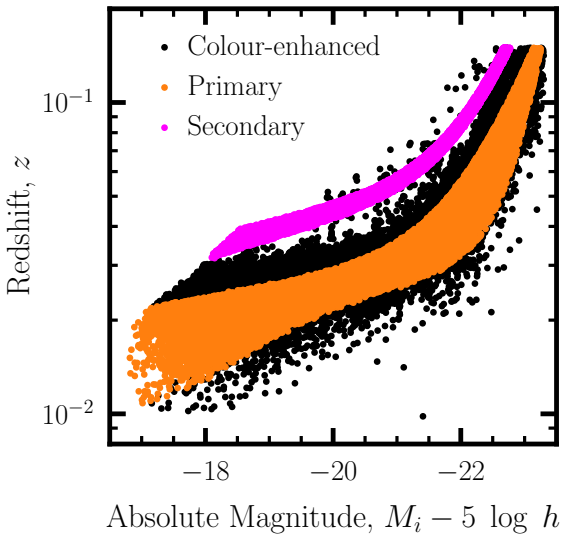

Figure 4.9 MaNGA survey galaxy selection and samples in the plane of redshift, z, and absolute $i-$band magnitude, M_i. Coloured points show individual candidate MaNGA targets from the Primary (orange), Secondary (magenta), and Colour-Enhanced (black) samples of the MaNGA survey.

galaxies. IFU design allocation for the Colour-Enhanced sample is the same as for the Primary sample – covering galaxies out to 1.5 $R_{\rm eff}$. The Secondary sample covers galaxies in the same luminosity range as for the Primary sample but at slightly higher redshifts. Galaxies in the Secondary samples are assigned IFU designs such that they receive greater coverage, out to 2.5 $R_{\rm eff}$.

Figure 4.10 shows the one-dimensional redshift distributions of each sample. In the demonstrations which follow, I marginalize over the luminosity axis of Figure 4.9 by selecting multiple redshifts for each TNG100-1 galaxy (regardless of its luminosity). Specifically, for each TNG100-1 galaxy, I randomly draw a redshift from the empirical probability density functions for each sample in Figure 4.10. In addition, I draw a redshift from the distribution defined by all galaxies in all samples. The result is four redshifts for each TNG100-1 galaxy: one for each of the Primary (PRI), Secondary (SEC), and Colour-Enhanced (CEN) samples and one drawn from their combined redshift distribution (ALL). By making synthetic observations at all of these redshifts (with the specific IFU allocations corresponding to each sample), the luminosity-selection and matching to the percentages in each sample can be performed later. This final step is not performed here as it is unnecessary for the demonstrations with RealSim-IFS. The primary objective of the following demonstrations is to show that

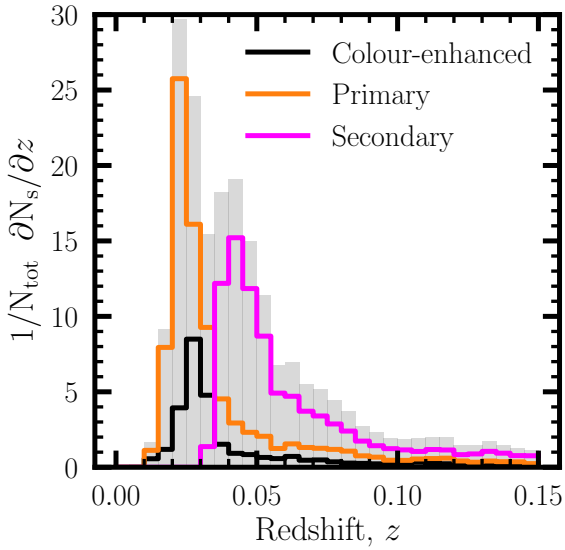

Figure 4.10 One-dimensional redshift histograms for each of the the Primary (orange), Secondary (magenta), and Colour-Enhanced (black) samples, $\partial N_s/\partial z$. The histograms are normalized by the total number of galaxies in all samples, N_{tot}, to show the relative contribution of each sample in each redshift bin.

RealSim-IFS can be applied to a simulated galaxy to produce synthetic observations in any of the MaNGA samples and with any IFU design.

Atmospheric seeing selection

Radiation from real galaxies is spatially broadened by the atmospheric seeing before any interaction with ground-based instruments. Consequently, selection and convolution with the atmospheric seeing directly follows redshift selection. Given a redshift for a particular synthetic observation, the physical FOV of the corresponding idealized synthetic LOSVD datacube is converted to an angular size for the chosen cosmology. Then, an atmospheric PSF is drawn randomly from the distribution of atmospheric seeing FWHM (arcseconds) estimates for real MaNGA observations using the `drpall` tables[8]. These atmospheric seeing values are *effective* estimates quoted at around 5000 Å and average around 1.5 arcseconds for MaNGA observations so far. The true atmospheric seeing has a weak $\lambda^{-0.2}$ dependence as measured by MaNGA.

[8]The seeing values used here derive from the MaNGA `drpall` tables which include minimum, maximum, and median guide-star atmospheric seeing estimates for every combined set of exposures. I adopt the median seeing FWHM over all exposures from these tables. Data model for `drpall` tables: https://data.sdss.org/datamodel/files/MANGA_SPECTRO_REDUX/DRPVER/drpall.html.

Since the data products used here are LOSVD cubes, the wavelength dependence in the atmospheric seeing is neglected and the seeing kernel is constructed from the effective values quoted in the `drpall` tables. In MaNGA, the guide-star seeing is empirically modelled as a combination of two circular Gaussian distributions (private communication with David Law):

$$k(r) = \frac{9}{13}G(r, \sigma_1) + \frac{4}{13}G(r, \sigma_2), \quad \text{where} \quad G(r, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(\frac{-r^2}{2\sigma_i^2}\right) \tag{4.8}$$

and for which:

$$\sigma_1 = \frac{\mathrm{FWHM}_{\mathrm{drpall}}}{1.05 \times 2\sqrt{2\ln 2}} \quad \text{and} \quad \sigma_2 = 2\sigma_1 \tag{4.9}$$

Therefore, given the quasi-randomly selected MaNGA redshift, z, and seeing, $\mathrm{FWHM}_{\mathrm{drpall}}$, for the synthetic observation, the atmospheric seeing kernel is constructed following Eqs. 4.8 and 4.9 and convolved with every velocity channel of the idealized LOSVD cube.

Assignment of IFU designs and observation

An IFU design is assigned to each synthetic observation based on: (1) the galaxy's angular effective radius at the target redshift and (2) the target sample (PRI, SEC, CEN, ALL), an approach that is similar to the MaNGA IFU assignment. However, I neglect the observational limitation that comes from the finite number of IFUs of each design (see Wake et al. 2017 for an exact description of the MaNGA IFU design selection strategy on a per-plate basis). For each "observation", the IFU design with the smallest difference between its own observational footprint diameter and $2 \times N_{R_{\mathrm{eff}}} \times R_{\mathrm{eff}}$ is chosen, where $N_{R_{\mathrm{eff}}}$ is 1.5 for the PRI, CEN, and ALL samples and 2.5 for the SEC sample. The stellar half-mass radius of the galaxy is adopted as R_{eff} for the TNG100-1 galaxies. Galaxies for which $2 \times N_{R_{\mathrm{eff}}} \times R_{\mathrm{eff}}$ is larger than the N127 design footprint are assigned the N127 design. Similarly, galaxies for which $2 \times N_{R_{\mathrm{eff}}} \times R_{\mathrm{eff}}$ is smaller than the N19 design are assigned the N19 design footprint.

The fibre measurements follow the method outlined in Section 4.3.2 and are distributed onto a fixed (100, 100) spaxel grid with 0.5 arcseconds spaxel^{-1} resolution regardless of IFU design (50 arcseconds FOV). The redistribution of flux follows the modified Shepard approach described in Section 4.3.3 used by the MaNGA DRP (Law et al., 2016). Moments are derived from the resulting cubes by assuming the LOSVDs

in each spaxel are described by single-component Gaussian distributions as in Figure 4.2.

Synthetic MaNGA kinematics

Figure 4.11 breaks down the production of synthetic MaNGA kinematics with RealSim-IFS for the large spiral galaxy from the upper row of Figure 4.2. The top row shows the idealized kinematic moment maps before any realism (identical to Figure 4.2). The middle row shows the moment maps derived from the atmospheric seeing-convolved cubes – following redshift selection for the Primary (PRI) sample and seeing selection. The sample selection, redshift, seeing, and IFU design are indicated in the middle left panel. The three-point dither observational footprint of the IFU design (N127) assigned to this PRI sample galaxy at the target redshift ($z = 0.0374$) is overlaid in each of the middle panels.

The lower row of panels in Figure 4.11 show the synthetic MaNGA kinematic moments derived from the cube produced by RealSim-IFS. The observational $N_{R_{\rm eff}} = R_{\rm IFU}/R_{\rm eff} = 1.16$ is shown in the lower left panel. Since this galaxy is part of the PRI sample, the target coverage was $N_{R_{\rm eff}} = 1.5$. However, there is no larger IFU design than N127 and so this IFU design was assigned following Section 4.4.2. Figure 4.11 shows that regular kinematic structure is reasonably well-preserved in the synthetic MaNGA maps. In particular, both the high-dispersion bulge and kinematically cold disc of this galaxy are visible in the final velocity dispersion map. However, the finer spatial structures in the kinematic moments ($\lesssim 2.5$ arcseconds) are completely washed out by the effective PSF (combined atmospheric and instrumental response).

Figure 4.12 shows the same breakdown for a synthetic MaNGA observation of the recent post-merger galaxy from the third row of Figure 4.2. This galaxy underwent a merger of mass ratio $\mu = 0.65$ within the last simulation snapshot ($t_{\rm pm} \leq 162$ Myr). As noted in Figure 4.2, the signature of the merger is imprinted on both the idealized morphology and kinematics of this galaxy. The morphology of the galaxy includes tails or plumes extending from its main body. The kinematics are heavily disturbed. These highly irregular and asymmetric kinematics are impressively well-preserved in the synthetic MaNGA kinematic maps. In particular, this result highlights the viability of exploiting the spatially resolved kinematics of galaxies for visual and automatic merger and post-merger classification.

Figure 4.13 shows the breakdown for a SEC sample observation for the post-

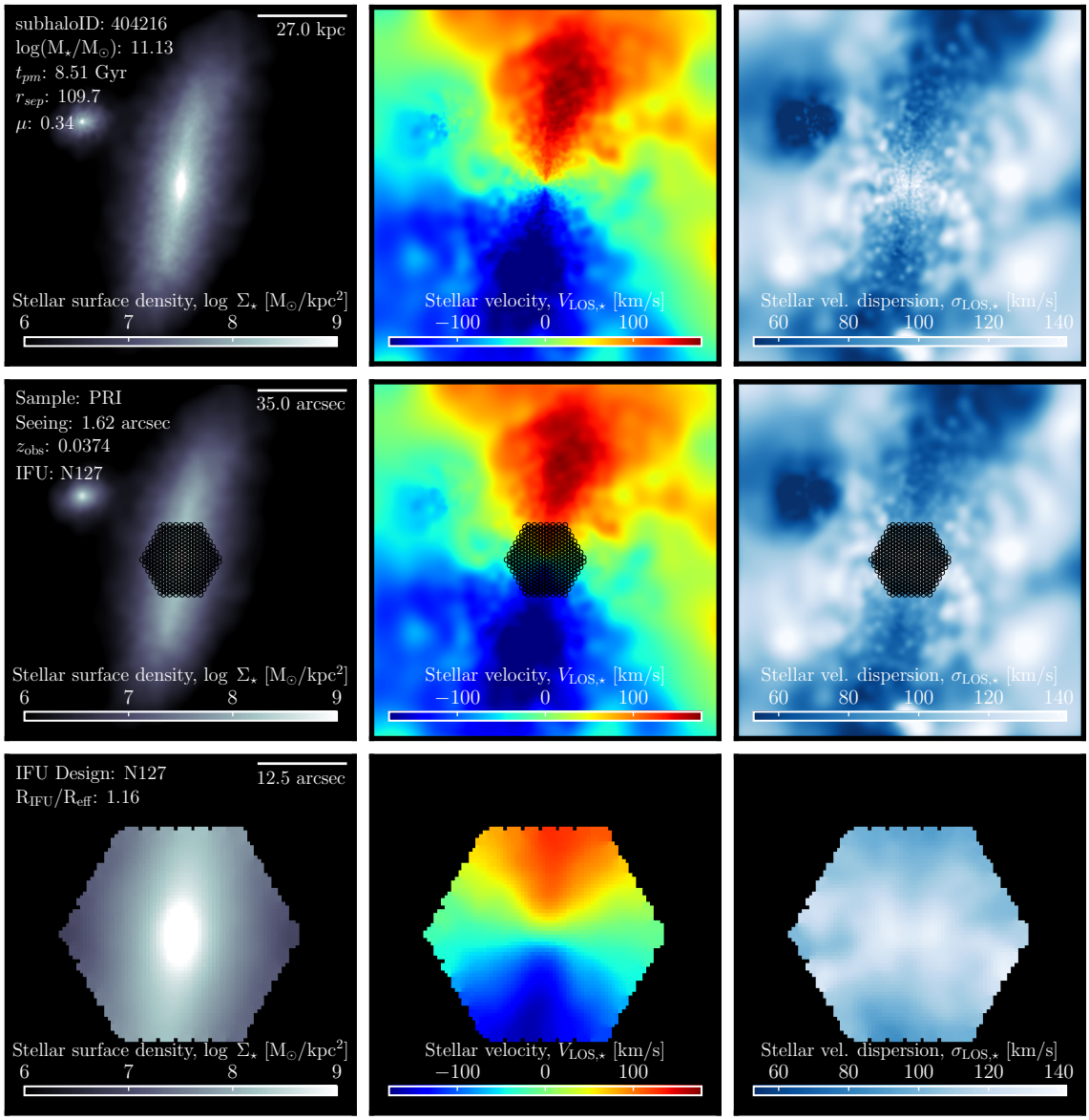

Figure 4.11 Idealized (upper row), redshift-selected and seeing convolved (middle row), and MaNGA-realistic (lower row) levels of realism for TNG100-1 galaxy ID: 404216. As in Figure 4.2, the horizontal scale in the upper left and middle left panels measure $2.5 \times R_{\rm eff}$ in physical and angular units, respectively. The middle panel shows the moments derived from the idealized cubes after convolution with the atmospheric seeing. The redshift and seeing FWHM (arcseconds) are indicated in the middle left panel along with the sample (Primary, PRI). The three-point dither pattern of the corresponding IFU design (N127) is overlaid on the the moment maps. The lower row shows the RealSim-IFS synthetic MaNGA observation with the N127 IFU design. The observational coverage of the galaxy by the IFU is also shown, $N_{R_{\rm eff}} = R_{\rm IFU}/R_{\rm eff}$. Given that this observation is for the PRI sample, it falls short of the target coverage of $N_{R_{\rm eff}} = 1.5$ – but there is no larger MaNGA IFU design.

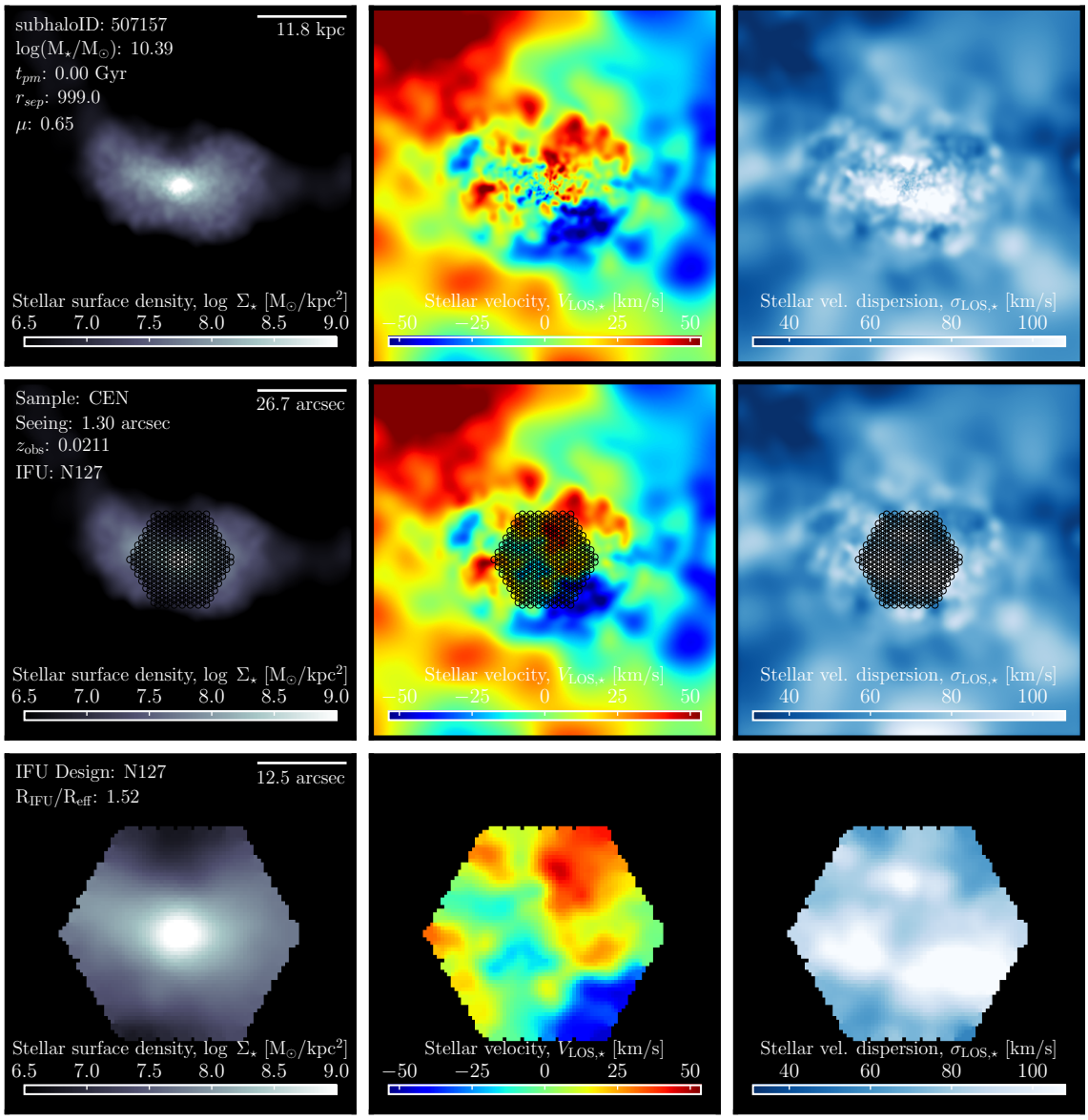

Figure 4.12 Similar to Figure 4.11 but for galaxy ID: 507157 observed for the CEN sample. The galaxy is the recent post-merger from the third row of Figure 4.2 ($t_{\rm pm} = 0$ Gyr, $\mu = 0.65$). The merger drives highly irregular kinematic structure that is visible in the idealized and synthetic MaNGA kinematic maps – both in velocity and velocity dispersion. This result highlights the potential of survey-realistic synthetic kinematic observations for merger classification.

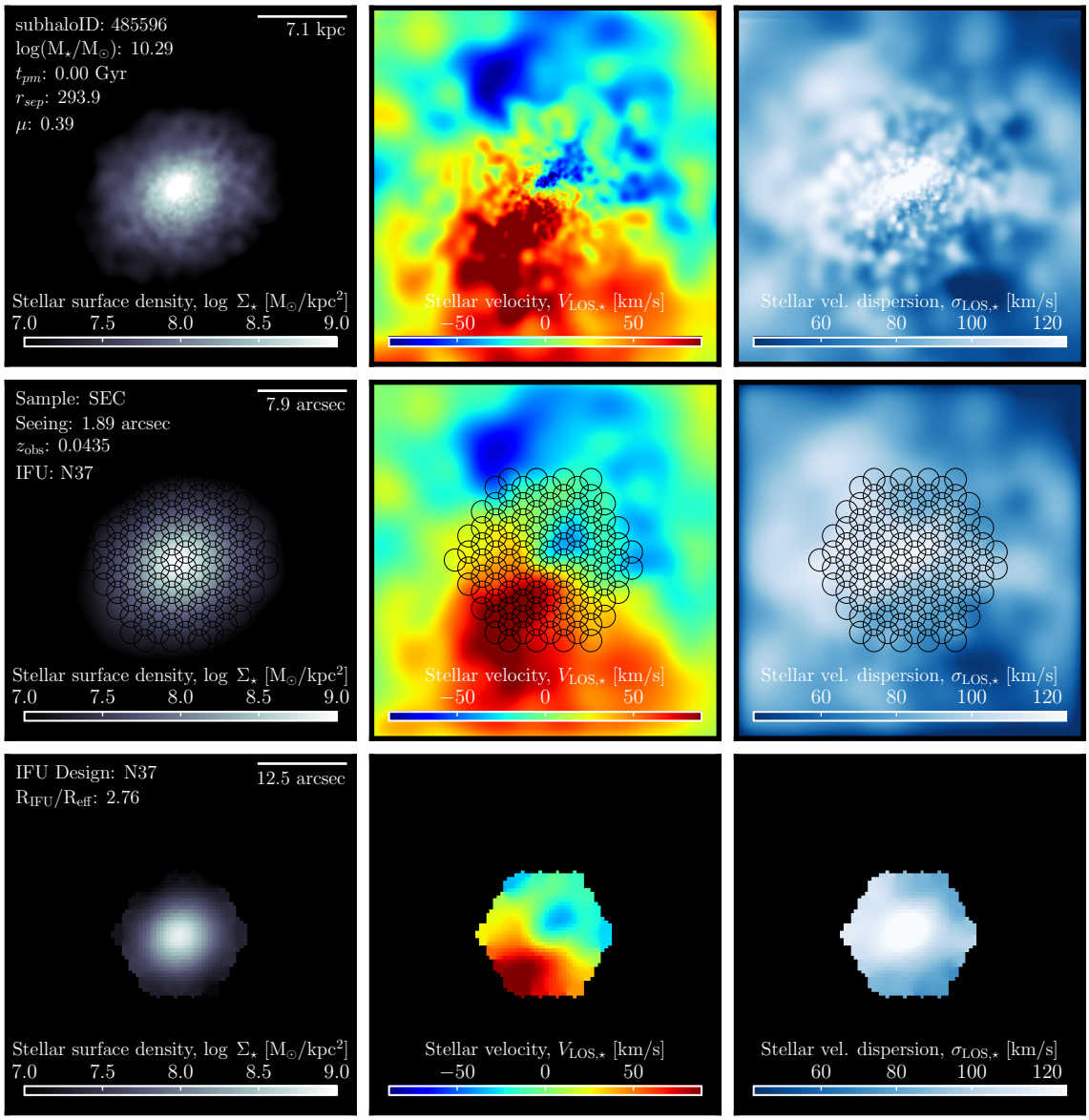

Figure 4.13 Similar to Figure 4.11 but for galaxy ID: 485596 observed for the SEC sample. In the SEC sample, IFUs are assigned to galaxies such that they cover a target $R_{\rm IFU}/R_{\rm eff} = 2.5$. The galaxy is the other recent post-merger from the fourth row of Figure 4.2 ($t_{\rm pm} = 0$ Gyr, $\mu = 0.39$). The galaxy shows no strong morphological disturbances but is heavily disturbed kinematically. Crucially, these kinematic disturbances are preserved in the realistic synthetic MaNGA kinematic maps.

merger galaxy from the bottom row of Figure 4.2. The SEC sample has a higher target coverage of $N_{R_{\rm eff}} = R_{\rm IFU}/R_{\rm eff} = 2.5$ than the other samples. In Section 4.2.2, I highlighted that the idealized morphology of this galaxy tells a different story than its kinematics. The galaxy is morphologically regular, but its kinematics are highly disturbed. In the RealSim-IFS synthetic MaNGA kinematic maps, this kinematic signature of the merger is preserved. The results of Figures 4.12 and 4.13 highlight the potential to achieve more accurate post-merger classifications by using imaging information in tandem with kinematics.

Lastly, Figure 4.14 shows a galaxy that is currently undergoing an interaction with a companion that is at least a tenth of its mass (following the definition of $r_{\rm sep}$ in Patton et al. 2020). It is possible that the interaction has driven the slight asymmetry and very loose winding of the spiral arms of this galaxy (e.g. Casteels et al. 2013). The idealized velocity map is also asymmetric in the magnitudes of the velocities about the minimum gravitational potential. Whether this galaxy is truly interacting or not (which can be directly determined by examining the history of this galaxy in the simulations), the crucial outcome is that the the signatures of kinematic irregularity are preserved in the final synthetic MaNGA observations.

Kinematic asymmetries may play an important role in calibrating the next generation of merger classification models. Therefore, the fact that the kinematic irregularities and asymmetries are well-preserved by RealSim-IFS after incorporating instrumental effects is promising. It is particularly exciting given the results of the tests in Section 4.4.1 showing the precision with which RealSim-IFS emulates the observational response of fibre based IFS instruments – meaning that these signatures should be accessible and exploitable in real (and current) fibre-based IFS observations of galaxies. A number of supplementary examples from the TNG100-1 synthetic MaNGA kinematic survey described in this section are provided in Appendix G following the same structure as Figures 4.11−4.14. In the next chapter, I forecast future research which will exploit spatially resolved kinematics (alongside imaging) from current and forthcoming observational programmes for accurate identification and detailed characterization of galaxy mergers.

4.5 Summary

In this chapter, I have introduced RealSim-IFS: a suite for generating realistic synthetic fibre-based IFS observations of galaxies from hydrodynamical simulations.

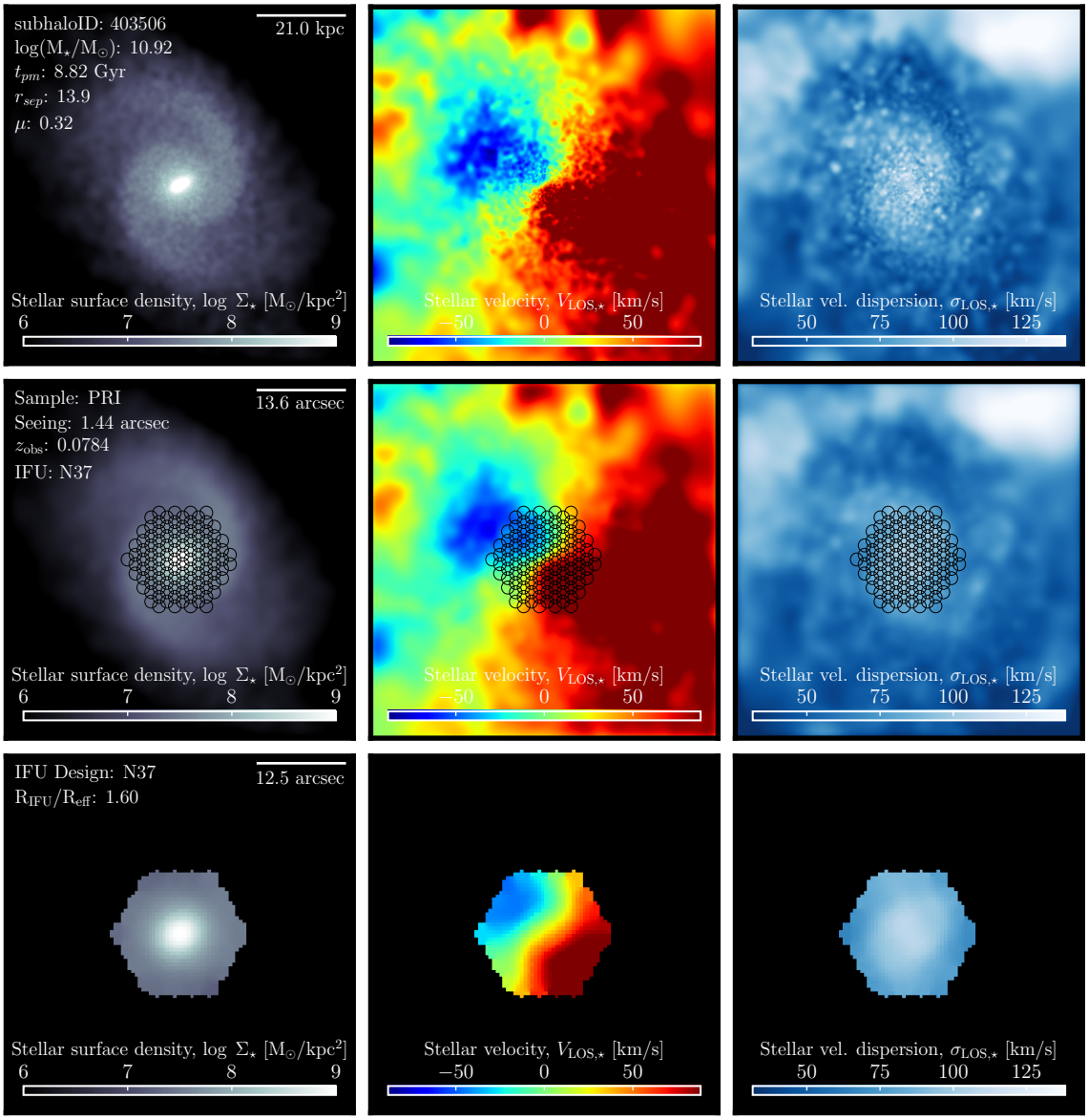

Figure 4.14 Similar to Figure 4.11 but for galaxy ID: 403506 observed for the PRI sample. The galaxy is undergoing a merger (minimum mass ratio $\mu_{\rm comp} \geq 0.1$) with a companion at relative separation $r_{\rm sep} = 13.9$ but that is not within the FOV in this camera angle. $\mu_{\rm comp}$ is not not to be confused with μ shown in the figure – which is the mass ratio for the most recently completed merger at $t_{\rm pm}$.

Extra-galactic astronomy has entered an era of large-scale IFS surveys providing spatially resolved kinematics for $10^3 - 10^4$ galaxies. These numbers will continue to grow with forthcoming observational IFS programmes. Similarly, we have seen vast improvements in the fidelity of galaxies produced in modern state-of-the-art hydrodynamical simulations – which model the co-evolution of dark and baryonic matter in statistically representative volumes from cosmological initial conditions. The primary goal of RealSim-IFS is to unify the major advances on these two fronts by enabling: (1) even-handed comparisons between simulated and observed galaxy IFS properties and (2) calibration of models which connect *observable* kinematic signatures of galaxies to their evolutionary histories and origins.

RealSim-IFS is designed to accurately emulate the observational signatures of fibre-based IFS instruments. The suite is highly flexible and can be applied to produce synthetic observations for any existing or future fibre-based IFU design. The three core functionalities of RealSim-IFS for producing the synthetic IFS observations are:

1. **Reproduction and/or creation of arbitrary IFU designs and observing strategies** (Section 4.3.1, Figure 4.3). RealSim-IFS accepts an arbitrary set of fibre positions, dithered exposure offsets or rotations, and corresponding fibre core aperture diameters.

2. **Integration of spatially discretized data into arbitrary fibre apertures** (either spectral or LOSVD data cubes, see Section 4.3.2, Figure 4.4). RealSim-IFS effectively uses an inverted *drizzle* algorithm to integrate flux within fibre apertures.

3. **Spatial reconstruction of irregularly sampled fibre data into regular Cartesian grid spaxels** (Section 4.3.3). RealSim-IFS contains modules for both *drizzle* and modified Shepard spatial reconstruction algorithms employed in the DRPs of current IFS surveys.

Section 4.4 showed tests and demonstrations with RealSim-IFS on both real and simulated data. I tested the precision with which RealSim-IFS reproduces the spatial reconstruction and cube creation component of the MaNGA DRP using real MaNGA fibre data (Section 4.4.1, Figure 4.8). I showed that after incorporating the role of DAR and other astrometry effects in the fibre positions (Figure 4.7), the data

cubes are produced with exceptional precision with respect to the MaNGA cubes (median $\Delta \log f_\lambda \sim 10^{-8}$ dex).

In Section 4.4.2, I demonstrated the application of RealSim-IFS to creating a synthetic MaNGA kinematic survey of galaxies the IllustrisTNG TNG100-1 cosmological hydrodynamical simulation. First, the demonstration illustrates how to emulate the statistics for redshift (Section 4.4.2), atmospheric seeing (Section 4.4.2), and IFU design assignment (Section 4.4.2) for the various MaNGA samples. Then, Figures 4.11-4.14 from Section 4.4.2 show examples which break down the synthetic observations for individual targets. In particular, I discuss the preservation of kinematic irregularities in the final synthetic MaNGA kinematic maps that are driven by galaxy mergers. Given that galaxy mergers can drive transient irregularities in the kinematics of their constituents and remnants, I highlight the potential of realistic synthetic kinematic observations such as these (for which the true statuses and properties of mergers are known) for calibrating more accurate merger classification models.

Chapter Acknowledgements

I am grateful for the support of a National Sciences and Engineering Research Council of Canada (NSERC) Graduate Scholarship. I thank Maan H. Hani (MHH) and Sara L. Ellison who provided valued and regular feedback and advice on this project. MHH was responsible for generating the LOSVD cubes for the IllustrisTNG simulations. I thank David Law for valued technical discussions and insight on MaNGA data reduction and Nic Scott for providing a set of SAMI fibre positions. The data processing and analysis conducted for this project were enabled by the computational resources provided by Compute Canada (www.computecanada.ca). This research made use of Astropy[9], a community-developed core Python package for Astronomy (Astropy Collaboration et al., 2013, 2018).

The MaNGA survey is part of SDSS-IV. Funding for SDSS-IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Partici-

[9]http://www.astropy.org

pating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Chapter 5

The future of galaxy and galaxy merger characterization

The enterprise of morphological and kinematic characterization of galaxies is to reveal the connections between their observed physical states and the events and processes which have shaped them. The underlying requirements to satisfy this goal are to (1) make robust distinctions between unique physical states and (2) identify and characterize galaxies that are transitioning. In this thesis, I have presented research that takes on both of these topics. Briefly, I have:

- Quantitatively evaluated the sensitivities of morphological measurements to the quality of the images from which the measurements derive (specifically, the depth) [Chapter 2].
- Developed a novel framework for identifying and distinguishing between galaxies that (a) are currently undergoing mergers and (b) have recently merged. In this framework, deep learning models are calibrated on synthetic images of galaxies from hydrodynamical simulations – where the true merger statuses of galaxies are known [Chapter 3].
- Shown that the synthetic images used to calibrate merger classification models *must* incorporate a high level of *realism* if the models are to be expected to identify and characterize galaxy mergers in real survey images [Chapter 3].
- Developed a comprehensive tool for generating survey-realistic synthetic IFS observations of galaxies from hydrodynamical simulations. The tool was designed

to generate realistic synthetic kinematic data for galaxies from hydrodynamical simulations to enable/improve merger classification and characterization [Chapter 4].

In this chapter, I outline how the methods, tools, and results I have presented in this thesis build toward a more refined understanding of the role of mergers in galaxy evolution. I also discuss the roles of current and forthcoming observational programmes in realizing this goal.

The role of photometric depth and quality

The comparison of galaxy morphologies derived from SDSS Legacy and Stripe 82 images revealed the sensitivities of measurements such as stellar bulge and disc fractions (Figures 2.8 and 2.9), their respective properties (Figures 2.11 and 2.12, and galaxy asymmetries (Figure 2.13) to depth at fixed angular resolution. In particular, asymmetric morphological structures such as tidal tails, shells, streams, and plumes can be exceptionally useful indicators of recent and ongoing merger activity. Chapter 2 showed that these asymmetric features, which are often at low surface-brightnesses with respect to their hosts, are enhanced (and even revealed) in deep co-add image relative to the single-epoch images in the SDSS Stripe 82.

From a theoretical perspective, the asymmetric morphological features triggered by mergers encode the initial morphologies, orbital scenarios, and stages of mergers (e.g. Hernquist & Quinn 1989; Helmi & White 1999; Johnston et al. 2008; Hendel & Johnston 2015; Hendel et al. 2019). The latter, that merger *stages* can be estimated from observed galaxy morphologies, was demonstrated explicitly in Chapter 3. However, the sensitivity goes beyond stage. Mergers between galaxies with initially high-eccentricity orbits tend to produce stellar streams. In contrast, mergers with low-eccentricity (more radial) orbits tend to yield stellar shells as each galaxy passes through a much deeper part of the gravitational potential of its companion. These theoretical studies have enormous implications for the task of estimating the initial conditions of galaxy mergers – particularly using the methods and tools presented in this thesis. Accurate estimations of the stages and initial conditions of observed mergers will open a new frontier in the study of galaxy mergers and refine their role in galaxy evolution. Specifically, the *observational* sensitivities of merger-triggered phenomena such as star-formation, chemical evolution, and AGN activity to merger

stages and initial conditions may be robustly quantified for the first time. However, the morphological features which may enable such detailed characterizations must obviously be *observable* if they are to be useful – not buried in sky noise, for example.

Current and forthcoming wide-field, deep imaging programmes are poised to enhance low surface-brightness structures with unprecedented sensitivity for large numbers of galaxies in the local Universe. Figure 5.1 shows what is offered by one such current and ongoing survey, the Canada France Imaging Survey (CFIS, Ibata et al. 2017; Cuillandre et al. in prep). The upper panels show the evolution of star formation rate and three-dimensional separation as functions of time along the interaction sequence (where $t = 0$ Gyr corresponds to the time at which the merging galaxies are 100 kpc apart). The lower panels show synthetic images for five snapshots along the sequence. The upper row of synthetic observations shows images with SDSS $r-$band realism as detailed in Chapter 3. The RealSim code used to generate these synthetic SDSS images can be adapted to any imaging instrument. For example, the lower row shows synthetic images with CFIS realism – generated using a CFIS expansion[1] I developed for the RealSim suite.

The contrast in the visibility of asymmetric and disturbed morphological structures between the SDSS and CFIS synthetic images is stark (see also Figure 1 of Ellison et al. 2019, for example). Upon completion, CFIS will cover $5,000$ square degrees in the $r-$band with a sensitivity ~ 2 magnitudes deeper than the SDSS (a factor of ~ 6 improvement in signal-to-noise ratio) and with significantly improved angular resolution (PSF FHWM ~ 0.6 arcseconds). Similarly, the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP, Aihara et al. 2018a,b) will offer similar photometric depth and resolution to CFIS but in five broadband filters covering $1,500$ square degrees (Komiyama et al., 2018; Miyazaki et al., 2018). These current deep imaging surveys are probing significantly fainter structures for galaxies in the local Universe and will offer more information that can be exploited by merger identification and characterization models. One of primary advantages of these two surveys is that they each have significant overlap with the SDSS and so can exploit a wealth of ancillary data: fibre spectra, redshifts, environmental characteristics, and integral field spectroscopy.

In addition to these current and ongoing surveys, the future Euclid space-based programme (expected launch in 2022) includes wide-field optical and near-infrared imaging components that will cover $15,000$ square degrees at even greater depths

[1] `https://github.com/cbottrell/RealSimCFIS`

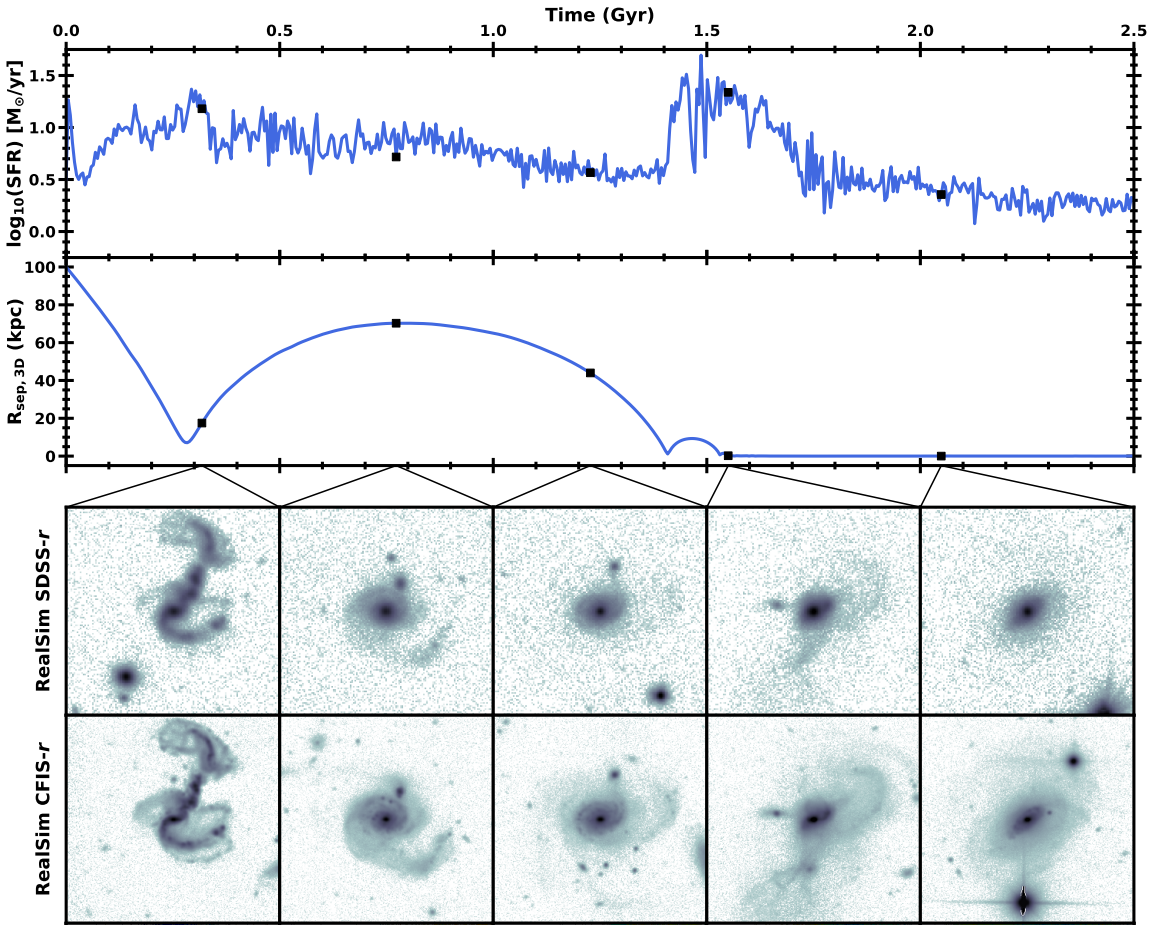

Figure 5.1 The role of depth and resolution in the observability of asymmetric, low surface-brightness substructure driven by galaxy mergers. The upper panels show the evolution of star-formation rate (SFR) and three-dimensional radial separation, $R_{\rm sep}$, between two merging galaxies from one of the Moreno et al. (2019) FIRE merger simulations used in Chapter 3. The lower panels show synthetic observations corresponding to particular snapshots along the interaction sequence. The upper row of synthetic images incorporate SDSS realism. The lower row incorporates realism from the Canada France Imaging Survey (CFIS, Ibata et al. 2017; Cuillandre et al. in prep) generated with a CFIS expansion to RealSim. The contrast in the amount of visible asymmetric substructure is stark – particularly in the third and fifth column which show the merger 1 Gyr after first pericentre and 0.5 Gyr after coalescence, respectively. The improved observability of asymmetric features stands to improve the accuracy of merger classification models such as those presented in this thesis.

and 0.1 and 0.3 arcseconds angular resolution, respectively – uninhibited by sky brightness and atmospheric blurring (Refregier et al., 2010; Laureijs et al., 2011). Taken together, the improved breadth, depth, and resolution of images from current programmes such as CFIS and HSC-SSP and future programmes such as Euclid offer a rich future for galaxy and galaxy merger characterization.

The role of galaxy kinematics

Chapter 3 showed that mergers can be identified and characterized by stage from images and the last section forecasted the role of deep, wide-field imaging programmes in improving the accuracy and detail to which mergers can be characterized in the local Universe. However, Chapter 4 showed that galaxy kinematics, derived from integral field spectroscopy observations, may provide a highly complimentary basis for realizing this goal (e.g. see Figure 4.2 and Appendix G).

Current multi-object IFS surveys such as SAMI (Croom et al., 2012) and MaNGA (Bundy et al., 2015) have each already obtained spatially resolved kinematics for thousands of galaxies. Furthermore, we can expect to see further expansion and refinement in IFS observations with future programmes such as Hector (Bryant et al., 2018). Galaxy kinematics may enable improved merger characterization either independently or in tandem with photometry. From an information theory perspective, it is *expected* that a merger characterization model that employs photometry and kinematics in tandem can do no poorer than either method independently. In particular, given that the information provided by photometry and kinematics are intrinsically unique, it is expected that models that combine kinematics and photometry should perform *better* than models that use only one type of data.

Therefore, incorporating kinematic information may be essential to calibrating models that can accurately estimate the stages and initial conditions of observed mergers – thereby enabling robust quantification of the sensitivities of star formation, chemical evolution, and AGN activity to these estimates. Indeed, the *relative* importance of kinematic information can be tested explicitly using the combination of hydrodynamical simulations, synthetic observations, and deep learning outlined in Chapter 3. Specifically, deep merger classification models can be trained on synthetic kinematic and photometric observations independently and in tandem. The resulting relative accuracies of models that are calibrated on (1) photometry alone, (2) kinematics alone, and (3) photometry *and* kinematics would reveal which data are most

valuable to the task and the *added* value (if any) of combining the data[2].

Figure 5.2 shows a thematic deep learning model combining photometric and kinematic data within the same architecture. The convolutional model is divided into separate branches for multi-band photometry (left) and gas/stellar kinematics (right). Each branch comprises a unique set of feature extraction layers (e.g. see Figure 1.10) – optimized to extract the most useful information from each data type independently. The output from each feature-extraction branch are then flattened and concatenated into a shared, one-dimensional feature array and fed into the classifier. The output from the model may be simple or complex. For example, the output may be merger status or stage as demonstrated in Chapter 3. However, as outlined in the last paragraph, it is of great value to establish the extent to which other merger parameters (such as their initial conditions) can be extracted. Therefore, the output could be the stage of the merger (expressed as time before/since coalescence ($t - t_{\rm c}$), the mass ratio (μ), the gas fraction ($f_{\rm gas}$), the orbital angular momentum relative to a stable circular orbit ($J/J_{\rm circ}$), and others.

Naturally, there may be a limit to what can be extracted given the large parameter space of initial conditions – which also includes the initial morphologies (e.g. bulge/disc fractions). But given that Chapter 3 showed that stage can be estimated with reasonable accuracy, it is promising that other, more detailed merger parameters may be extracted from the data – provided that sufficient information is given to the model. Crucially, these models *must* be calibrated on realistic synthetic images and kinematic observations of galaxies from hydrodynamical simulations (e.g. Chapters 3 and 4) – where all of these initial conditions are known *a priori*.

With the tools now in place for making the realistic synthetic images to be used to calibrate merger classification models (RealSim: Bottrell et al. 2017b; Chapter 3), Chapter 4 presented a tool, RealSim-IFS, that can be used to construct realistic synthetic kinematic observations. In particular, the core functionality of RealSim-IFS is to emulate the instrumental responses of current fibre-based IFS instruments such as SAMI and MaNGA for simulations – also anticipating the instrumental designs of future programmes such as Hector. Deep learning models are exceptionally sensitive and the data upon which a model is calibrated must be highly comparable to the data to which it is applied. The purpose of RealSim-IFS is to unify the advances in modern state-of-the-art simulations with current and forthcoming IFS instruments by putting their data products on even ground.

[2]Such as study is currently being carried out in Bottrell et al. in prep.

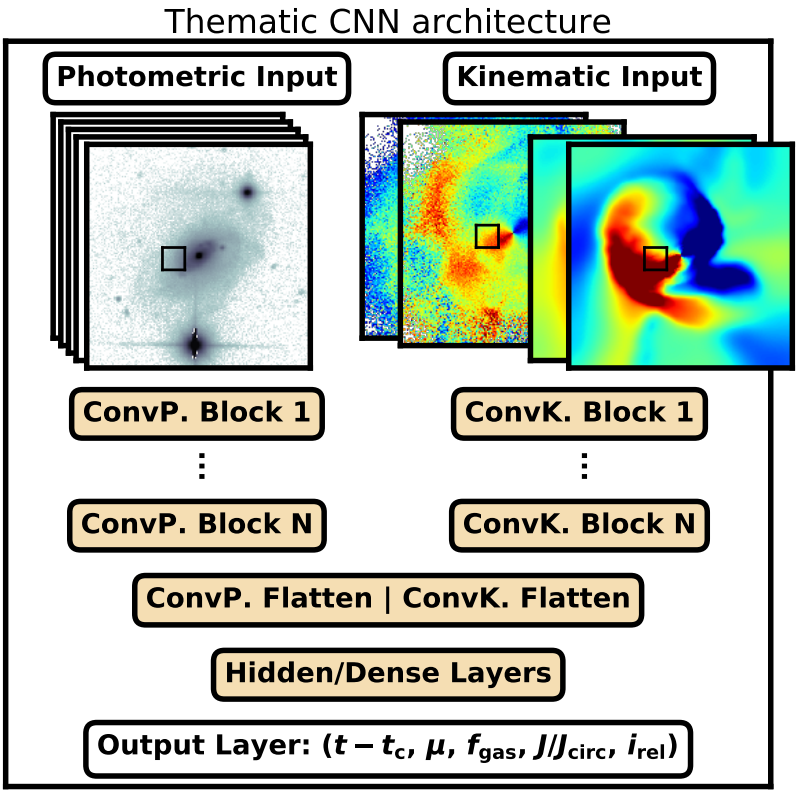

Figure 5.2 A *thematic* architecture for a deep learning model combining photometric and kinematic data simultaneously. In the left branch, the model takes multi-band photometry as input. The right branch takes kinematic input. Each branch comprises a series of unique feature extraction layers (e.g. see Figure 1.10). The output from each branch is flattened and concatenated in a one-dimensional feature array and fed into the classification and/or regression component (Hidden/Dense). The output from the model may include merger status (merger/non-merger) or more detailed information for mergers such as stage and initial conditions of the merger – including mass ratio, gas fraction, orbits, and relative inclinations.

Mergers beyond the local Universe

Detailed characterizations of galaxy mergers and their consequences in the local Universe will lay the foundation for merger studies at higher redshifts – when the relative role of mergers in galaxy-formation and evolution may have been very different. Though it is observationally well-established that the merger rate grows with increasing redshift as the average matter density of the Universe increases (e.g. Patton et al. 2002; Conselice et al. 2003; Lin et al. 2004; De Propris et al. 2007; Conselice et al. 2008; Patton & Atfield 2008; de Ravel et al. 2009; Lotz et al. 2011; López-Sanjuan et al. 2013b; Casteels et al. 2014; Mundy et al. 2017; Duncan et al. 2019; Ferreira et al. 2020), the relative role of mergers in the build-up of galaxy stellar mass, their morphologies, chemical evolution, and black-hole activity and growth in the early Universe is currently uncertain. For example, the relatively abundant supply of gas in the early Universe – tunnelling into galaxies along cosmic filaments – may mean that mergers played a more minor role in these phenomena (e.g. Dekel & Birnboim 2006; Birnboim et al. 2007; Dekel et al. 2009; Bournaud et al. 2011).

Combined with existing and future deep imaging programmes, forthcoming IFS surveys that probe samples of galaxies beyond the local Universe will be major cornerstones in understanding the relative role and importance of galaxy mergers throughout cosmic time. For example, the Middle Ages Galaxy Properties with Integral Field Spectroscopy (MAGPI) survey[3] will provide IFS observations of 60 massive galaxies and their combined 118 satellites at $z \sim 0.3$ (lookback time, $t_{\rm L} \sim 3.5$ Gyr) with the *same physical resolution* as MaNGA and SAMI in the local Universe ($z \sim 0.04$). MAGPI and other future IFS surveys that will push to higher redshifts will open the doors to entirely unexplored epochs for detailed morphological and kinematic characterization of galaxies – enabling connections to chemical evolution, star-formation, and AGN activity (among other properties) and yielding further insight into their origins and transformation.

[3] `https://magpisurvey.org/`

Bibliography

Abadi, M. G., Navarro, J. F., Steinmetz, M., & Eke, V. R. 2003a, Astrophysical Journal, 591, 499

— 2003b, Astrophysical Journal, 597, 21

Abazajian, K. N. et al. 2009, Astrophysical Journal, Supplement Series, 182, 543

Abraham, R. G., Valdes, F., Yee, H. K. C., & van den Bergh, S. 1994, Astrophysical Journal, 432, 75

Abraham, R. G., van den Bergh, S., Glazebrook, K., Ellis, R. S., Santiago, B. X., Surma, P., & Griffiths, R. E. 1996, Astrophysical Journal, Supplement Series, 107, 1

Abraham, R. G., van den Bergh, S., & Nair, P. 2003, Astrophysical Journal, 588, 218

Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, Monthly Notices of the RAS, 479, 415

Agertz, O. & Kravtsov, A. V. 2015, Astrophysical Journal, 804, 18

Agertz, O., Teyssier, R., & Moore, B. 2011, Monthly Notices of the RAS, 410, 1391

Aguado, D. S. et al. 2019, Astrophysical Journal, Supplement Series, 240, 23

Aihara, H. et al. 2018a, Publications of the ASJ, 70, S4

— 2018b, Publications of the ASJ, 70, S8

Albrecht, A. & Steinhardt, P. J. 1982, Phys. Rev. Lett., 48, 1220

Allen, P. D., Driver, S. P., Graham, A. W., Cameron, E., Liske, J., & de Propris, R. 2006, Monthly Notices of the RAS, 371, 2

Alonso, M. S., Lambas, D. G., Tissera, P., & Coldwell, G. 2007, Monthly Notices of the RAS, 375, 1017

Andredakis, Y. C., Peletier, R. F., & Balcells, M. 1995, Monthly Notices of the RAS, 275, 874

Andredakis, Y. C. & Sanders, R. H. 1994, Monthly Notices of the RAS, 267, 283

Annis, J. et al. 2014, Astrophysical Journal, 794, 120

Aragon-Salamanca, A., Ellis, R. S., Couch, W. J., & Carter, D. 1993, Monthly Notices of the RAS, 262, 764

Astropy Collaboration et al. 2018, Astronomical Journal, 156, 123

— 2013, Astronomy and Astrophysics, 558, A33

Bacon, R. et al. 2010, in Proc. SPIE, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 773508

Baes, M., Verstappen, J., De Looze, I., Fritz, J., Saftly, W., Vidal Pérez, E., Stalevski, M., & Valcke, S. 2011, Astrophysical Journal, Supplement Series, 196, 22

Balcells, M., Graham, A. W., Domínguez-Palmero, L., & Peletier, R. F. 2003, Astrophysical Journal, Letters to the Editor, 582, L79

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, Astrophysical Journal, 600, 681

Balogh, M. L., Baldry, I. K., Nichol, R., Miller, C., Bower, R., & Glazebrook, K. 2004, Astrophysical Journal, Letters to the Editor, 615, L101

Bamford, S. P. et al. 2009, Monthly Notices of the RAS, 393, 1324

Barnes, J. E. 1988, Astrophysical Journal, 331, 699

— 2011, Monthly Notices of the RAS, 413, 2860

Barnes, J. E. & Hernquist, L. 1992, Annual Review of Astronomy and Astrophysics, 30, 705

Barrera-Ballesteros, J. K. et al. 2014, Astronomy and Astrophysics, 568, A70

— 2015, Astronomy and Astrophysics, 582, A21

Barrows, R. S., Comerford, J. M., Greene, J. E., & Pooley, D. 2017, Astrophysical Journal, 838, 129

Barton, E. J., Geller, M. J., & Kenyon, S. J. 2000, The Astrophysical Journal, 530, 660

Bayes, T. 1763, Philosophical transactions of the Royal Society of London, 370

Bell, E. F. 2008, Astrophysical Journal, 682, 355

Benson, A. J., Džanović, D., Frenk, C. S., & Sharples, R. 2007, Monthly Notices of the RAS, 379, 841

Berg, T. A. M., Simard, L., Mendel Trevor, J., & Ellison, S. L. 2014, Monthly Notices of the RAS, 440, L66

Bernardi, M., Meert, A., Vikram, V., Huertas-Company, M., Mei, S., Shankar, F., & Sheth, R. K. 2014, Monthly Notices of the RAS, 443, 874

Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, Astronomical Journal, 119, 2645

Bertin, E. & Arnouts, S. 1996, Astronomy and Astrophysics, Supplement Series, 117, 393

Bertin, E., Mellier, Y., Radovich, M., Missonnier, G., Didelon, P., & Morin, B. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228

Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Birnboim, Y., Dekel, A., & Neistein, E. 2007, Monthly Notices of the RAS, 380, 339

Blanton, M. R. et al. 2003a, Astrophysical Journal, 594, 186

— 2003b, Astrophysical Journal, 592, 819

Blanton, M. R. & Roweis, S. 2007, Astronomical Journal, 133, 734

Bloom, J. V. et al. 2018, Monthly Notices of the RAS, 476, 2339

— 2017, Monthly Notices of the RAS, 465, 123

Bluck, A. F. L. et al. 2019, Monthly Notices of the RAS, 485, 666

Bluck, A. F. L., Conselice, C. J., Buitrago, F. o., Grützbauch, R., Hoyos, C., Mortlock, A., & Bauer, A. E. 2012, Astrophysical Journal, 747, 34

Bluck, A. F. L., Mendel, J. T., Ellison, S. L., Moreno, J., Simard, L., Patton, D. R., & Starkenburg, E. 2014, Monthly Notices of the RAS, 441, 599

Bluck, A. F. L. et al. 2016, Monthly Notices of the RAS, 462, 2559

Blumenthal, G. R., Faber, S. M., Flores, R., & Primack, J. R. 1986, Astrophysical Journal, 301, 27

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Blumenthal, K. A. & Barnes, J. E. 2018, Monthly Notices of the RAS, 479, 3952

Blumenthal, K. A. et al. 2020, Monthly Notices of the RAS, 492, 2075

Bottrell, C. et al. 2019a, Monthly Notices of the RAS, 490, 5390

Bottrell, C., Simard, L., Mendel, J. T., & Ellison, S. L. 2019b, Monthly Notices of the RAS, 486, 390

Bottrell, C., Torrey, P., Simard, L., & Ellison, S. L. 2017a, Monthly Notices of the RAS, 467, 1033

— 2017b, Monthly Notices of the RAS, 467, 2879

Bournaud, F., Dekel, A., Teyssier, R., Cacciato, M., Daddi, E., Juneau, S., & Shankar, F. 2011, Astrophysical Journal, Letters to the Editor, 741, L33

Brinchmann, J. et al. 1998, Astrophysical Journal, 499, 112

Brooks, A. M. et al. 2011, Astrophysical Journal, 728, 51

Bryant, J. J. et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10702, Proc. SPIE, 107021H

Bryant, J. J. et al. 2019, Monthly Notices of the RAS, 483, 458

— 2015, Monthly Notices of the RAS, 447, 2857

Buda, M., Maki, A., & Mazurowski, M. A. 2017, arXiv e-prints, arXiv:1710.05381

Bundy, K. et al. 2015, Astrophysical Journal, 798, 7

Bundy, K., Ellis, R. S., & Conselice, C. J. 2005, Astrophysical Journal, 625, 621

Bushouse, H. A. 1987, Astrophysical Journal, 320, 49

Butcher, H. & Oemler, Jr., A. 1984, Astrophysical Journal, 285, 426

Camps, P. & Baes, M. 2015, Astronomy and Computing, 9, 20

Cappellari, M. 2017, Monthly Notices of the RAS, 466, 798

Cappellari, M. et al. 2011, Monthly Notices of the RAS, 416, 1680

Cardie, C. & Howe, N. 1997, in ICML, 57–65

Carlberg, R. G., Pritchet, C. J., & Infante, L. 1994, Astrophysical Journal, 435, 540

Casteels, K. R. V. et al. 2013, Monthly Notices of the RAS, 429, 1051

Casteels, K. R. V. et al. 2014, Monthly Notices of the RAS, 445, 1157

Chan, P. K. & Stolfo, S. J. 1998, in KDD, Vol. 1998, 164–168

Cheng, T.-Y. et al. 2020, Monthly Notices of the RAS, 493, 4209

Cheung, E. et al. 2012, Astrophysical Journal, 760, 131

Chollet, F. et al. 2015, Keras, `https://keras.io`

Christensen, C. R., Brooks, A. M., Fisher, D. B., Governato, F., McCleary, J., Quinn, T. R., Shen, S., & Wadsley, J. 2014, Monthly Notices of the RAS, 440, L51

Ciambur, B. C. 2015, Astrophysical Journal, 810, 120

Clowe, D., Luppino, G. A., Kaiser, N., & Gioia, I. M. 2000, Astrophysical Journal, 539, 540

Cole, S., Hatton, S., Weinberg, D. H., & Frenk, C. S. 1998, Monthly Notices of the RAS, 300, 945

Cole, S. et al. 2005, Monthly Notices of the RAS, 362, 505

Colina, L., Arribas, S., & Monreal-Ibero, A. 2005, Astrophysical Journal, 621, 725

Colless, M. et al. 2001, Monthly Notices of the RAS, 328, 1039

Comerford, J. M., Schluns, K., Greene, J. E., & Cool, R. J. 2013, Astrophysical Journal, 777, 64

Conselice, C. J. 2003, Astrophysical Journal, Supplement Series, 147, 1

— 2006, Monthly Notices of the RAS, 373, 1389

Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, Astronomical Journal, 126, 1183

Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, Astrophysical Journal, 529, 886

Conselice, C. J., Rajgor, S., & Myers, R. 2008, Monthly Notices of the RAS, 386, 909

Courteau, S. 1997, Astronomical Journal, 114, 2402

Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, Monthly Notices of the RAS, 384, 386

Crain, R. A. et al. 2009, Monthly Notices of the RAS, 399, 1773

Croft, R. A. C., Di Matteo, T., Springel, V., & Hernquist, L. 2009, Monthly Notices of the RAS, 400, 43

Croom, S. M. et al. 2012, Monthly Notices of the RAS, 421, 872

Dalton, G. et al. 2012, in Proc. SPIE, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84460P

Danovich, M., Dekel, A., Hahn, O., Ceverino, D., & Primack, J. 2015, Monthly Notices of the RAS, 449, 2087

Darg, D. W. et al. 2010, Monthly Notices of the RAS, 401, 1043

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, Astrophysical Journal, 292, 371

Davis, M. & Geller, M. J. 1976, Astrophysical Journal, 208, 13

Davis, M., Huchra, J., Latham, D. W., & Tonry, J. 1982, Astrophysical Journal, 253, 423

de Jong, R. S. 1996, Astronomy and Astrophysics, 313, 45

De Propris, R., Conselice, C. J., Liske, J., Driver, S. P., Patton, D. R., Graham, A. W., & Allen, P. D. 2007, Astrophysical Journal, 666, 212

de Ravel, L. et al. 2009, Astronomy and Astrophysics, 498, 379

de Souza, R. E., Gadotti, D. A., & dos Anjos, S. 2004, Astrophysical Journal, Supplement Series, 153, 411

de Vaucouleurs, G. 1959a, Handbuch der Physik, 53, 275

— 1959b, Handbuch der Physik, 53, 311

Dechter, R. 1986, 178–185

Dekel, A. & Birnboim, Y. 2006, Monthly Notices of the RAS, 368, 2

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

Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604

Dickinson, H. et al. 2018, Astrophysical Journal, 853, 194

Dieleman, S., Willett, K. W., & Dambre, J. 2015, Monthly Notices of the RAS, 450, 1441

Djorgovski, S. & Davis, M. 1987, Astrophysical Journal, 313, 59

Doi, M. et al. 2010, Astronomical Journal, 139, 1628

Domínguez Sánchez, H. et al. 2019, Monthly Notices of the RAS, 484, 93

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, Monthly Notices of the RAS, 476, 3661

Dressler, A. 1980, Astrophysical Journal, 236, 351

Dressler, A., Lynden-Bell, D., Burstein, D., Davies, R. L., Faber, S. M., Terlevich, R., & Wegner, G. 1987, Astrophysical Journal, 313, 42

Driver, S. P. et al. 2006, Monthly Notices of the RAS, 368, 414

Drory, N. et al. 2015, Astronomical Journal, 149, 77

Du, M., Ho, L. C., Debattista, V. P., Pillepich, A., Nelson, D., Zhao, D., & Hernquist, L. 2020, arXiv e-prints, arXiv:2002.04182

Duncan, K. et al. 2019, Astrophysical Journal, 876, 110

Ellison, S. L., Mendel, J. T., Patton, D. R., & Scudder, J. M. 2013, Monthly Notices of the RAS, 435, 3627

Ellison, S. L., Patton, D. R., & Hickox, R. C. 2015, Monthly Notices of the RAS, 451, L35

Ellison, S. L., Patton, D. R., Mendel, J. T., & Scudder, J. M. 2011, Monthly Notices of the RAS, 418, 2043

Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, Astronomical Journal, 135, 1877

Ellison, S. L., Patton, D. R., Simard, L., McConnachie, A. W., Baldry, I. K., & Mendel, J. T. 2010, Monthly Notices of the RAS, 407, 1514

Ellison, S. L., Viswanathan, A., Patton, D. R., Bottrell, C., McConnachie, A. W., Gwyn, S., & Cuillandre, J.-C. 2019, Monthly Notices of the RAS, 487, 2491

Elmegreen, B. G., Elmegreen, D. M., & Montenegro, L. 1992, Astrophysical Journal, Supplement Series, 79, 37

Emsellem, E. et al. 2007, Monthly Notices of the RAS, 379, 401

Emsellem, E., Greusard, D., Combes, F., Friedli, D., Leon, S., Pécontal, E., & Wozniak, H. 2001, Astronomy and Astrophysics, 368, 52

Faber, S. M. & Jackson, R. E. 1976, Astrophysical Journal, 204, 668

Faber, S. M. et al. 2007, Astrophysical Journal, 665, 265

Falcón-Barroso, J. et al. 2006, Monthly Notices of the RAS, 369, 529

Fall, S. M. & Efstathiou, G. 1980, Monthly Notices of the RAS, 193, 189

Fang, J. J., Faber, S. M., Koo, D. C., & Dekel, A. 2013, Astrophysical Journal, 776, 63

Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, Astrophysical Journal, 703, 1416

Feng, S., Shen, S.-Y., Yuan, F.-T., Riffel, R. A., & Pan, K. 2020, Astrophysical Journal, Letters to the Editor, 892, L20

Ferreira, L., Conselice, C. J., Duncan, K., Cheng, T.-Y., Griffiths, A., & Whitney, A. 2020, arXiv e-prints, arXiv:2005.00476

Filippenko, A. V. 1982, Publications of the ASP, 94, 715

Fischer, J. L., Domínguez Sánchez, H., & Bernardi, M. 2019, Monthly Notices of the RAS, 483, 2057

Fischer, P. et al. 2000, Astronomical Journal, 120, 1198

Fisher, D. B. & Drory, N. 2008, Astronomical Journal, 136, 773

— 2010, Astrophysical Journal, 716, 942

Fixsen, D. J. 2009, Astrophysical Journal, 707, 916

Fliri, J. & Trujillo, I. 2016, Monthly Notices of the RAS, 456, 1359

Foord, A., Gültekin, K., Nevin, R., Comerford, J. M., Hodges-Kluck, E., Barrows, R. S., Goulding, A. D., & Greene, J. E. 2020, Astrophysical Journal, 892, 29

Freeman, K. C. 1970, Astrophysical Journal, 160, 811

Frenk, C. S., White, S. D. M., Davis, M., & Efstathiou, G. 1988, Astrophysical Journal, 327, 507

Frenk, C. S., White, S. D. M., Efstathiou, G., & Davis, M. 1985, Nature, 317, 595

Frieman, J. A. et al. 2008a, Astronomical Journal, 135, 338

Frieman, J. A., Turner, M. S., & Huterer, D. 2008b, Annual Review of Astronomy and Astrophysics, 46, 385

Fruchter, A. S. & Hook, R. N. 2002, Publications of the ASP, 114, 144

Fu, H. et al. 2018, Astrophysical Journal, 856, 93

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, Astronomical Journal, 111, 1748

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

Furlong, M. et al. 2015, Monthly Notices of the RAS, 450, 4486

Gadotti, D. A. 2009, Monthly Notices of the RAS, 393, 1531

Geller, M. J. & Huchra, J. P. 1989, Science, 246, 897

Genel, S. et al. 2014, Monthly Notices of the RAS, 445, 175

Gerber, R. A. & Lamb, S. A. 1994, Astrophysical Journal, 431, 604

Giusarma, E., Reyes Hurtado, M., Villaescusa-Navarro, F., He, S., Ho, S., & Hahn, C. 2019, arXiv e-prints, arXiv:1910.04255

Glazebrook, K. 2013, PASA, 30, e056

Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep learning (MIT press)

Goulding, A. D. et al. 2018, Publications of the ASJ, 70, S37

Governato, F. et al. 2004, Astrophysical Journal, 607, 688

Graham, A. W. 2001, Astronomical Journal, 121, 820

— 2013, Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology, Vol. 6 (Springer Netherlands), 91

Graham, A. W., Ciambur, B. C., & Savorgnan, G. A. D. 2016, Astrophysical Journal, 831, 132

Graham, A. W. & Guzmán, R. 2003, Astronomical Journal, 125, 2936

Groves, B., Dopita, M. A., Sutherland, R. S., Kewley, L. J., Fischera, J., Leitherer, C., Brandl, B., & van Breugel, W. 2008, Astrophysical Journal, Supplement Series, 176, 438

Grzymala-Busse, J. W., Goodwin, L. K., Grzymala-Busse, W. J., & Zheng, X. 2004, in Rough-neural computing (Springer), 543–553

Guedes, J., Callegari, S., Madau, P., & Mayer, L. 2011, Astrophysical Journal, 742, 76

Gunn, J. E. & Gott, J. Richard, I. 1972, Astrophysical Journal, 176, 1

Gunn, J. E. et al. 2006, Astronomical Journal, 131, 2332

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. 2017, Expert Systems with Applications, 73, 220

Hani, M. H., Gosain, H., Ellison, S. L., Patton, D. R., & Torrey, P. 2020, Monthly Notices of the RAS, 493, 3716

Hani, M. H., Sparre, M., Ellison, S. L., Torrey, P., & Vogelsberger, M. 2018, Monthly Notices of the RAS, 475, 1160

Harrison, E. R. 1970, Phys. Rev. D, 1, 2726

Hausen, R. & Robertson, B. 2019, arXiv e-prints, arXiv:1906.11248

Häußler, B. et al. 2013, Monthly Notices of the RAS, 430, 330

Hayward, C. C. & Hopkins, P. F. 2017, Monthly Notices of the RAS, 465, 1682

He, K., Zhang, X., Ren, S., & Sun, J. 2015, arXiv e-prints, arXiv:1512.03385

Helmi, A. & White, S. D. M. 1999, Monthly Notices of the RAS, 307, 495

Hendel, D. & Johnston, K. V. 2015, Monthly Notices of the RAS, 454, 2472

Hendel, D., Johnston, K. V., Patra, R. K., & Sen, B. 2019, Monthly Notices of the RAS, 486, 3604

Hernández-Toledo, H. M., Avila-Reese, V., Conselice, C. J., & Puerari, I. 2005, Astronomical Journal, 129, 682

Hernández-Toledo, H. M., Avila-Reese, V., Salazar-Contreras, J. R., & Conselice, C. J. 2006, Astronomical Journal, 132, 71

Hernández-Toledo, H. M., Zendejas-Domínguez, J., & Avila-Reese, V. 2007, Astronomical Journal, 134, 2286

Hernquist, L. 1989, Nature, 340, 687

— 1990, Astrophysical Journal, 356, 359

— 1992, Astrophysical Journal, 400, 460

Hernquist, L. & Quinn, P. J. 1989, Astrophysical Journal, 342, 1

Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. 2017, Nature, 548, 555

Hinshaw, G. et al. 2013, Astrophysical Journal, Supplement Series, 208, 19

Hogg, D. W., Finkbeiner, D. P., Schlegel, D. J., & Gunn, J. E. 2001, Astronomical Journal, 122, 2129

Hopkins, P. F. 2015, Monthly Notices of the RAS, 450, 53

— 2017, arXiv e-prints, arXiv:1712.01294

Hopkins, P. F., Cox, T. J., Hernquist, L., Narayanan, D., Hayward, C. C., & Murray, N. 2013, Monthly Notices of the RAS, 430, 1901

Hopkins, P. F., Cox, T. J., Kereš, D., & Hernquist, L. 2008a, Astrophysical Journal, Supplement Series, 175, 390

Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009, Astrophysical Journal, 691, 1168

Hopkins, P. F., Hernquist, L., Cox, T. J., Dutta, S. N., & Rothberg, B. 2008b, Astrophysical Journal, 679, 156

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008c, Astrophysical Journal, Supplement Series, 175, 356

Hopkins, P. F., Kereš, D., Oñorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., & Bullock, J. S. 2014, Monthly Notices of the RAS, 445, 581

Hopkins, P. F. & Quataert, E. 2010, Monthly Notices of the RAS, 407, 1529

Hopkins, P. F. et al. 2018, Monthly Notices of the RAS, 480, 800

Howard, S., Keel, W. C., Byrd, G., & Burkey, J. 1993, Astrophysical Journal, 417, 502

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hubble, E. P. 1926, Astrophysical Journal, 64

— 1936, Realm of the Nebulae (Yale University Press, New Haven)

Huertas-Company, M. et al. 2015, Astrophysical Journal, Supplement Series, 221, 8

— 2019, Monthly Notices of the RAS, 489, 1859

Hung, C.-L., Hayward, C. C., Smith, H. A., Ashby, M. L. N., Lanz, L., Martínez-Galarza, J. R., Sanders, D. B., & Zezas, A. 2016, Astrophysical Journal, 816, 99

Hung, C.-L. et al. 2015, Astrophysical Journal, 803, 62

Ibata, R. A. et al. 2017, Astrophysical Journal, 848, 128

Ivezić, Ž. et al. 2019, Astrophysical Journal, 873, 111

Ivezić, Ž. et al. 2004, Astronomische Nachrichten, 325, 583

Jacobs, C. et al. 2019, Monthly Notices of the RAS, 484, 5330

Jesseit, R., Cappellari, M., Naab, T., Emsellem, E., & Burkert, A. 2009, Monthly Notices of the RAS, 397, 1202

Jesseit, R., Naab, T., Peletier, R. F., & Burkert, A. 2007, Monthly Notices of the RAS, 376, 997

Ji, I., Peirani, S., & Yi, S. K. 2014, Astronomy and Astrophysics, 566, A97

Jiang, L. et al. 2014, Astrophysical Journal, Supplement Series, 213, 12

Johnson, S. D., Chen, H.-W., & Mulchaey, J. S. 2015, Monthly Notices of the RAS, 449, 3263

Johnston, K. V., Bullock, J. S., Sharma, S., Font, A., Robertson, B. E., & Leitner, S. N. 2008, Astrophysical Journal, 689, 936

Jonsson, P. 2006, Monthly Notices of the RAS, 372, 2

Jonsson, P., Groves, B. A., & Cox, T. J. 2010, Monthly Notices of the RAS, 403, 17

Kartaltepe, J. S. et al. 2015, Astrophysical Journal, Supplement Series, 221, 11

— 2007, Astrophysical Journal, Supplement Series, 172, 320

Katz, N., Hernquist, L., & Weinberg, D. H. 1992, Astrophysical Journal, Letters to the Editor, 399, L109

Katz, N., Weinberg, D. H., & Hernquist, L. 1996, Astrophysical Journal, Supplement Series, 105, 19

Kauffmann, G., Heckman, T. M., De Lucia, G., Brinchmann, J., Charlot, S., Tremonti, C., White, S. D. M., & Brinkmann, J. 2006, Monthly Notices of the RAS, 367, 1394

Kauffmann, G. et al. 2003a, Monthly Notices of the RAS, 346, 1055

— 2003b, Monthly Notices of the RAS, 341, 33

Kaviraj, S. 2010, Monthly Notices of the RAS, 406, 382

Keel, W. C., Kennicutt, R. C., J., Hummel, E., & van der Hulst, J. M. 1985, Astronomical Journal, 90, 708

Kelvin, L. S. et al. 2012, Monthly Notices of the RAS, 421, 1007

Kennicutt, Jr., R. C. 1998a, Annual Review of Astronomy and Astrophysics, 36, 189

— 1998b, Astrophysical Journal, 498, 541

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, Monthly Notices of the RAS, 363, 2

Kereš, D., Vogelsberger, M., Sijacki, D., Springel, V., & Hernquist, L. 2012, Monthly Notices of the RAS, 425, 2027

Kewley, L. J., Geller, M. J., & Barton, E. J. 2006, Astronomical Journal, 131, 2004

Khosroshahi, H. G., Wadadekar, Y., Kembhavi, A., & Mobasher, B. 2000, Astrophysical Journal, Letters to the Editor, 531, L103

Kirshner, R. P., Oemler, A., J., & Schechter, P. L. 1978, Astronomical Journal, 83, 1549

Komiyama, Y. et al. 2018, Publications of the ASJ, 70, S2

Kormendy, J. 1977, Astrophysical Journal, 217, 406

Kormendy, J. & Bender, R. 1996, Astrophysical Journal, Letters to the Editor, 464, L119

— 2012, Astrophysical Journal, Supplement Series, 198, 2

Kormendy, J., Bender, R., & Cornell, M. E. 2011, Nature, 469, 374

Kormendy, J. & Djorgovski, S. 1989, Annual Review of Astronomy and Astrophysics, 27, 235

Kormendy, J. & Fisher, D. B. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 396, Formation and Evolution of Galaxy Disks, ed. J. G. Funes & E. M. Corsini, 297

Kormendy, J. & Ho, L. C. 2013, Annual Review of Astronomy and Astrophysics, 51, 511

Kormendy, J. & Kennicutt, Jr., R. C. 2004, Annual Review of Astronomy and Astrophysics, 42, 603

Koss, M., Mushotzky, R., Veilleux, S., & Winter, L. 2010, Astrophysical Journal, Letters to the Editor, 716, L125

Krajnović, D., Cappellari, M., de Zeeuw, P. T., & Copin, Y. 2006, Monthly Notices of the RAS, 366, 787

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12 (USA: Curran Associates Inc.), 1097–1105

Kronberger, T., Kapferer, W., Schindler, S., & Ziegler, B. L. 2007, Astronomy and Astrophysics, 473, 761

Kronberger, T., Kapferer, W., Unterguggenberger, S., Schindler, S., & Ziegler, B. L. 2008, Astronomy and Astrophysics, 483, 783

Kroupa, P. 2001, Monthly Notices of the RAS, 322, 231

Kruk, S. J. et al. 2018, Monthly Notices of the RAS, 473, 4731

Lacey, C. & Cole, S. 1993, Monthly Notices of the RAS, 262, 627

Lackner, C. N. & Gunn, J. E. 2012, Monthly Notices of the RAS, 421, 2277

Lang, P. et al. 2014, Astrophysical Journal, 788, 11

Lange, R. et al. 2016, Monthly Notices of the RAS, 462, 1470

Laureijs, R. et al. 2011, arXiv e-prints, arXiv:1110.3193

Laurikainen, E., Salo, H., Athanassoula, E., Bosma, A., Buta, R., & Janz, J. 2013, Monthly Notices of the RAS, 430, 3489

Laurikainen, E., Salo, H., & Buta, R. 2005, Monthly Notices of the RAS, 362, 1319

Laurikainen, E., Salo, H., Buta, R., Knapen, J. H., & Comerón, S. 2010, Monthly Notices of the RAS, 405, 1089

Law, D. R. et al. 2016, Astronomical Journal, 152, 83

— 2015, Astronomical Journal, 150, 19

Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436

LeCun, Y., Bengio, Y., et al. 1995, The handbook of brain theory and neural networks, 3361, 1995

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. 1989, Neural computation, 1, 541

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al. 1998, Proceedings of the IEEE, 86, 2278

Leitherer, C. et al. 1999, Astrophysical Journal, Supplement Series, 123, 3

Lilly, S. J., Le Fevre, O., Crampton, D., Hammer, F., & Tresse, L. 1995, Astrophysical Journal, 455, 50

Lin, L. et al. 2007, Astrophysical Journal, Letters to the Editor, 660, L51

— 2004, Astrophysical Journal, Letters to the Editor, 617, L9

— 2008, Astrophysical Journal, 681, 232

Linde, A. D. 1982, Physics Letters B, 108, 389

Lintott, C. J. et al. 2008, Monthly Notices of the RAS, 389, 1179

Lisker, T. 2008, Astrophysical Journal, Supplement Series, 179, 319

López-Sanjuan, C. et al. 2013a, Astronomy and Astrophysics, 558, A135

— 2011, Astronomy and Astrophysics, 530, A20

— 2012, Astronomy and Astrophysics, 548, A7

— 2013b, Astronomy and Astrophysics, 553, A78

Lotz, J. M., Jonsson, P., Cox, T. J., Croton, D., Primack, J. R., Somerville, R. S., & Stewart, K. 2011, Astrophysical Journal, 742, 103

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, Monthly Notices of the RAS, 391, 1137

— 2010a, Monthly Notices of the RAS, 404, 590

— 2010b, Monthly Notices of the RAS, 404, 575

Lotz, J. M., Madau, P., Giavalisco, M., Primack, J., & Ferguson, H. C. 2006, Astrophysical Journal, 636, 592

Lotz, J. M., Primack, J., & Madau, P. 2004, Astronomical Journal, 128, 163

Lucey, M., Ting, Y.-S., Ramachandra, N. S., & Hawkins, K. 2020, arXiv e-prints, arXiv:2002.02961

Lupton, R., Blanton, M. R., Fekete, G., Hogg, D. W., O'Mullane, W., Szalay, A., & Wherry, N. 2004, Publications of the ASP, 116, 133

Lupton, R., Gunn, J. E., Ivezić, Z., Knapp, G. R., & Kent, S. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne, 269

Lupton, R. H., Ivezic, Z., Gunn, J. E., Knapp, G., Strauss, M. A., & Yasuda, N. 2002, in Proc. SPIE, Vol. 4836, Survey and Other Telescope Technologies and Discoveries, ed. J. A. Tyson & S. Wolff, 350–356

Lupton, R. H., Ivezić, Ž., Gunn, J. E., Knapp, G. R., & Strauss, M. A. 2012, The photo-lite draft, plus other notes at RHL's Web site `http://www.astro.princeton.edu/~rhl/photo-lite.pdf`

Lynden-Bell, D. 1967, Monthly Notices of the RAS, 136, 101

Mac Namee, B., Cunningham, P., Byrne, S., & Corrigan, O. I. 2002, Artificial intelligence in medicine, 24, 51

MacArthur, L. A., Courteau, S., & Holtzman, J. A. 2003, Astrophysical Journal, 582, 689

Marinacci, F., Fraternali, F., Binney, J., Nipoti, C., Ciotti, L., & Londrillo, P. 2012, in European Physical Journal Web of Conferences, Vol. 19, European Physical Journal Web of Conferences, 08008

Marinacci, F. et al. 2018, Monthly Notices of the RAS, 480, 5113

Mármol-Queraltó, E. et al. 2011, Astronomy and Astrophysics, 534, A8

Márquez, I., Masegosa, J., Durret, F., González Delgado, R. M., Moles, M., Maza, J., Pérez, E., & Roth, M. 2003, Astronomy and Astrophysics, 409, 459

Martin, C. L. 2005, Astrophysical Journal, 621, 227

Martin, G., Kaviraj, S., Devriendt, J. E. G., Dubois, Y., & Pichon, C. 2018, Monthly Notices of the RAS, 480, 2266

Meert, A., Vikram, V., & Bernardi, M. 2013, Monthly Notices of the RAS, 433, 1344

— 2015, Monthly Notices of the RAS, 446, 3943

Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, Astrophysical Journal, Supplement Series, 210, 3

Méndez-Abreu, J., Aguerri, J. A. L., Corsini, E. M., & Simonneau, E. 2008, Astronomy and Astrophysics, 478, 353

Michel-Dansac, L., Lambas, D. G., Alonso, M. S., & Tissera, P. 2008, Monthly Notices of the RAS, 386, L82

Mihos, J. C., Dubinski, J., & Hernquist, L. 1998, Astrophysical Journal, 494, 183

Mihos, J. C. & Hernquist, L. 1994, Astrophysical Journal, Letters to the Editor, 431, L9

— 1996, Astrophysical Journal, 464, 641

Miyazaki, S. et al. 2018, Publications of the ASJ, 70, S1

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Mo, H. J., Mao, S., & White, S. D. M. 1998, Monthly Notices of the RAS, 295, 319

Möllenhoff, C. & Heidt, J. 2001, Astronomy and Astrophysics, 368, 16

Monaghan, J. J. 1992, Annual Review of Astronomy and Astrophysics, 30, 543

Moorthy, B. K. & Holtzman, J. A. 2006, Monthly Notices of the RAS, 371, 583

Moreno, J., Torrey, P., Ellison, S. L., Patton, D. R., Bluck, A. F. L., Bansal, G., & Hernquist, L. 2015, Monthly Notices of the RAS, 448, 1107

Moreno, J. et al. 2019, Monthly Notices of the RAS, 485, 1320

Moster, B. P., Naab, T., & White, S. D. M. 2013, Monthly Notices of the RAS, 428, 3121

Müller-Sánchez, F., Nevin, R., Comerford, J. M., Davies, R. I., Privon, G. C., & Treister, E. 2018, Nature, 556, 345

Mundy, C. J., Conselice, C. J., Duncan, K. J., Almaini, O., Häußler, B., & Hartley, W. G. 2017, Monthly Notices of the RAS, 470, 3507

Murali, C., Katz, N., Hernquist, L., Weinberg, D. H., & Davé, R. 2002, Astrophysical Journal, 571, 1

Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, Publications of the ASP, 131, 118002

Naab, T. & Burkert, A. 2003, Astrophysical Journal, 597, 893

Naab, T. et al. 2014, Monthly Notices of the RAS, 444, 3357

Naiman, J. P. et al. 2018, Monthly Notices of the RAS, 477, 1206

Nair, P. B. & Abraham, R. G. 2010, Astrophysical Journal, Supplement Series, 186, 427

Nakoneczny, S., Bilicki, M., Solarz, A., Pollo, A., Maddox, N., Spiniello, C., Brescia, M., & Napolitano, N. R. 2019, Astronomy and Astrophysics, 624, A13

Navarro, J. F. & White, S. D. M. 1994, Monthly Notices of the RAS, 267, 401

Negroponte, J. & White, S. D. M. 1983, Monthly Notices of the RAS, 205, 1009

Nelson, D. et al. 2019a, Monthly Notices of the RAS, 490, 3234

— 2018, Monthly Notices of the RAS, 475, 624

— 2019b, Computational Astrophysics and Cosmology, 6, 2

Nevin, R., Blecha, L., Comerford, J., & Greene, J. 2019, Astrophysical Journal, 872, 76

Noguchi, M. 1991, Monthly Notices of the RAS, 251, 360

Ntampaka, M. et al. 2019, Astrophysical Journal, 876, 82

Ocvirk, P., Pichon, C., & Teyssier, R. 2008, Monthly Notices of the RAS, 390, 1326

Oemler, Augustus, J. 1974, Astrophysical Journal, 194, 1

Oh, S. et al. 2016, Astrophysical Journal, 832, 69

Oke, J. B. & Gunn, J. E. 1983, Astrophysical Journal, 266, 713

Oman, K. A. et al. 2015, Monthly Notices of the RAS, 452, 3650

Omand, C. M. B., Balogh, M. L., & Poggianti, B. M. 2014, Monthly Notices of the RAS, 440, 843

Oppenheimer, B. D., Davé, R., Kereš, D., Fardal, M., Katz, N., Kollmeier, J. A., & Weinberg, D. H. 2010, Monthly Notices of the RAS, 406, 2325

Padmanabhan, N. et al. 2008, Astrophysical Journal, 674, 1217

Pan, H.-A. et al. 2019, Astrophysical Journal, 881, 119

Patton, D. R. & Atfield, J. E. 2008, Astrophysical Journal, 685, 235

Patton, D. R., Ellison, S. L., Simard, L., McConnachie, A. W., & Mendel, J. T. 2011, Monthly Notices of the RAS, 412, 591

Patton, D. R., Grant, J. K., Simard, L., Pritchet, C. J., Carlberg, R. G., & Borne, K. D. 2005, Astronomical Journal, 130, 2043

Patton, D. R. et al. 2002, Astrophysical Journal, 565, 208

Patton, D. R., Qamar, F. D., Ellison, S. L., Bluck, A. F. L., Simard, L., Mendel, J. T., Moreno, J., & Torrey, P. 2016, Monthly Notices of the RAS, 461, 2589

Patton, D. R., Torrey, P., Ellison, S. L., Mendel, J. T., & Scudder, J. M. 2013, Monthly Notices of the RAS, 433, L59

Patton, D. R. et al. 2020, arXiv e-prints, arXiv:2003.00289

Pawlik, M. M., Wild, V., Walcher, C. J., Johansson, P. H., Villforth, C., Rowlands, K., Mendez-Abreu, J., & Hewlett, T. 2016, Monthly Notices of the RAS, 456, 3032

Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., & van der Tak, F. F. S. 2019, Astronomy and Astrophysics, 626, A49

Peebles, P. J. E. 1965, Astrophysical Journal, 142, 1317

Peebles, P. J. E. & Yu, J. T. 1970, Astrophysical Journal, 162, 815

Peletier, R. F. et al. 2007, Monthly Notices of the RAS, 379, 445

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, Astronomical Journal, 124, 266

— 2010, Astronomical Journal, 139, 2097

Penzias, A. A. & Wilson, R. W. 1965, Astrophysical Journal, 142, 419

Percival, W. J. et al. 2001, Monthly Notices of the RAS, 327, 1297

Perez, J., Michel-Dansac, L., & Tissera, P. B. 2011, Monthly Notices of the RAS, 417, 580

Perlmutter, S. 2003, Physics Today, 56, 53

Pier, J. R., Munn, J. A., Hindsley, R. B., Hennessy, G. S., Kent, S. M., Lupton, R. H., & Ivezić, Ž. 2003, Astronomical Journal, 125, 1559

Pillepich, A. et al. 2018a, Monthly Notices of the RAS, 475, 648

— 2018b, Monthly Notices of the RAS, 473, 4077

Pilyugin, L. S., Grebel, E. K., Zinchenko, I. A., Vílchez, J. M., Sakhibov, F., Nefedyev, Y. A., & Berczik, P. P. 2020, Astronomy and Astrophysics, 634, A26

Planck Collaboration et al. 2016, Astronomy and Astrophysics, 594, A1

— 2018, arXiv e-prints, arXiv:1807.06209

Press, W. H. & Schechter, P. 1974, Astrophysical Journal, 187, 425

Radivojac, P., Chawla, N. V., Dunker, A. K., & Obradovic, Z. 2004, Journal of Biomedical Informatics, 37, 224

Rampazzo, R., Plana, H., Amram, P., Bagarotto, S., Boulesteix, J., & Rosado, M. 2005, Monthly Notices of the RAS, 356, 1177

Rees, M. J. & Ostriker, J. P. 1977, Monthly Notices of the RAS, 179, 541

Refregier, A. 2003, Annual Review of Astronomy and Astrophysics, 41, 645

Refregier, A., Amara, A., Kitching, T. D., Rassat, A., Scaramella, R., & Weller, J. 2010, arXiv e-prints, arXiv:1001.0061

Ribli, D., Pataki, B. Á., & Csabai, I. 2019a, Nature Astronomy, 3, 93

Ribli, D., Pataki, B. Á., Zorrilla Matilla, J. M., Hsu, D., Haiman, Z., & Csabai, I. 2019b, Monthly Notices of the RAS, 490, 1843

Riess, A. G. et al. 1998, Astronomical Journal, 116, 1009

Robertson, B., Bullock, J. S., Cox, T. J., Di Matteo, T., Hernquist, L., Springel, V., & Yoshida, N. 2006, Astrophysical Journal, 645, 986

Robitaille, T. P. 2011, Astronomy and Astrophysics, 536, A79

Robotham, A. S. G. et al. 2014, Monthly Notices of the RAS, 444, 3986

Robotham, A. S. G., Taranu, D. S., Tobar, R., Moffett, A., & Driver, S. P. 2017, Monthly Notices of the RAS, 466, 1513

Rodriguez-Gomez, V. et al. 2015, Monthly Notices of the RAS, 449, 49

— 2017, Monthly Notices of the RAS, 467, 3083

— 2019, Monthly Notices of the RAS, 483, 4140

Rubin, V. C. & Ford, Jr., W. K. 1970, Astrophysical Journal, 159, 379

Rubin, V. C., Ford, W. K. J., & . Thonnard, N. 1980, Astrophysical Journal, 238, 471

Rupke, D. S., Veilleux, S., & Sanders, D. B. 2005a, Astrophysical Journal, 632, 751

— 2005b, Astrophysical Journal, Supplement Series, 160, 115

Rupke, D. S. N., Kewley, L. J., & Barnes, J. E. 2010a, Astrophysical Journal, Letters to the Editor, 710, L156

Rupke, D. S. N., Kewley, L. J., & Chien, L. H. 2010b, Astrophysical Journal, 723, 1255

Saintonge, A. et al. 2016, Monthly Notices of the RAS, 462, 1749

Sales, L. V., Navarro, J. F., Theuns, T., Schaye, J., White, S. D. M., Frenk, C. S., Crain, R. A., & Dalla Vecchia, C. 2012, Monthly Notices of the RAS, 423, 1544

Sánchez, S. F. et al. 2012, Astronomy and Astrophysics, 538, A8

Sandage, A. 1961, The Hubble atlas of galaxies (Carnegie Institution, Washington)

— 1975, Classification and Stellar Content of Galaxies Obtained from Direct Photography, ed. A. Sandage, M. Sandage, & J. Kristian (the University of Chicago Press), 1

Satyapal, S., Ellison, S. L., McAlpine, W., Hickox, R. C., Patton, D. R., & Mendel, J. T. 2014, Monthly Notices of the RAS, 441, 1297

Sawala, T. et al. 2016, Monthly Notices of the RAS, 457, 1931

Schade, D., Lilly, S. J., Crampton, D., Hammer, F., Le Fevre, O., & Tresse, L. 1995, Astrophysical Journal, Letters to the Editor, 451, L1

Schaye, J. et al. 2015, Monthly Notices of the RAS, 446, 521

— 2010, Monthly Notices of the RAS, 402, 1536

Schiminovich, D. et al. 2007, Astrophysical Journal, Supplement Series, 173, 315

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, Astrophysical Journal, 500, 525

Schmidt, M. 1959, Astrophysical Journal, 129, 243

Scudder, J. M., Ellison, S. L., Torrey, P., Patton, D. R., & Mendel, J. T. 2012, Monthly Notices of the RAS, 426, 549

Sellwood, J. A. 2013, Dynamics of Disks and Warps, ed. T. D. Oswalt & G. Gilmore, 923

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shapiro, K. L. et al. 2008, Astrophysical Journal, 682, 231

Sharp, R. et al. 2015, Monthly Notices of the RAS, 446, 1551

Sharples, R. M. et al. 2004, in Proc. SPIE, Vol. 5492, Ground-based Instrumentation for Astronomy, ed. A. F. M. Moorwood & M. Iye, 1179–1186

Silk, J. 1977, Astrophysical Journal, 211, 638

Silverman, J. D. et al. 2011, Astrophysical Journal, 743, 2

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, Astrophysical Journal, Supplement Series, 196, 11

Simard, L. et al. 2002, Astrophysical Journal, Supplement Series, 142, 1

Simmons, B. D. et al. 2017, Monthly Notices of the RAS, 464, 4420

Smee, S. A. et al. 2013, Astronomical Journal, 146, 32

Smith, B. J., Struck, C., Hancock, M., Appleton, P. N., Charmandaris, V., & Reach, W. T. 2007, Astronomical Journal, 133, 791

Smith, J. A. et al. 2002, Astronomical Journal, 123, 2121

Smoot, G. F. et al. 1992, Astrophysical Journal, Letters to the Editor, 396, L1

Snyder, G. F., Rodriguez-Gomez, V., Lotz, J. M., Torrey, P., Quirk, A. C. N., Hernquist, L., Vogelsberger, M., & Freeman, P. E. 2019, Monthly Notices of the RAS, 486, 3702

Snyder, G. F. et al. 2015, Monthly Notices of the RAS, 454, 1886

Sol Alonso, M., Michel-Dansac, L., & Lambas, D. G. 2010, Astronomy and Astrophysics, 514, A57

Somerville, R. S. & Davé, R. 2015, Annual Review of Astronomy and Astrophysics, 53, 51

Sparre, M. & Springel, V. 2016, Monthly Notices of the RAS, 462, 2418

— 2017, Monthly Notices of the RAS, 470, 3946

Springel, V. 2000, Monthly Notices of the RAS, 312, 859

— 2010, Monthly Notices of the RAS, 401, 791

Springel, V., Di Matteo, T., & Hernquist, L. 2005a, Monthly Notices of the RAS, 361, 776

Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137

Springel, V. & Hernquist, L. 2003, Monthly Notices of the RAS, 339, 289

Springel, V. et al. 2018, Monthly Notices of the RAS, 475, 676

— 2005b, Nature, 435, 629

Steinacker, J., Baes, M., & Gordon, K. D. 2013, Annual Review of Astronomy and Astrophysics, 51, 63

Steinmetz, M. & Navarro, J. F. 2002, New Astronomy, 7, 155

Stewart, K. R., Brooks, A. M., Bullock, J. S., Maller, A. H., Diemand, J., Wadsley, J., & Moustakas, L. A. 2013, Astrophysical Journal, 769, 74

Stoughton, C. et al. 2002, Astronomical Journal, 123, 485

Strateva, I. et al. 2001, Astronomical Journal, 122, 1861

Strauss, M. A. et al. 2002, Astronomical Journal, 124, 1810

Strickland, D. K. & Heckman, T. M. 2009, Astrophysical Journal, 697, 2030

Tacchella, S. et al. 2019, Monthly Notices of the RAS, 487, 5416

Tal, T. & van Dokkum, P. G. 2011, Astrophysical Journal, 731, 89

Taranu, D. S., Dubinski, J., & Yee, H. K. C. 2013, Astrophysical Journal, 778, 61

Taylor, P., Federrath, C., & Kobayashi, C. 2018, Monthly Notices of the RAS, 479, 141

Teimoorinia, H., Bluck, A. F. L., & Ellison, S. L. 2016, Monthly Notices of the RAS, 457, 2086

Teimoorinia, H. & Ellison, S. L. 2014, Monthly Notices of the RAS, 439, 3526

Teimoorinia, H., Kavelaars, J. J., Gwyn, S., Durand, D., Rolston, K., & Ouellette, A. 2020a, arXiv e-prints, arXiv:2003.01785

Teimoorinia, H., Toyonaga, R. D., Fabbro, S., & Bottrell, C. 2020b, Publications of the ASP, 132, 044501

Thorp, M. D., Ellison, S. L., Simard, L., Sánchez, S. F., & Antonio, B. 2019, Monthly Notices of the RAS, 482, L55

Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. G. Larson, D. Campbell, Yale University Observatory, 401

Toomre, A. & Toomre, J. 1972, Astrophysical Journal, 178, 623

Torrey, P., Cox, T. J., Kewley, L., & Hernquist, L. 2012a, Astrophysical Journal, 746, 108

Torrey, P., Hopkins, P. F., Faucher-Giguère, C.-A., Vogelsberger, M., Quataert, E., Kereš, D., & Murray, N. 2017, Monthly Notices of the RAS, 467, 2301

Torrey, P. et al. 2015, Monthly Notices of the RAS, 447, 2753

Torrey, P., Vogelsberger, M., Genel, S., Sijacki, D., Springel, V., & Hernquist, L. 2014, Monthly Notices of the RAS, 438, 1985

Torrey, P., Vogelsberger, M., Sijacki, D., Springel, V., & Hernquist, L. 2012b, Monthly Notices of the RAS, 427, 2224

Tucker, D. L. et al. 2006, Astronomische Nachrichten, 327, 821

van de Voort, F., Davis, T. A., Kereš, D., Quataert, E., Faucher-Giguère, C.-A., & Hopkins, P. F. 2015, Monthly Notices of the RAS, 451, 3269

van den Bergh, S. 1998, Galaxy Morphology and Classification

Veilleux, S. et al. 2013, Astrophysical Journal, 776, 27

Vika, M., Bamford, S. P., Häußler, B., Rojas, A. L., Borch, A., & Nichol, R. C. 2013, Monthly Notices of the RAS, 435, 623

Vogelsberger, M., Genel, S., Sijacki, D., Torrey, P., Springel, V., & Hernquist, L. 2013, Monthly Notices of the RAS, 436, 3031

Vogelsberger, M. et al. 2014a, Nature, 509, 177

— 2014b, Monthly Notices of the RAS, 444, 1518

Vogelsberger, M., Sijacki, D., Kereš, D., Springel, V., & Hernquist, L. 2012, Monthly Notices of the RAS, 425, 3024

Wake, D. A. et al. 2017, Astronomical Journal, 154, 86

Walmsley, M., Ferguson, A. M. N., Mann, R. G., & Lintott, C. J. 2019, Monthly Notices of the RAS, 483, 2968

Weinberg, D. H., Hernquist, L., & Katz, N. 1997, Astrophysical Journal, 477, 8

Weinberger, R. et al. 2018, Monthly Notices of the RAS, 479, 4056

Wen, Z. Z., Zheng, X. Z., & An, F. X. 2014, Astrophysical Journal, 787, 130

Westfall, K. B. et al. 2019, Astronomical Journal, 158, 231

White, S. D. M. & Frenk, C. S. 1991, Astrophysical Journal, 379, 52

White, S. D. M., Frenk, C. S., & Davis, M. 1983, Astrophysical Journal, Letters to the Editor, 274, L1

White, S. D. M. & Rees, M. J. 1978, Monthly Notices of the RAS, 183, 341

Willett, K. W. et al. 2013, Monthly Notices of the RAS, 435, 2835

— 2015, Monthly Notices of the RAS, 449, 820

Wilson, G., Kaiser, N., Luppino, G. A., & Cowie, L. L. 2001, Astrophysical Journal, 555, 572

Wong, K. C. et al. 2011, Astrophysical Journal, 728, 119

Woo, J., Dekel, A., Faber, S. M., & Koo, D. C. 2015, Monthly Notices of the RAS, 448, 237

Woo, J.-H., Son, D., & Bae, H.-J. 2017, Astrophysical Journal, 839, 120

Yang, Y. et al. 2008, Astronomy and Astrophysics, 477, 789

York, D. G. et al. 2000, Astronomical Journal, 120, 1579

Zel'Dovich, Y. B. 1970, Astronomy and Astrophysics, 500, 13

Zeldovich, Y. B. 1972, Monthly Notices of the RAS, 160, 1P

Zschaechner, L. K. et al. 2016, Astrophysical Journal, 832, 142

Zubko, V., Dwek, E., & Arendt, R. G. 2004, Astrophysical Journal, Supplement Series, 152, 211

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Appendix A

Stripe 82: measurement uncertainties

In this appendix to Chapter 2, I compute the photometric uncertainties on several model parameters both as (1) a reference and (2) to illustrate the superior constraints on measurements using Stripe 82 images with respect to Legacy. The uncertainties I report are propagated uncertainties from the local mean sky measurement, $\sigma_{\langle \mathrm{sky} \rangle}$, and the GIM2D parametric uncertainties computed by marginally sampling the posterior probability distribution for each parameter about the best-fitting `n4` model. Uncertainties from covariances between model parameters are not accounted for, however these should be more or less similar in each dataset given that the same decomposition routines and models are used in the comparison. The catalogs include all of the necessary information needed to reproduce these uncertainty estimates.

Figure A.1 compares the u (lower row) and r (upper row) band total magnitude uncertainties, $\sigma(\mathrm{m}_{x,\mathrm{galaxy}})$, in Stripe 82 (left panels, magenta curves) and Legacy (right panels, cyan curves). For visual impression, the magenta curves for Stripe 82 are redrawn on the right panels. Background greyscale shows the 2D histogram of computed uncertainties. Curves show the median (solid, thick) and 16-84 percentile range (dashed). From Figure A.1 it is clear that the constraints on total magnitudes are substantially better in the deep co-adds. In the u-band, for example, we can see that a galaxy with $m_{u\mathrm{galaxy}} \approx 19$ in Stripe 82 has the same uncertainty (around 0.03 mag) as one at $m_{u\mathrm{galaxy}} \approx 16$ in the single-epoch Legacy images.

Figure A.2 shows the photometric uncertainties for the bulge, $\sigma(m_{r,\mathrm{bulge}})$, and disc, $\sigma(m_{r,\mathrm{disc}})$, component magnitudes and the bulge-to-total ratio, $\sigma(B/T)_r$ in the

r-band. The bottom right panel shows that Legacy B/T uncertainties are small out to $m_{r,\mathrm{galaxy}} \approx 15.5$ mag – after which the uncertainties begin to rise steadily. Stripe 82, however, has median B/T uncertainties that are typically confined to $\sigma(B/T)_r \lesssim 0.01$ over the full magnitude range. This result is consistent with my assertion that the improved depth in Stripe 82 enables characterization of B/T in galaxies for which the photometry was insufficient to discriminate between bulge and disc light in Legacy. It also supports the claim that Stripe 82 should be more discriminating in statistical comparisons designed to show whether a two-component bulge+disc decomposition is favoured over a single-component fit – particularly for faint targets. However, there is an interesting offset in the median Stripe 82 B/T uncertainties relative to Legacy where $m_{r,\mathrm{galaxy}} \lesssim 15.5$ mag and which appears to increase with brightness. To test whether these high B/T uncertainties came from the Stripe 82 sky or GIM2D uncertainties I turned off the sky uncertainties in my calculations and found that the systematic persisted. The large scatter in the Stripe 82 B/T uncertainties of bright galaxies therefore comes from parametric GIM2D uncertainty in B/T.

To understand what might be driving the inflated scatter and corresponding offset in Stripe 82 B/T parametric uncertainties for bright sources, I inspected the Stripe 82 and Legacy decomposition mosaics of objects with $\sigma(B/T)_{r,\mathrm{S82}} > 0.025$ and $m_{r,\mathrm{S82}} < 14.5$ mag – finding 41 in Stripe 82 and 2 in Legacy (both targets from Legacy were also satisfied the cut in Stripe 82). The uncertainties among the Stripe 82 targets that met this criterion were also generally larger than the two from Legacy. The mosaics of these sources revealed that they all had exceptionally asymmetric isophotes. Many were barred. Some were edge-on discs with long dust lanes. Others were spiral galaxies with strong twists and whose inner structures differed greatly in axis ratio or isophotal position angle from their outer structures. But all also had visible bulges. It might be that for these bright, asymmetric objects the bulge model is bouncing between fitting the bulge itself and fitting what the disc cannot. In other words, the bulge model is trying to compensate for the inadequacy of the disc model to the task of fitting a highly asymmetric disc or other extended structure.

Following this intuition, the reason that the uncertainty is higher in Stripe 82 is because the outer structure has higher S/N and therefore greater weight in the fit than in Legacy. If there are then twists in this outer structure then the bulge has to try to find a saddle point in likelihood space between properly modelling the bulge and modelling part of that outer structure – driving up its parametric uncertainty relative to the Legacy images. I have not designed an experiment to explicitly test

this assertion. Whether it is correct or not, the inflated B/T uncertainties at the bright end of Stripe 82 propagate into the corresponding bulge and disc uncertainty calculations shown in the upper two rows of Figure A.2. Ultimately, this systematic highlights the limitations of modelling highly asymmetric structures with symmetric models and the value of more sophisticated (but also more sensitive) models which can can both fit galaxies with complex isophotes *and* return realistic structural estimates (e.g., Ciambur 2015). By finding inventive ways to inject priors into these models and find reasonable initial conditions, they may soon be scalable to large surveys.

The upper two rows of Figure A.2 show the bulge, $\sigma(m_{r,\mathrm{bulge}})$, and disc, $\sigma(m_{r,\mathrm{disc}})$, component magnitudes and the bulge-to-total ratio, $\sigma(B/T)_r$ in the r-band. Again, in each case, we see the mark made by imaging depth on the Stripe 82 uncertainties. In general, the uncertainties on component magnitudes are much larger than for total magnitudes – as expected since each presents a fraction of the whole. Taking $\sigma_{r,\mathrm{tol}} = 0.1$ magnitudes ($\sim 10\%$ flux difference) as a toy tolerance, disc magnitudes in Stripe 82 are robust out to $m_{r,\mathrm{disc}} \approx 20$ mag, whereas they exceed the tolerance at $m_{r,\mathrm{disc}} \approx 18$ mag with Legacy photometry. Similarly, the limit for bulges is extended from $m_{r,\mathrm{bulge}} \approx 17.8$ mag to $m_{r,\mathrm{bulge}} \approx 19.8$ mag. The ~ 2 magnitude improvement in the limiting magnitude for a given tolerance generally holds for tolerances less than $\sigma_{r,\mathrm{tol}} = 0.2$ magnitudes.

I caution that one should be careful not to over-interpret these uncertainties. These uncertainties are the statistical uncertainties *given* an assumed model. Systematics that may arise from assuming models that may not be suited to real galaxies may not be captured (as highlighted in the discussion of the B/T uncertainties). One may interpret the uncertainties I report as encoding the probabilities that a new measurement made using the same model and optimization routine in an image of similar quality would achieve similar results.

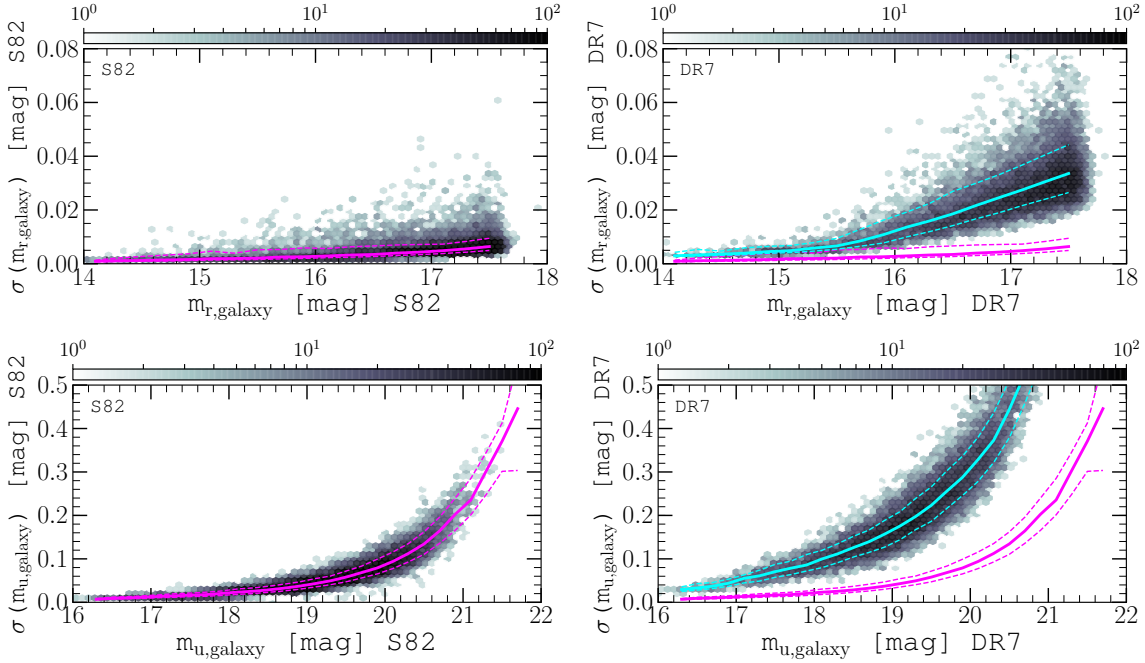

Figure A.1 Comparison of photometric total magnitude uncertainties, $\sigma^2(\mathrm{m}_{x,\mathrm{galaxy}}) = \sigma^2_{\langle \mathrm{sky} \rangle} + \sigma^2_{\mathrm{GIM2D}}$, for Stripe 82 (left panels, magenta curves) and Legacy (right panels, cyan curves). Upper and lower rows show the r and u band magnitude uncertainties. Stripe 82 curves are redrawn on the Legacy panels for visual impression. It is important to note the difference in scale between the upper and lower rows.

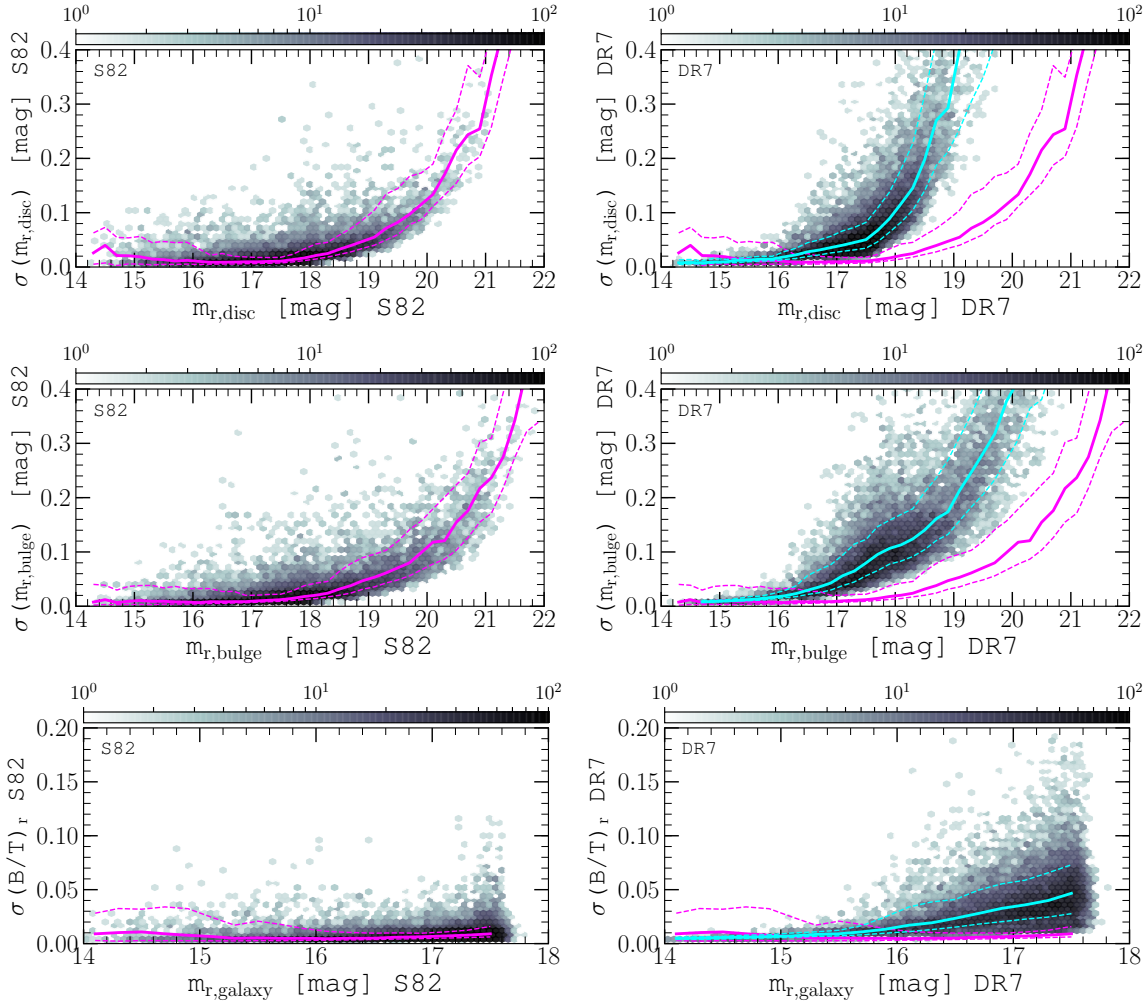

Figure A.2 Comparison of photometric component magnitude and bulge-to-total fraction uncertainties for Stripe 82 (left panels, magenta curves) and Legacy (right panels, cyan curves) in the r-band. Stripe 82 curves are redrawn on the Legacy panels for visual impression.

Appendix B

Stripe 82: catalog structure and schema

Table B.1 is an example table schema for the decomposition catalogs presented as part of Chapter 2. Twelve tables were released in total – one for each unique combination of model and $Xr-$band pairing[1]. Tables for fits with the same decomposition model have identical schema. Tables in which structural parameters are fixed to results from gr-band decompositions include those forced values. A user may refer to Table 2.1 to determine whether structural parameters in a given table are measured independently or fixed. One must again recall that the decompositions in each band are always performed pairwise with the $r-$band. The structural measurements in a given table are either: (1) covariant with the r-band structures (as in the case of Sérsic index in the `ps` fits); or (2) fixed to the results of the simultaneous $gr-$band decompositions and as such are never *truly* independent in any given band.

The `ps`, `n4`, and `fn` tables are identical in structure with a few exceptions. First, the `n4` tables contain a P_{pS} parameter. Second, the `fn` tables contain both P_{pS} and P_{n4} parameters. Each catalog also includes an estimate of the local sky background (relative to the pre-subtracted SExtractor background pixel mode in the full frame) and its uncertainty around each galaxy in each band. As tabulated, the sky uncertainties are normalized by the square root of the number of pixels used to estimate the sky level. These can be converted to calibrated sky surface brightness uncertainties,

[1]All Stripe 82 morphology tables can be found at `http://orca.phys.uvic.ca/~cbottrel/share/Stripe82/Catalogs/`

$\sigma_{\rm sky}$ (in magnitudes/arcsecond^2), as follows:

$$\sigma_{\rm sky} = -2.5\log_{10}\left(\frac{\texttt{skySigNorm}\times\sqrt{\texttt{skyNpix}}}{\texttt{pixel\_scale}^2}\right) + 30.0 \qquad \text{(B.1)}$$

where `pixel_scale` is the pixel scale, 0.396127 arcsec/pixel, of the SDSS camera. The local sky backgrounds themselves (`db` in the tables) are the relative offset between the global sky in the frame measured by SExtractor (which is then subtracted from the image) and the gim2d local estimate around the target galaxy. The example schema in Table B.1 is for a *gr* decomposition table with the `n4` model.

Table B.1: Stripe 82 galaxy morphology catalog example schema for the `n4` simultaneous *gr* band decompositions (online supplementary information). The `fn` tables are identical to the `n4` tables but include the additional P_{n4} statistic. The single-component `ps` model tables contain neither P_{pS} nor P_{n4} statistics and columns which pertained to the discs in the two-component tables are either NULL or are filled with nonsensical values in the `ps` tables: disc position angle, scale length, inclination and their associated uncertainties. Bulge-to-total fractions are also meaningless in the `ps` tables. Rest-frame quantities assume $(H_0, \Omega_m, \Omega_\Lambda) = (70$ km/s/Mpc, 0.3, 0.7). Absolute magnitudes incorporate $k-$corrections computed using `kcorrect` version 4.2 (Blanton & Roweis, 2007).

Column	Type	Units	Description
s82_objID	varchar(25)	n/a	Unique Stripe 82 object ID
objID	varchar(25)	n/a	Matched SDSS DR7 object ID
ra	float	degrees	Right ascension
decl	float	degrees	Declination
run	int(11)	n/a	Stripe 82 run: S-106, N-206
rerun	int(11)	n/a	Processing/calibration identifier
camcol	int(11)	n/a	CCD camera column
field	int(11)	n/a	Stripe 82 field
ID	int(11)	n/a	Object ID within given field
petroMag_r	float	mag	SDSS Petrosian magnitude
extinction_r	float	mag	Galactic extinction
z	float	n/a	Spectroscopic redshift
sciim_nx	int(11)	pixels	Size of science cut-out x
sciim_ny	int(11)	pixels	Size of science cut-out y
prchost	varchar(15)	n/a	Processing host (decomposition)

Continued on next page

Table B.1 – *Continued from previous page*

Column	Type	Units	Description
start_time	datetime	n/a	Processing start-time
end_time	datetime	n/a	Processing end-tim
dfm_r	float	counts/s	(16 - 50)% range f
f_r	float	counts/s	Galaxy total flux
dfp_r	float	counts/s	(50 - 84)% range f
dg2dmagm_r	float	mag	(16 - 50)% range g2dmag
g2dmag_r	float	mag	Galaxy apparent magnitude
dg2dmagp_r	float	mag	(50 - 84)% range g2dmag
dbtm_r	float	n/a	(16 - 50)% range bt
bt_r	float	n/a	Bulge-to-total fraction
dbtp_r	float	n/a	(50 - 84)% range bt
drem	float	arcsec	(16 - 50)% range re
re	float	arcsec	Bulge effective radius
drep	float	arcsec	(50 - 84)% range re
dem	float	n/a	(16 - 50)% range e
e	float	n/a	Bulge ellipticity
dep	float	n/a	(50 - 84)% range e
dphibm	float	degrees	(16 - 50)% range phib
phib	float	degrees	Bulge position angle
dphibp	float	degrees	(50 - 84)% range phib
drdm	float	arcsec	(16 - 50)% range rd
rd	float	arcsec	Disc scale length
drdp	float	arcsec	(50 - 84)% range rd
didm	float	degrees	(16 - 50)% range incd
incd	float	degrees	Disc inclination
didp	float	degrees	(50 - 84)% range incd
dphidm	float	degrees	(16 - 50)% range phib
phid	float	degrees	Disc position angle
dphidp	float	degrees	(50 - 84)% range phid
ddxm_g	float	pixels	(16 - 50)% range dx
dx_g	float	pixels	Centroid offset from PHOTO x
ddxp_g	float	pixels	(50 - 84)% range dx

Continued on next page

Table B.1 – *Continued from previous page*

Column	Type	Units	Description
ddym_g	float	pixels	(16 - 50)% range dy
dy_g	float	pixels	Centroid offset from PHOTO y
ddyp_g	float	pixels	(50 - 84)% range dy
ddbm_r	float	counts/s	(16 - 50)% range db
db_r	float	counts/s	Residual background
ddbp_r	float	counts/s	(50 - 84)% range db
dnm	float	n/a	(16 - 50)% range n
n	float	n/a	Sersic Index
dnp	float	n/a	(50 - 84)% range n
rhalf_r	float	arcsec	Galaxy half-light radius (HLR)
rchi2_r	float	n/a	Reduced CHI2 statistic
c1_r	float	n/a	Concentration within 1 HLR
c2_r	float	n/a	Concentration within 2 HLR
c3_r	float	n/a	Concentration within 3 HLR
c4_r	float	n/a	Concentration within 4 HLR
a_r	float	n/a	Asymmetry parameter
da_r	float	n/a	Uncertainty in asymmetry
az2_1_r	float	n/a	Az for n=2 within 1 HLR
az2_2_r	float	n/a	Az for n=2 within 2 HLR
az3_1_r	float	n/a	Az for n=3 within 1 HLR
az3_2_r	float	n/a	Az for n=3 within 2 HLR
az5_1_r	float	n/a	Az for n=5 within 1 HLR
az5_2_r	float	n/a	Az for n=5 within 2 HLR
dz_r	float	n/a	Dz parameter
rt1_1_r	float	n/a	Res. asymmetry R_T in 1 HLR
rt1_2_r	float	n/a	Res. asymmetry R_T in 2 HLR
rt1_3_r	float	n/a	Res. asymmetry R_T in 3 HLR
ra1_1_r	float	n/a	Res. asymmetry R_A in 1 HLR
ra1_2_r	float	n/a	Res. asymmetry R_A in 2 HLR
ra1_3_r	float	n/a	Res. asymmetry R_A in 3 HLR
prcflag	int(11)	n/a	Processing flag (0: converged)
rd_kpc	float	kpc	Disc scale length [kpc]

Continued on next page

Table B.1 – *Continued from previous page*

Column	Type	Units	Description
re_kpc	float	kpc	Bulge effective radius [kpc]
rhalf_kpc_r	float	kpc	Galaxy HLR [kpc]
petroR50_r	float	arcsec	Petrosian radius R50
mu50_r	float	mag/arcsec2	Avg. surface brightness in R50
Vmax	float	Mpc3	Volume correction
Mr_galaxy	float	mag	Galaxy absolute magnitude
Mr_bulge	float	mag	Bulge absolute magnitude
Mr_disk	float	mag	Disc absolute magnitude
Mg_galaxy	float	mag	Galaxy absolute magnitude
Mg_bulge	float	mag	Bulge absolute magnitude
Mg_disk	float	mag	Disc absolute magnitude
extinction_g	float	mag	Galactic extinction
petroMag_g	float	mag	SDSS Petrosian magnitude
dfm_g	float	counts/s	(16 - 50)% range f
f_g	float	counts/s	Galaxy total flux
dfp_g	float	counts/s	(50 - 84)% range f
dbtm_g	float	n/a	(16 - 50)% range bt
bt_g	float	n/a	Bulge-to-total fraction
dbtp_g	float	n/a	(50 - 84)% range bt
dg2dmagm_g	float	mag	(16 - 50)% range g2dmag
g2dmag_g	float	mag	Galaxy apparent magnitude
dg2dmagp_g	float	mag	(50 - 84)% range g2dmag
rhalf_g	float	arcsec	Galaxy half-light radius (HLR)
rhalf_kpc_g	float	kpc	Galaxy HLR [kpc]
rchi2_g	float	n/a	Reduced CHI2 statistic
ddbm_g	float	counts/s	(16 - 50)% range db
db_g	float	counts/s	Residual background
ddbp_g	float	counts/s	(50 - 84)% range db
c1_g	float	n/a	Concentration within 1 HLR
c2_g	float	n/a	Concentration within 2 HLR
c3_g	float	n/a	Concentration within 3 HLR
c4_g	float	n/a	Concentration within 4 HLR

Continued on next page

Table B.1 – *Continued from previous page*

Column	Type	Units	Description
petroR50_g	float	arcsec	Petrosian radius R50
mu50_g	float	mag/arcsec2	Avg. surface brightness in R50
a_g	float	n/a	Asymmetry
da_g	float	n/a	Uncertainty in asymmetry
az2_1_g	float	n/a	Az for n=2 within 1 HLR
az2_2_g	float	n/a	Az for n=2 within 2 HLR
az3_1_g	float	n/a	Az for n=3 within 1 HLR
az3_2_g	float	n/a	Az for n=3 within 2 HLR
az5_1_g	float	n/a	Az for n=5 within 1 HLR
az5_2_g	float	n/a	Az for n=5 within 2 HLR
dz_g	float	n/a	Dz parameter
rt1_1_g	float	n/a	Res. asymmetry R_T in 1 HLR
rt1_2_g	float	n/a	Res. asymmetry R_T in 2 HLR
rt1_3_g	float	n/a	Res. asymmetry R_T in 3 HLR
ra1_1_g	float	n/a	Res. asymmetry R_A in 1 HLR
ra1_2_g	float	n/a	Res. asymmetry R_A in 2 HLR
ra1_3_g	float	n/a	Res. asymmetry R_A in 3 HLR
specclass	int(11)	n/a	Spectroscopic classification
ddxm_r	float	pixels	(16 - 50)% range dx
dx_r	float	pixels	Centroid offset from PHOTO x
ddxp_r	float	pixels	(50 - 84)% range dx
ddym_r	float	pixels	(16 - 50)% range dy
dy_r	float	pixels	Centroid offset from PHOTO y
ddyp_r	float	pixels	(50 - 84)% range dy
crhalf_g	float	arcsec	Circular half-light radius
crhalf_r	float	arcsec	Circular half-light radius
P_pS	float	n/a	F-test statistic PpS
npxfit0	int(11)	n/a	# sky pixels used in fitting
npxfit1	int(11)	n/a	# target pixels used in fitting
skySigNorm_r	float	counts/s/pixel	Sky standard error
skySigNorm_g	float	counts/s/pixel	Sky standard error
skyNpix_g	float	n/a	# pixels used to measure sky

Continued on next page

Table B.1 – *Continued from previous page*

Column	Type	Units	Description
skyNpix_r	float	n/a	# pixels used to measure sky

Appendix C

Merger stage predictions: single-band photometry results

This appendix to Chapter 3 provides the results of the singe-band photometry experiments from Chapter 3, Section 3.3.2. Figures C.1 and C.2 show the results for the single-channel handshake experiment in which the r-band and i-band images from PH, PHSR and PHFR datasets are used to train single-channel, colour-insensitive networks which are then applied to images of each type.

Figure C.1 Confusion matrices for the single-channel handshake experiment in which the r-band images from Ph, PhSR and PhFR datasets are used to train colour-insensitive networks which are then applied to images of each type. For reference, Figure 3.4 describes the information displayed by each individual confusion matrix in detail.

Figure C.2 Same as Figure C.2 but for networks trained and tested on i-band images from the Ph, PhSR and PhFR datasets.

Appendix D

Merger stage predictions: correlations between galaxy images

In this appendix to Chapter 3, I investigate the possibility that the snapshot sampling cadence for each interaction was too fine – resulting in galaxy images that could be strongly correlated (e.g. see Chapter 3, Figure 3.1). Strong correlations between images in neighbouring snapshots could result in overfitting to the training data and, consequently, loss of generalizability to test data from other datasets (e.g. train: PH, test: PHFR). Meanwhile, strong correlations between images in tests where a network is applied to test data from the same dataset (e.g. train: PH, test: PH) may lead to correlated training/test data and result in erroneously high test accuracies. Since the main investigation is focused on the sensitivity of network performance to realism, it is crucial to characterize (and preferably rule out) the sensitivity of the network performance to the snapshot selection cadence.

Figure D.1 shows PH (upper panels) and corresponding PHFR images (lower panels) for five neighbouring snapshots from the sampling of the fiducial G2G3e orbit 1 merger at a fixed camera angle. The snapshots are selected from near first apocenter ($t \approx 0.75$ Gyr in Figure 3.1) where the rate of morphological evolution would be expected to be at a minimum during the pair phase and, therefore, are most likely to be visually correlated. It should be noted that the smaller object in the images is not the companion (which is outside the field of view) but is a tidal dwarf that is produced in this particular interaction. The PH images in the upper panels of Figure D.1 show that there is still visual evolution in galaxy structure (both inner and outer) during this relatively calm part of the pair phase. The PHFR images in the lower panels are

even less visibly correlated due to the varying sky levels, resolution, and contamination by additional sources. Correlations between the images should therefore stand to be most problematic in the idealized images which do not incorporate any additional realism. However, since these visual assessments are subjective, I also devised a test which quantitatively investigates the possibility of correlations between images from neighbouring snapshots.

If images from neighbouring snapshots are strongly correlated, then randomly sampling the test data from a dataset (e.g. splitting Sm into training/validation/test data) will result in training, validation, and test data that are correspondingly correlated. There are two experiments which could be performed to investigate whether such correlations may be influencing the results of my other tests: (1) reduce the sampling cadence (e.g. discard every second snapshot from my initial selection) and (2) reserve all images from a full merger simulation for testing and train/validate on the remaining images. The problem with (1) is that this approach would reduce the data volume by an integer factor – which could lead to overfitting from sparseness in the dataset and make it impossible to assign any reduction in accuracy to either this sparseness or image correlations. Test (2) does not suffer from this problem because, for a given merger, it only reduces the data volume by relatively small factor of 1/23. Poor classification performance on the images from the merger that were reserved for testing (particularly for phases where the rate of morphological evolution is expected to be small) would confirm that the snapshot selection is too fine and yields correlations between images that would be affecting my results. The second test would simultaneously show whether a network is generalizable to images from a merger that it has never seen before (albeit, within the limitations outlined in Section 3.4.2). For these reasons, I carry out the second test.

Figure D.2 shows the results of my investigation into possible correlations between images from neighbouring snapshots. The investigation was performed using the Sm dataset. I removed all images from the fiducial G2G3e orbit 1 merger (Figure 3.1) from the Sm dataset and reserved them as a test set. I also removed images from the G2G3 merger in the mass ratio suite (which all use "e" orbit 1 initial conditions). This merger is not identical to the fiducial simulation (due to the chaotic nature of galaxy mergers) but has the same initial conditions and is consequently removed and discarded to eliminate any possibility that the training/validation data include images which may be correlated with the test images from the fiducial G2G3e orbit 1 merger simulation. I trained a network on the remaining data with a (70,30)%

training/validation split. Only the images for the fiducial merger were used as test data. The right panel of Figure D.2 shows the confusion matrix for this test. This test set contains no images with the Iso class because those images are drawn from separate isolated simulation runs as outlined in Section 3.2.1. Consequently, all elements in the first column are NaNs. The performance on the G2G3e orbit 1 test images, after removing all images from the training/validation data which might be correlated to these test images, is consistent with the results shown in the upper left panel of Figure 3.6. In other words, the network runs equally well on a merger for which it has seen seen zero images (this test), as mergers for which $\sim 70\%$ of their images are included in the training data (upper left panel of Figure 3.6). The results of this test demonstrate that possible correlations between images in neighbouring snapshots do not have a significant role in the successes of my networks.

The upper left panel of Figure D.2 shows the normalized classification scores as coloured bars for each snapshot in the G2G3e orbit 1 test images, averaged over all camera angles at each snapshot, $\overline{P}(X = \text{Class})$. The lower left panel shows the radial separation sequence for this merger with each selected snapshot and its true class indicated with coloured circles as in Figure 3.1. This classification sequence plot demonstrates the two main results of this appendix: (1) correlations between images from neighbouring snapshots is not a significant contributor to the accuracies reported in the main analyses (even near first apocenter, where such correlations would be expected to be strongest) and (2) the changes to classification scores as one approaches a temporal class boundary are continuous and not choppy or sporadic (e.g. between the Pair and Post-merger phase at $t \approx 1.54$ Gyr).

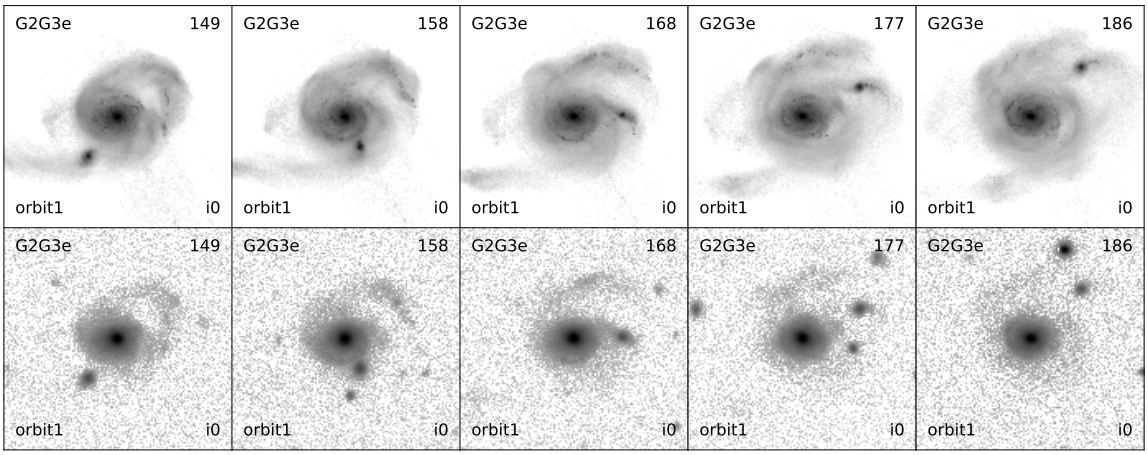

Figure D.1 Ph (top row) and corresponding PhFR (bottom row) images for five neighbouring snapshots at fixed camera angle from the sampling of the G2G3e orbit 1 merger near first apocenter ($t \approx 0.75$ Gyr in Figures 3.1 and D.2). If there were strong correlations between images from neighbouring snapshots, they would be most likely to occur here, at first apocenter in the merger sequence – since the rate of morphological evolution should be lowest relative to the more rapid changes at first pericenter and beyond second pericenter. Visually, inner and outer structures of the galaxy both evolve appreciably in this sequence of images. These visual differences are apparent in both the Ph and PhFR images. Figure D.2 shows the results of a more quantitative and robust experiment which demonstrates that images such as these are not so strongly correlated that they are influencing my main results.

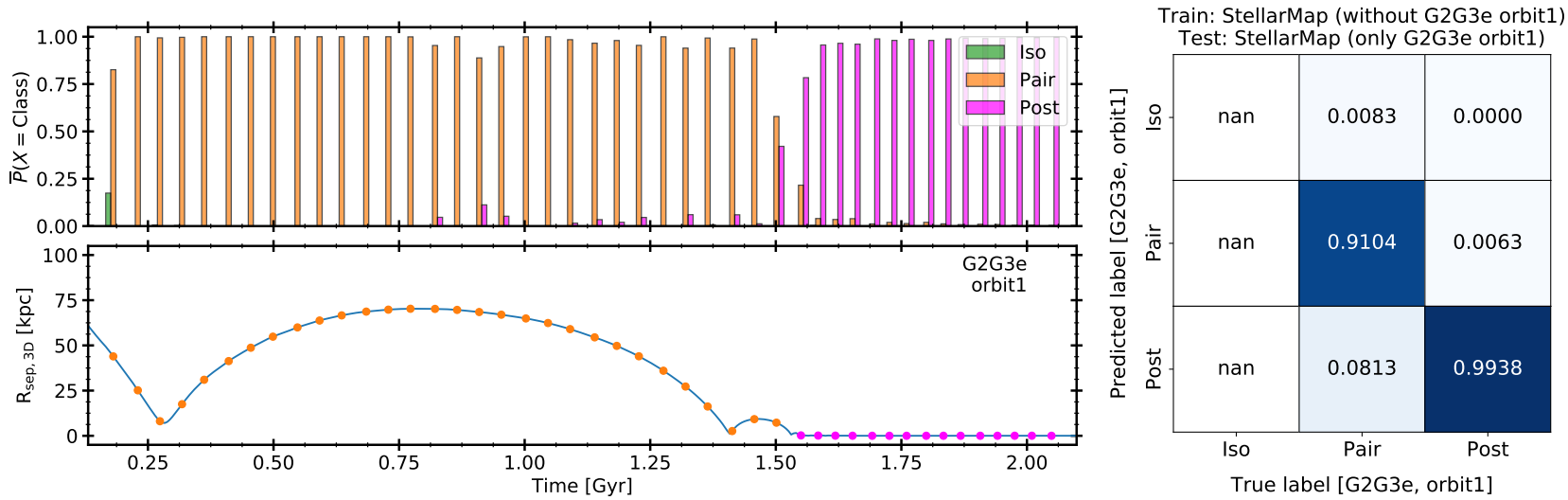

Figure D.2 Classification scores along fiducial G2G3e orbit 1 interaction sequence (left panels) and confusion matrix (right panel) for the experiment described in Appendix D. The experiment is designed to show whether the sampling cadence for each merger yields images that are too strongly correlated – resulting in test images that are too similar to the training/validation images. Applying a network to images from an *entire merger* that the network never saw during training would reveal whether such correlations are affecting my results. The images from the fiducial G2G3e orbit 1 merger were removed from the SM dataset and reserved for testing. A SM network was then trained using the remaining data SM and applied to the G2G3e orbit 1 test images. The right panel shows the confusion matrix for this test. These results are consistent with the results shown in the upper left panel of Figure 3.6 (train: SM, test: SM) – demonstrating that possible correlations between images from neighbouring snapshots are not affecting the results of my main experiments. The left panels show the classification scores as coloured bars for each snapshot in the G2G3e orbit 1 test sequence, averaged over all camera angles at each snapshot, $\overline{P}(X = \text{Class})$. The results of this test show that: (1) that the snapshot sampling cadence is not affecting the networks' performances even in phases where the rate of morphological evolution is expected to be minimal (e.g. near apocenter) and (2) there is a continuous transition between class scores at the temporal class boundaries (e.g. between Pair and Post-merger at $t \approx 1.54$ Gyr).

Appendix E

Merger stage predictions: main handshake results

This appendix to Chapter 3 provides the results of every test conducted as part of the main handshake experiment from Chapter 3, Section 3.3.1. Figure E.1 shows the combined confusion matrices for every test in the main handshake experiment. Each row corresponds to a type of training data. Each column corresponds to a type of test data. As in Figure 3.4, each matrix shown combines 10 individual tests performed with different random allocations of training, validation and test images.

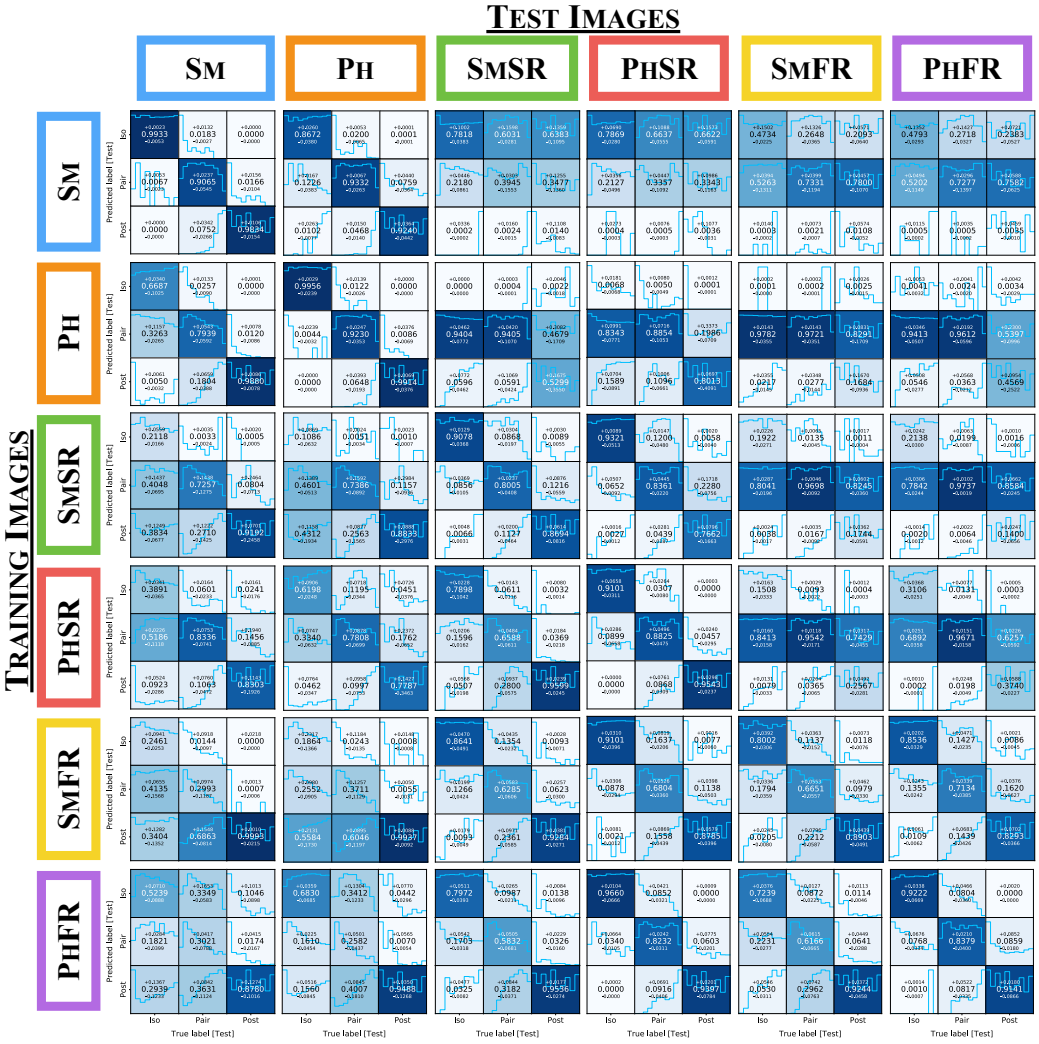

Figure E.1 Confusion matrices corresponding to every test carried out in Section 3.3.1. For reference, Figure 3.4 describes the information displayed by each individual confusion matrix in detail. The median overall performance of each result is shown in Figure 3.9. Refer to the corresponding publication (Bottrell et al., 2019a) or the electronic version of this thesis to examine the figure in high resolution.

Appendix F

Merger stage predictions: colour images

In this appendix, I show *gri* colour-composites of idealized and survey-realistic photometric images from Chapter 3. The *gri* composites are generated using the an Astropy visualization tool for combining RGB (Red, Green, Blue) images from CCDs. This tool derives originally from algorithms developed in Lupton et al. (2004).

Figure F.1 shows idealized (Ph) colour-composite images for isolated (upper row), pair (middle row), and post-merger (lower row) snapshots. Each row shows a single snapshot from the corresponding stage realized with four unique camera angles. The reddening and obscuration of light from the discs and nuclear regions of these galaxies is a consequence of the dust model employed in the radiative transfer. The effects of dust are particularly visible in the edge-on orientation of the isolated disc in the upper right panel.

Figure F.2 shows the survey-realistic counter-parts of the images in Figure F.1. Note the strong variation in the crowdedness of the images in terms of real foreground and background sources – including foreground stars which can generate diffraction spikes as seen in the lower right panel. The brightness of the sky and reduced angular resolution also limit the visibility of faint, low surface brightness features that are easily picked out in the idealized images from Figure F.1. These factors are some of the central motivations for the analyses conducted in Chapter 3. Figure F.3 directly contrasts four idealized and survey-realistic images from the G2G3e orbit 1 merger.

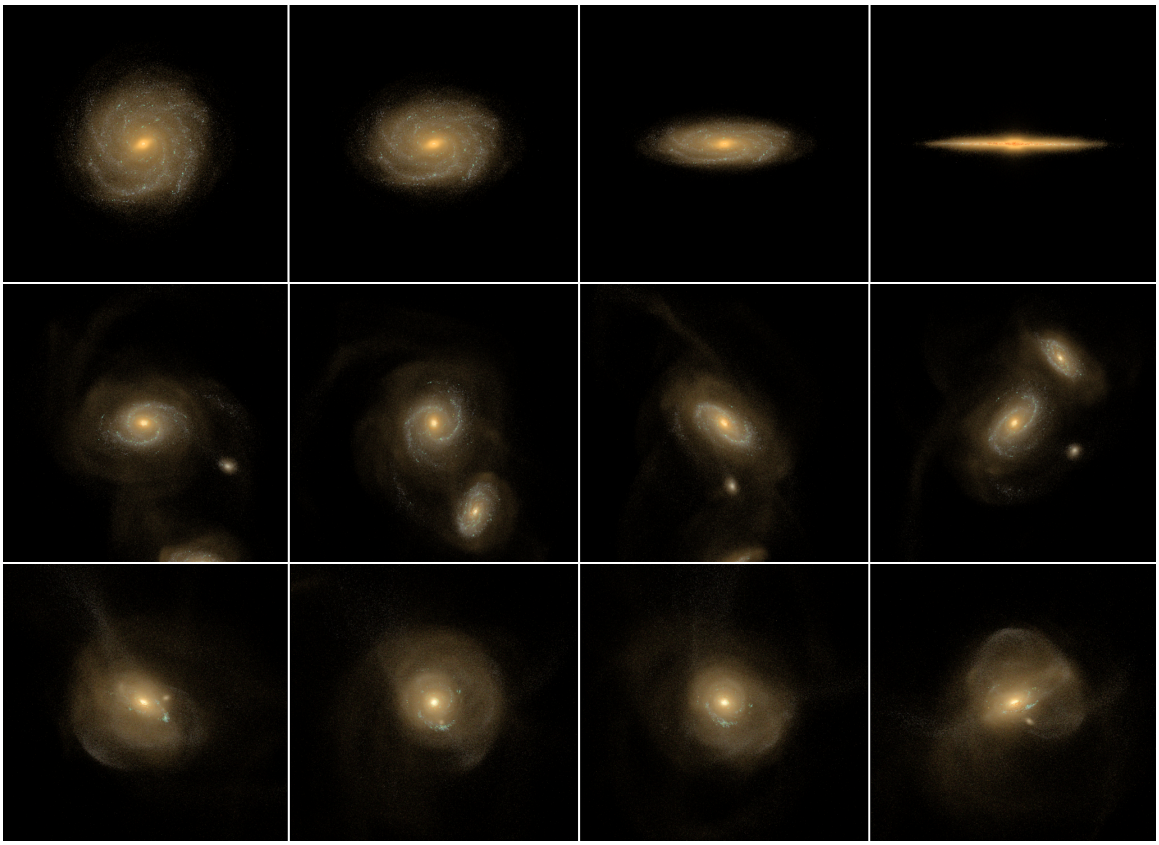

Figure F.1 *gri* colour composites of idealized synthetic images (PH) from the isolated G3 simulation (upper row) and G2G3e orbit 1 merger (lower rows). Each row shows four camera angle realizations of a single snapshot in the isolated phase (upper row), pair phase (middle row) and post-merger phase (lower row). Each field of view is 50 kpc or 55 arcsec at the redshift at which the galaxies are inserted, $z = 0.046$.

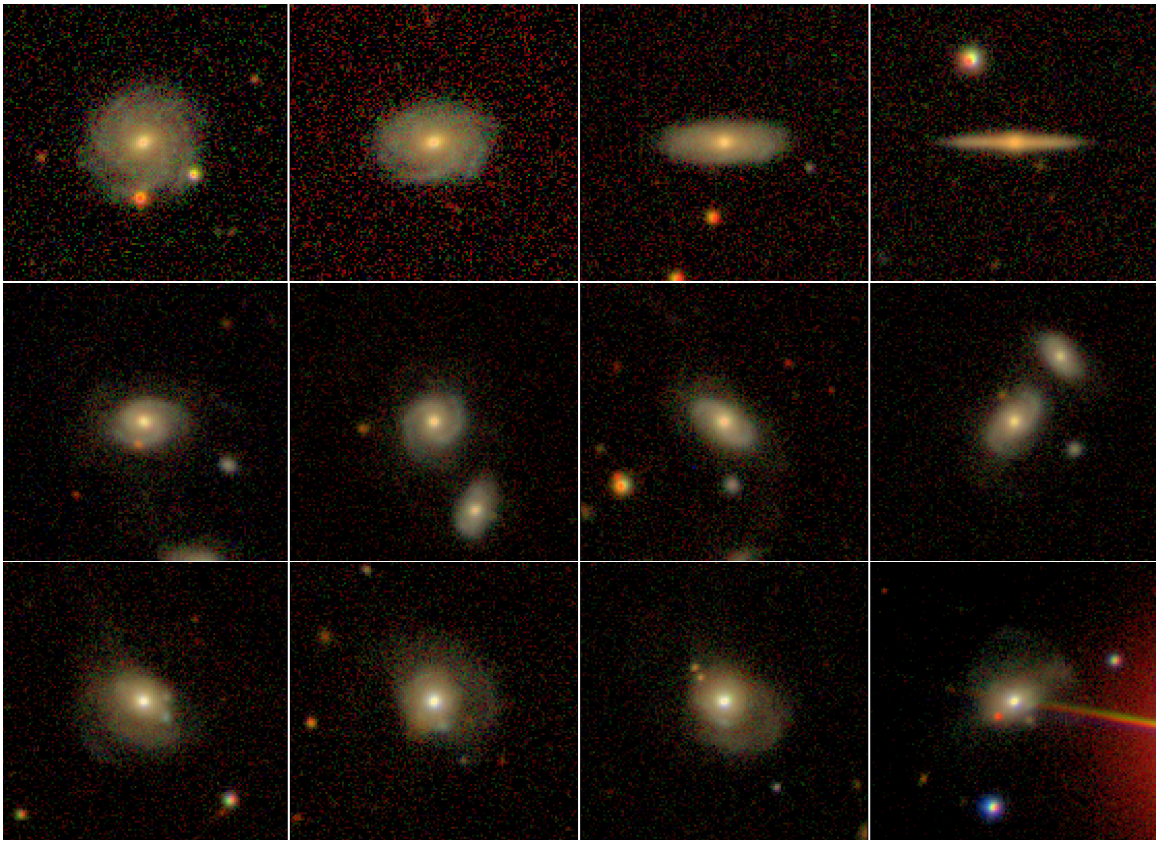

Figure F.2 Same snapshots and camera angles as Figure F.1 but with SDSS survey-realism produced with RealSim.

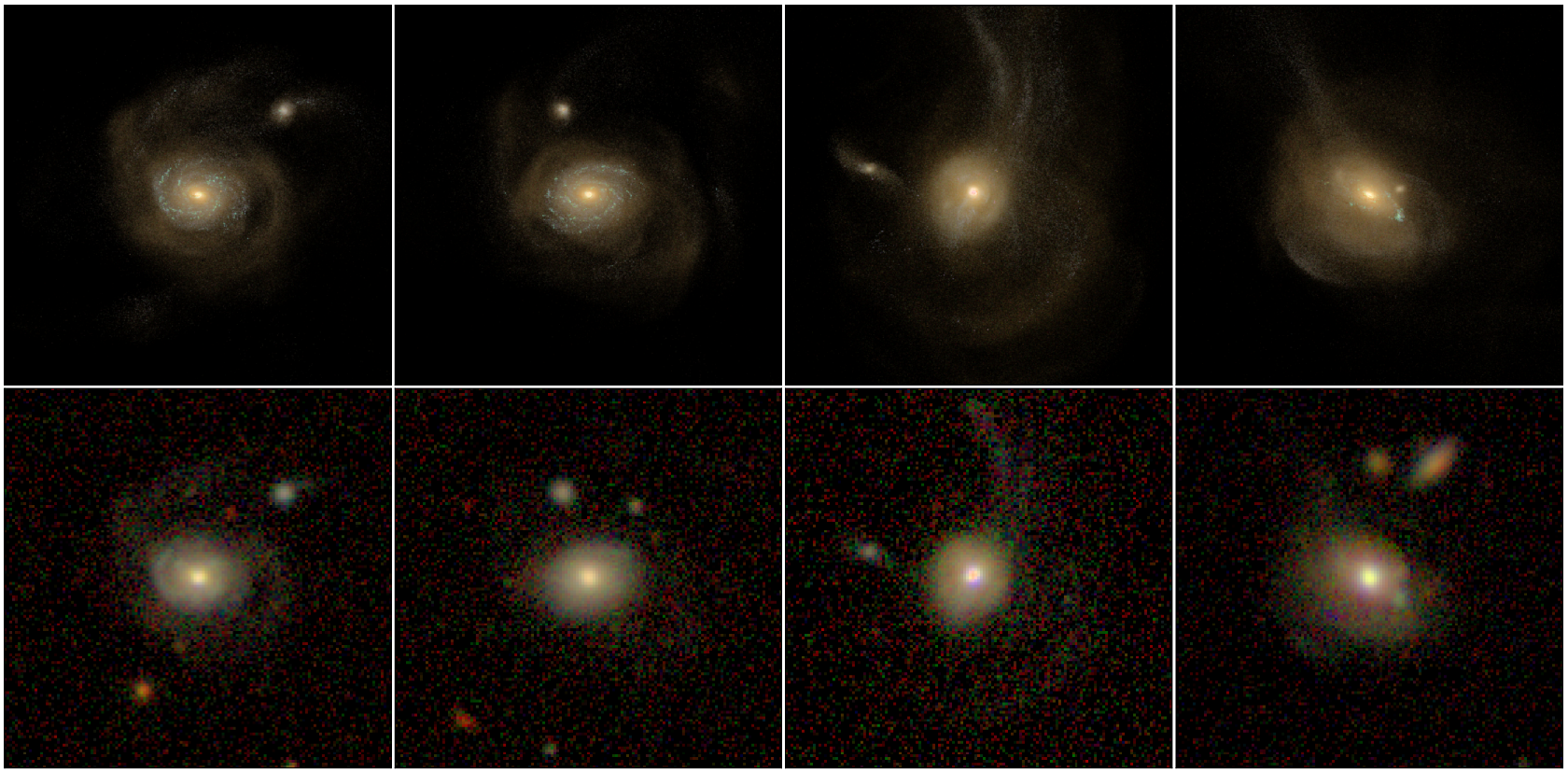

Figure F.3 *gri* colour composites of idealized and survey-realistic synthetic images from the G2G3e orbit 1 merger. The first two columns show the galaxy in the pair phase but in orientations where the companion is not within the 50 kpc field of view. The third and fourth column show the post-merger at the moment of coalescence shortly afterwards ($t - t_c \approx 0$ and $t - t_c \approx 450$ Myr, respectively, with t_c as defined in Equation 3.1).

Appendix G

Synthetic kinematics: TNG100 MaNGA observations

This appendix to Chapter 4 includes supplementary examples from the TNG100-1 synthetic MaNGA survey outlined in Section 4.4.2. 13 galaxies are highlighted including several post-merger galaxies with $t_{\rm pm} < 162$ Myr. The Primary (PRI), Secondary (SEC), and Colour-Enhanced (CEN) sample MaNGA observations for each galaxy are shown (where available) for two single camera angles each. One can see that the post-merger galaxies are a visually distinct population from the galaxies with large $t_{\rm pm}$ and dimensionless, relative companion separations $r_{\rm sep}$ based on their MaNGA kinematics. These examples highlight the viability of improving the accuracy of merger classification using spatially resolved galaxy kinematics.

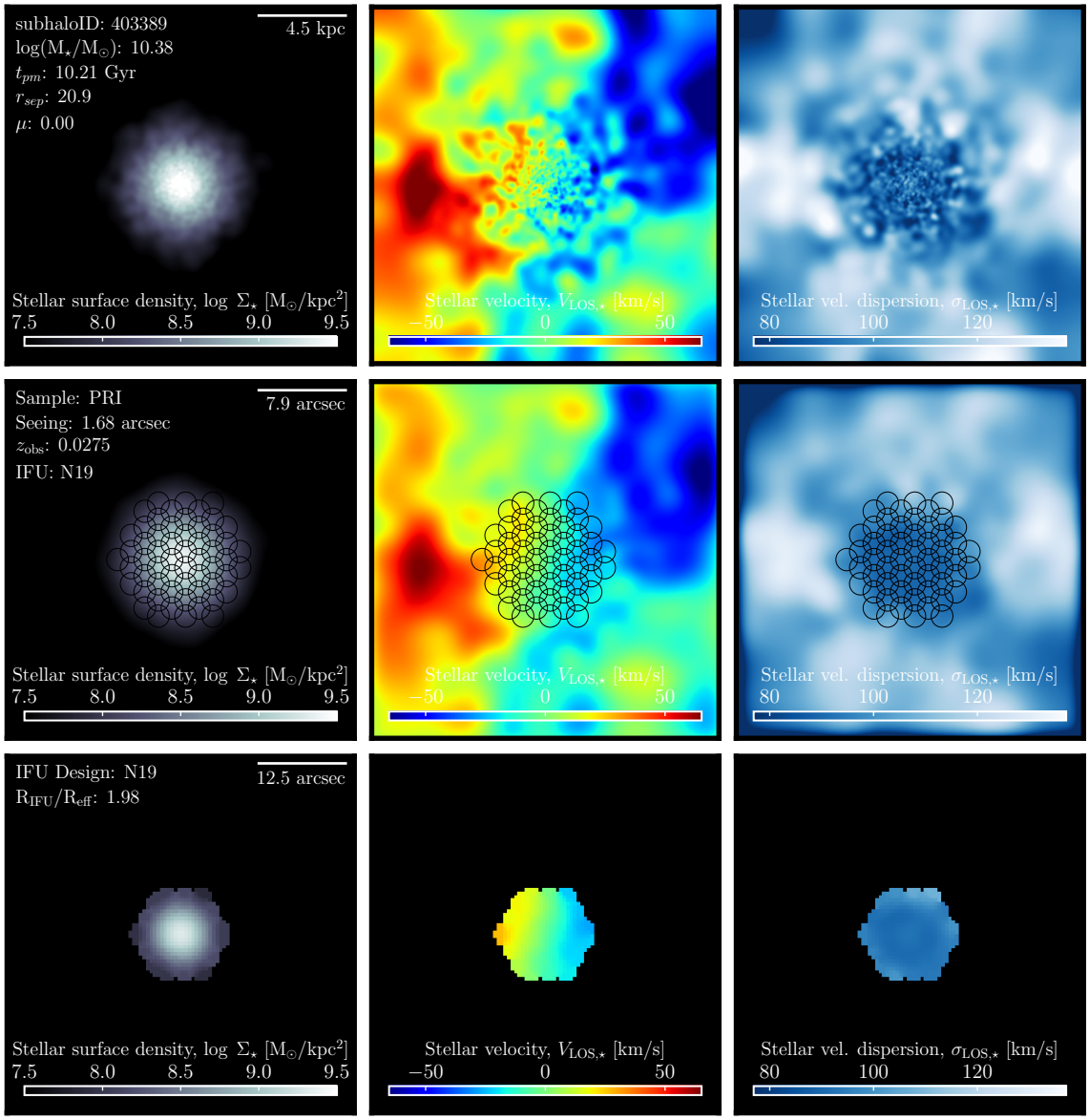

Figure G.1 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403389, camera: 1 observed for the PRI sample.

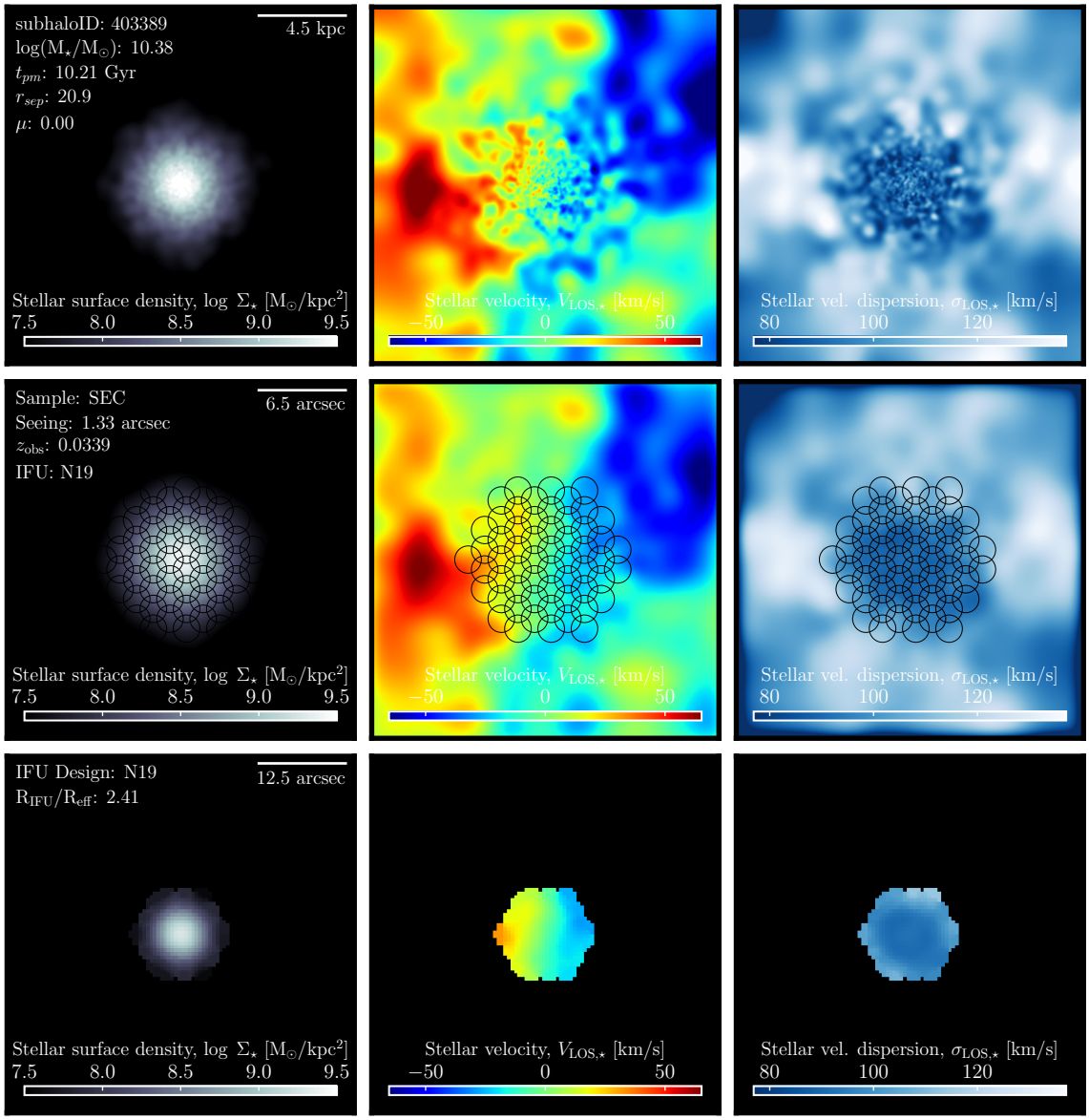

Figure G.2 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403389, camera: 1 observed for the SEC sample.

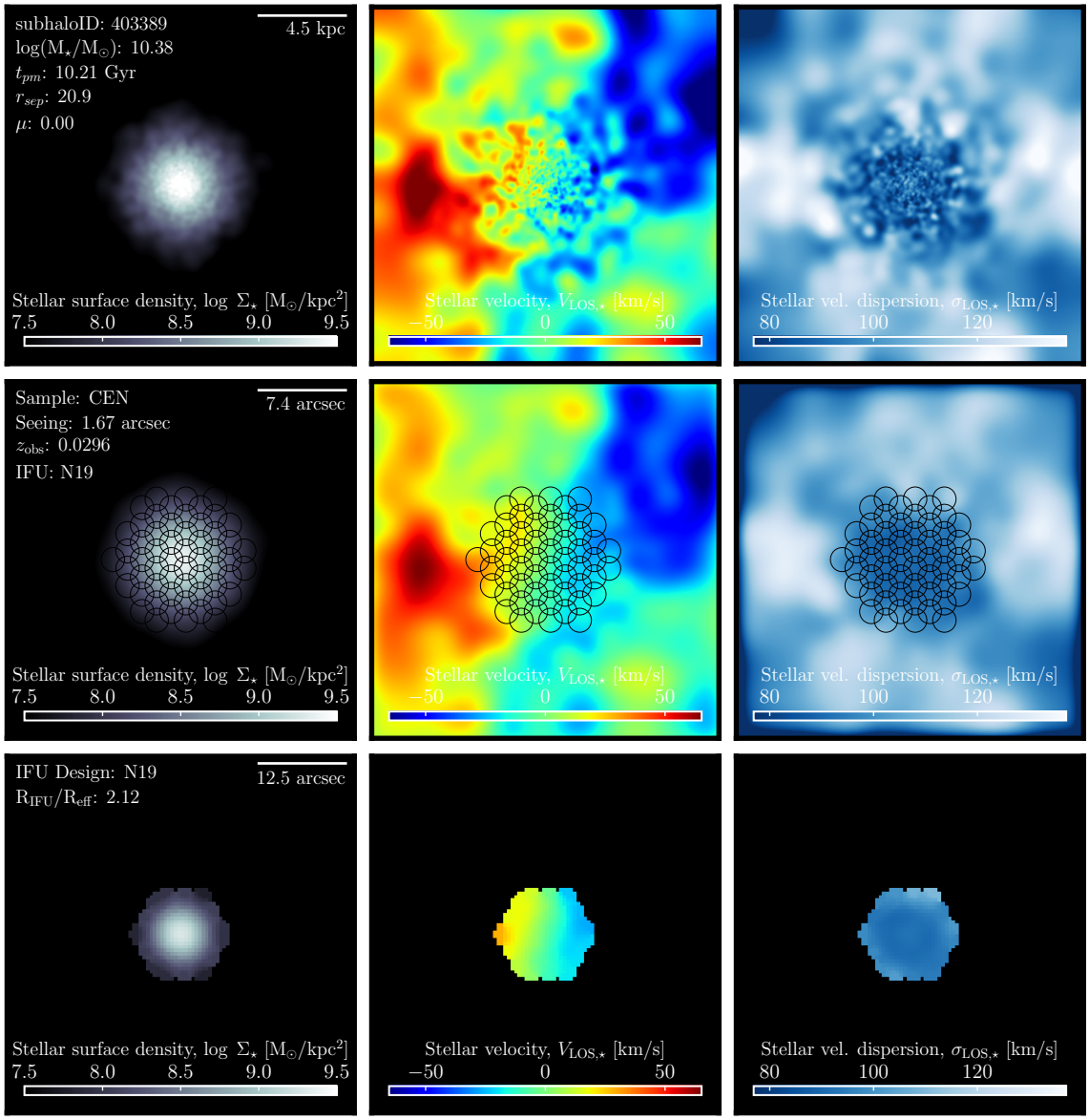

Figure G.3 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403389, camera: 1 observed for the CEN sample.

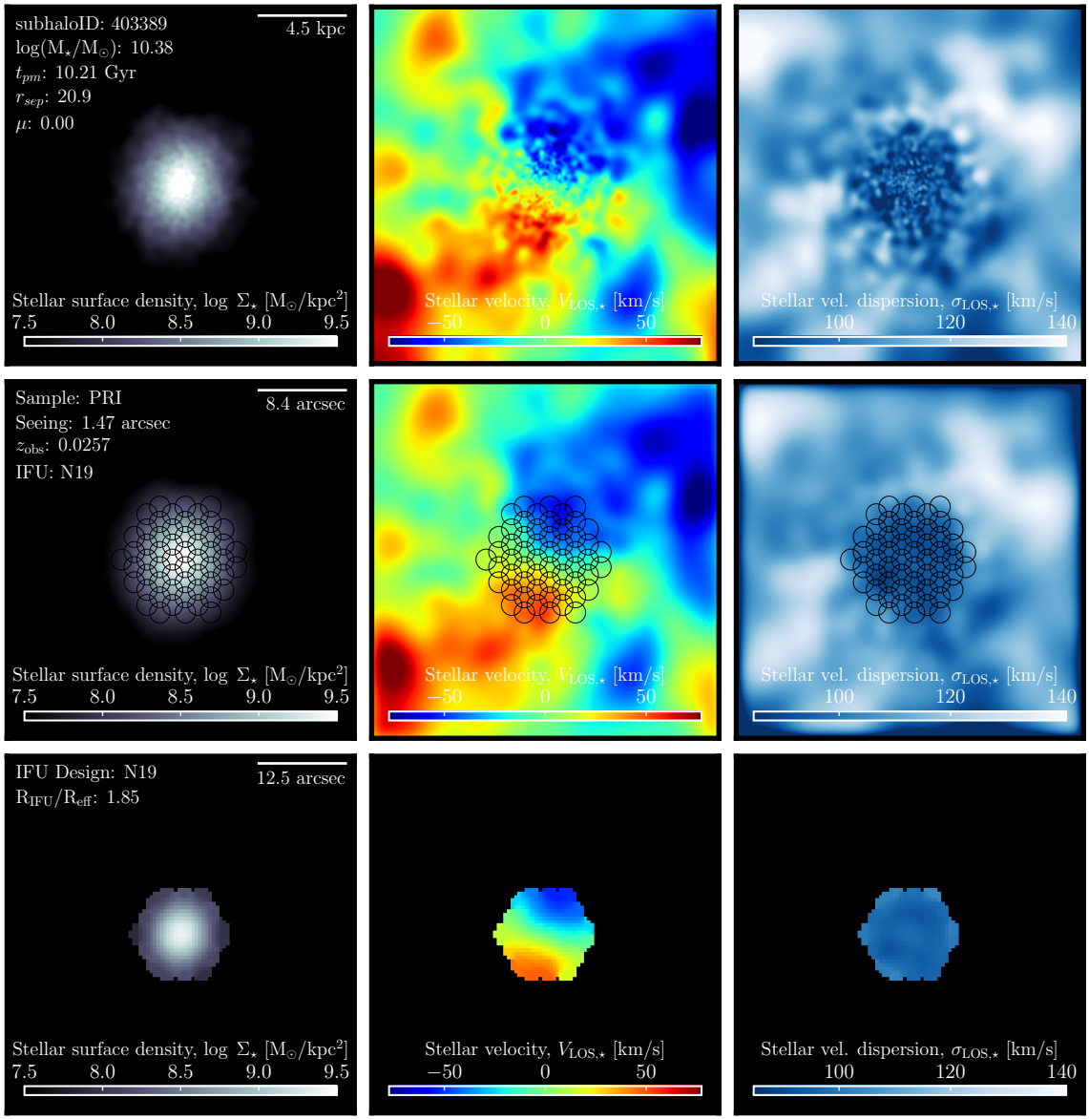

Figure G.4 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403389, camera: 3 observed for the PRI sample.

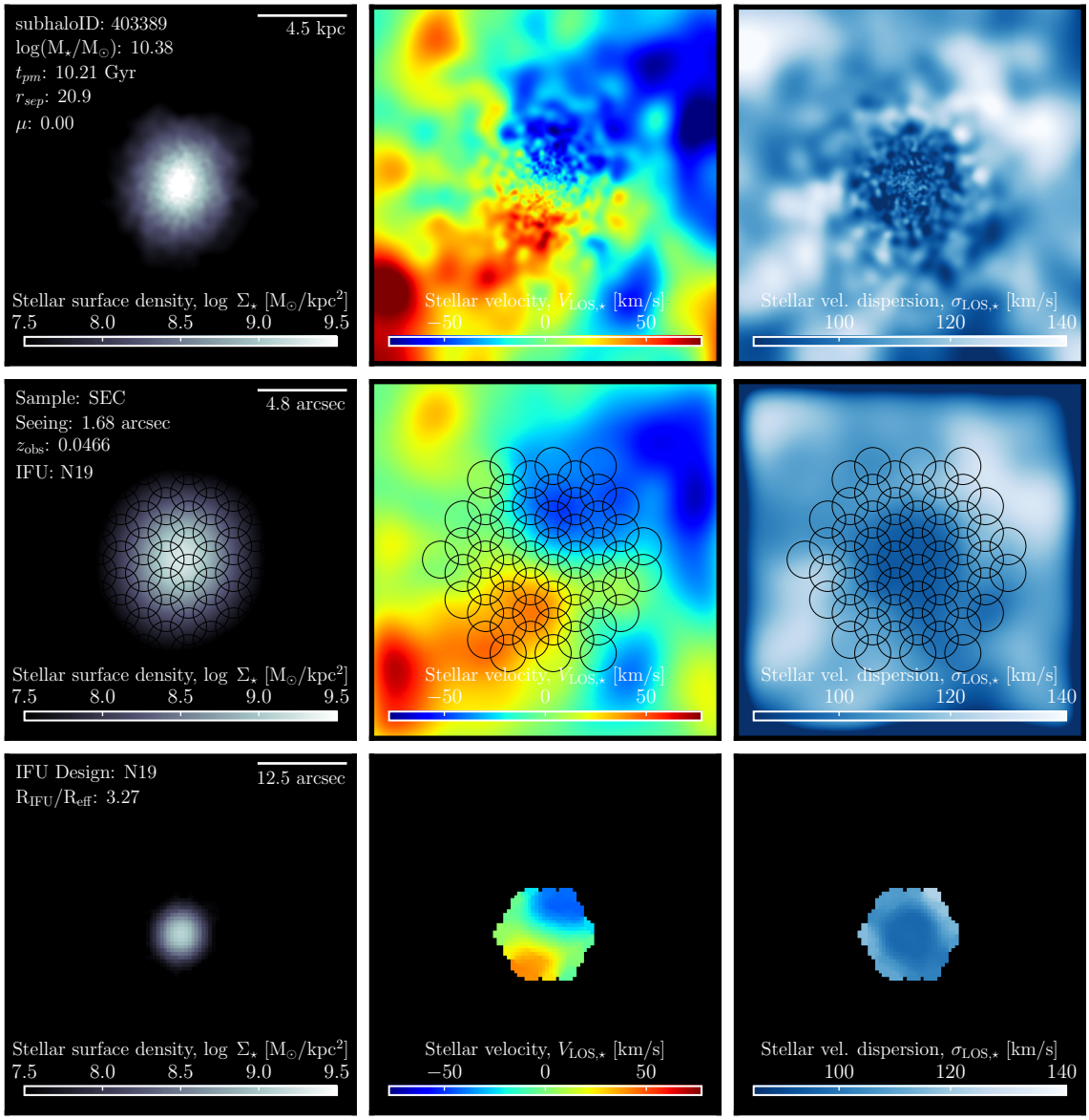

Figure G.5 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403389, camera: 3 observed for the SEC sample.

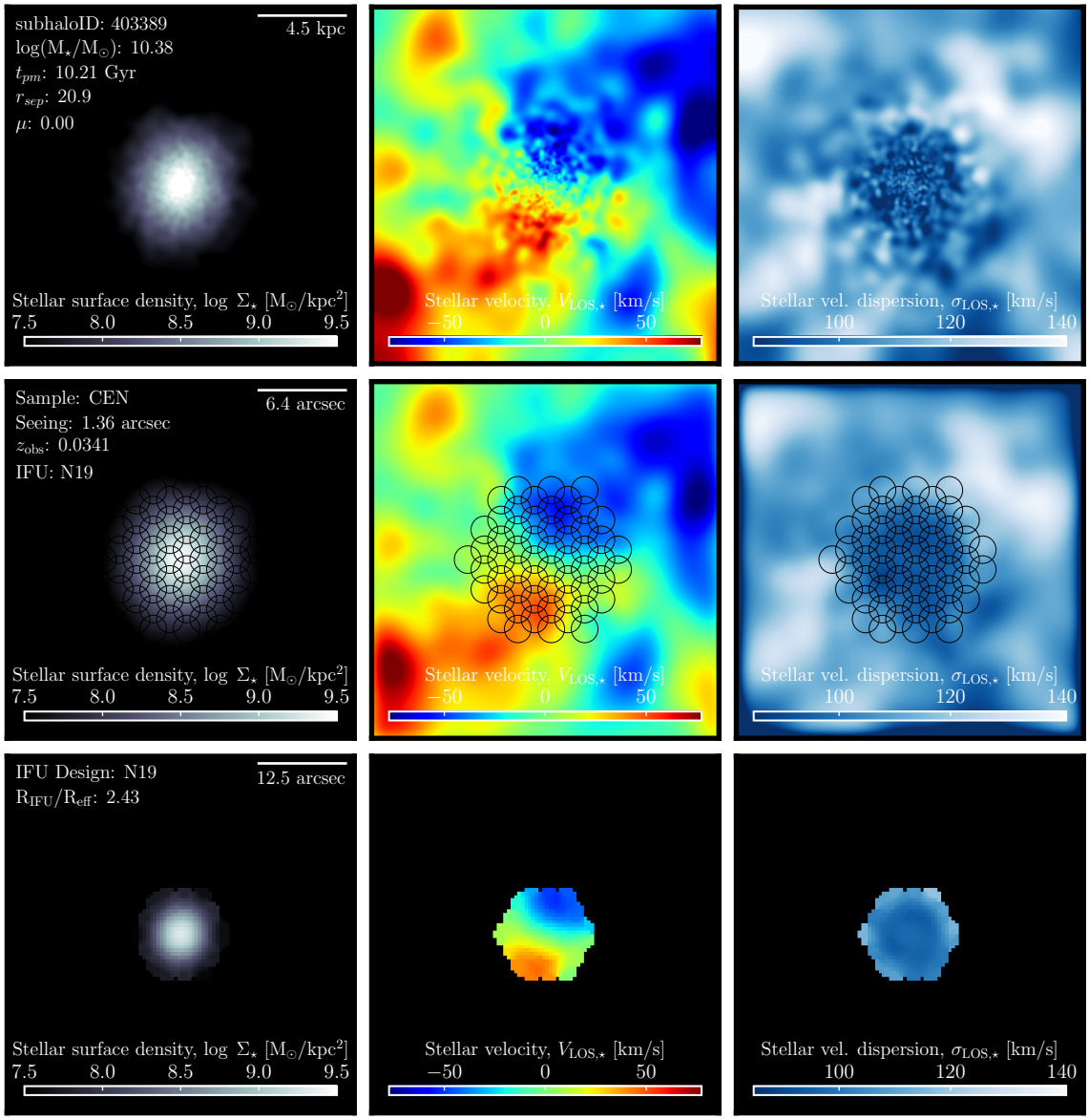

Figure G.6 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403389, camera: 3 observed for the CEN sample.

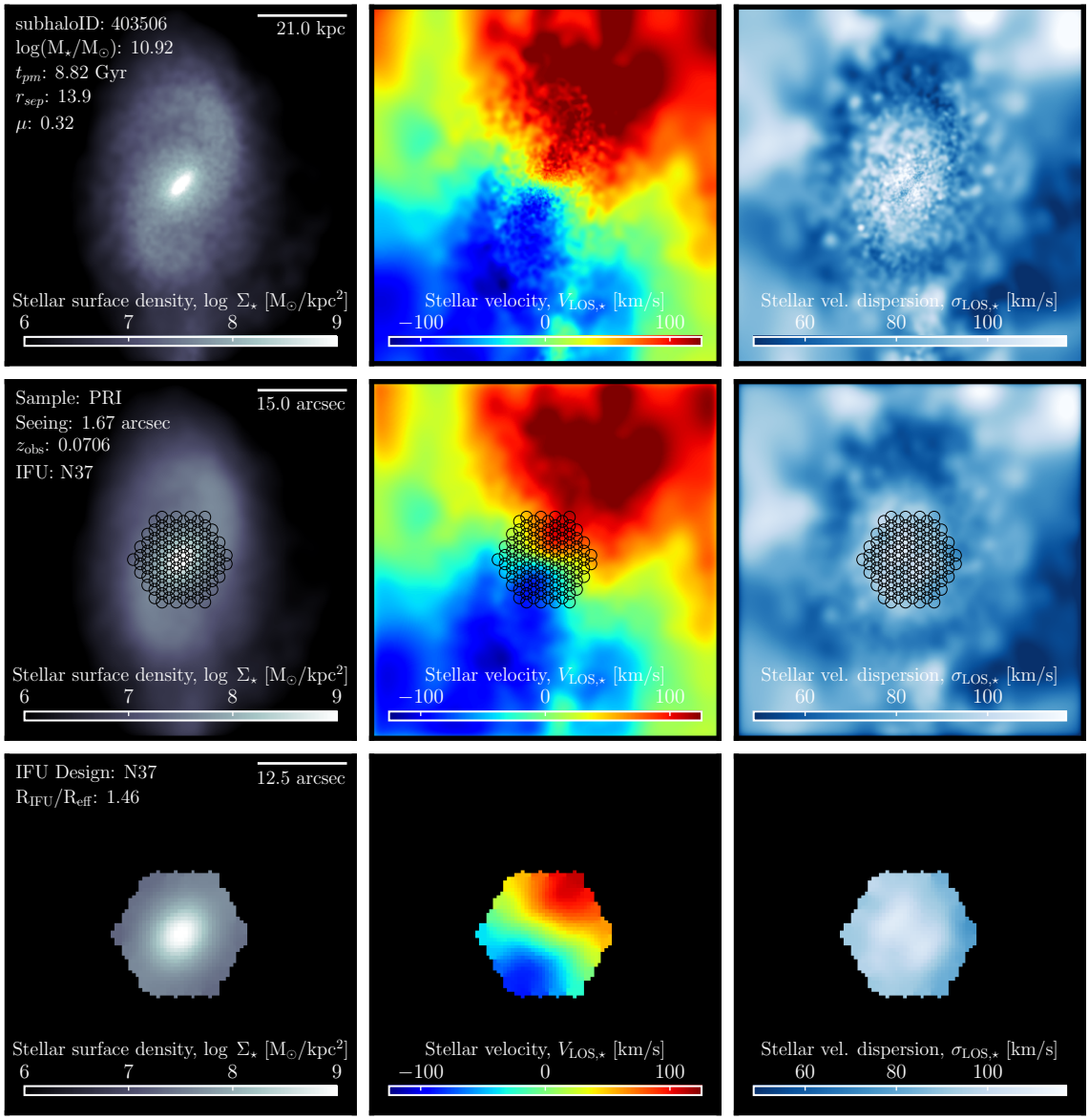

Figure G.7 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403506, camera: 1 observed for the PRI sample.

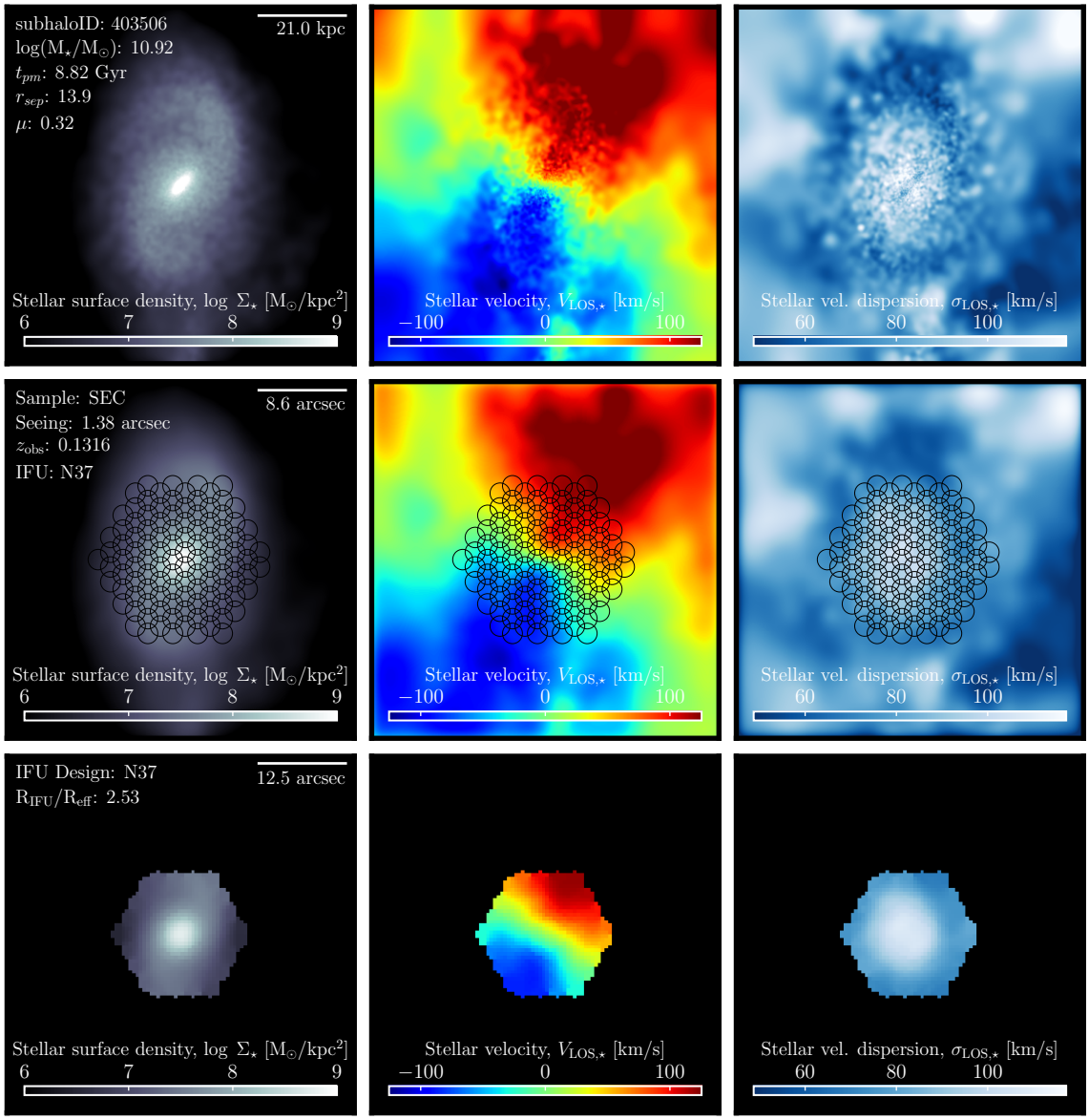

Figure G.8 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403506, camera: 1 observed for the SEC sample.

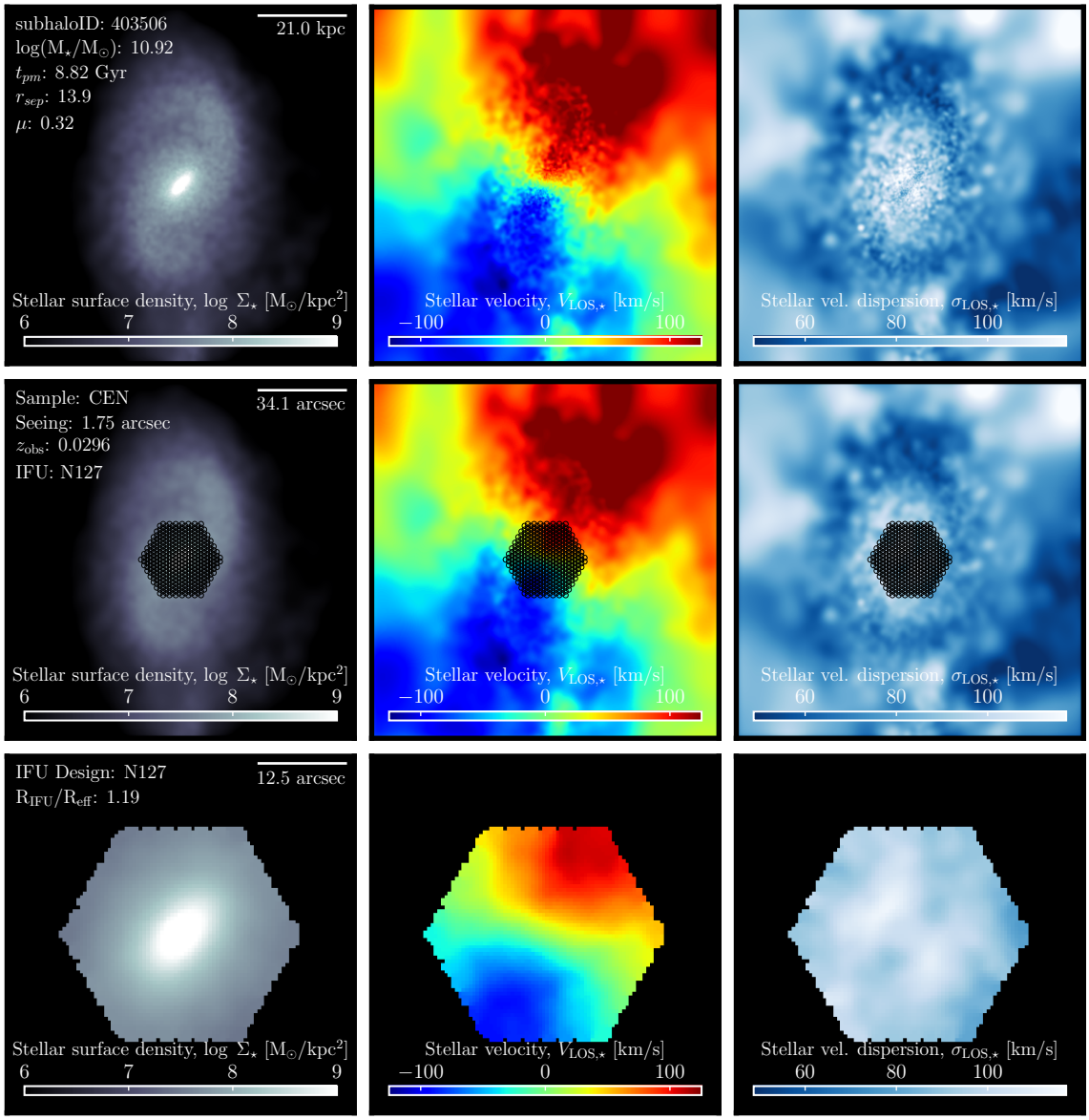

Figure G.9 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403506, camera: 1 observed for the CEN sample.

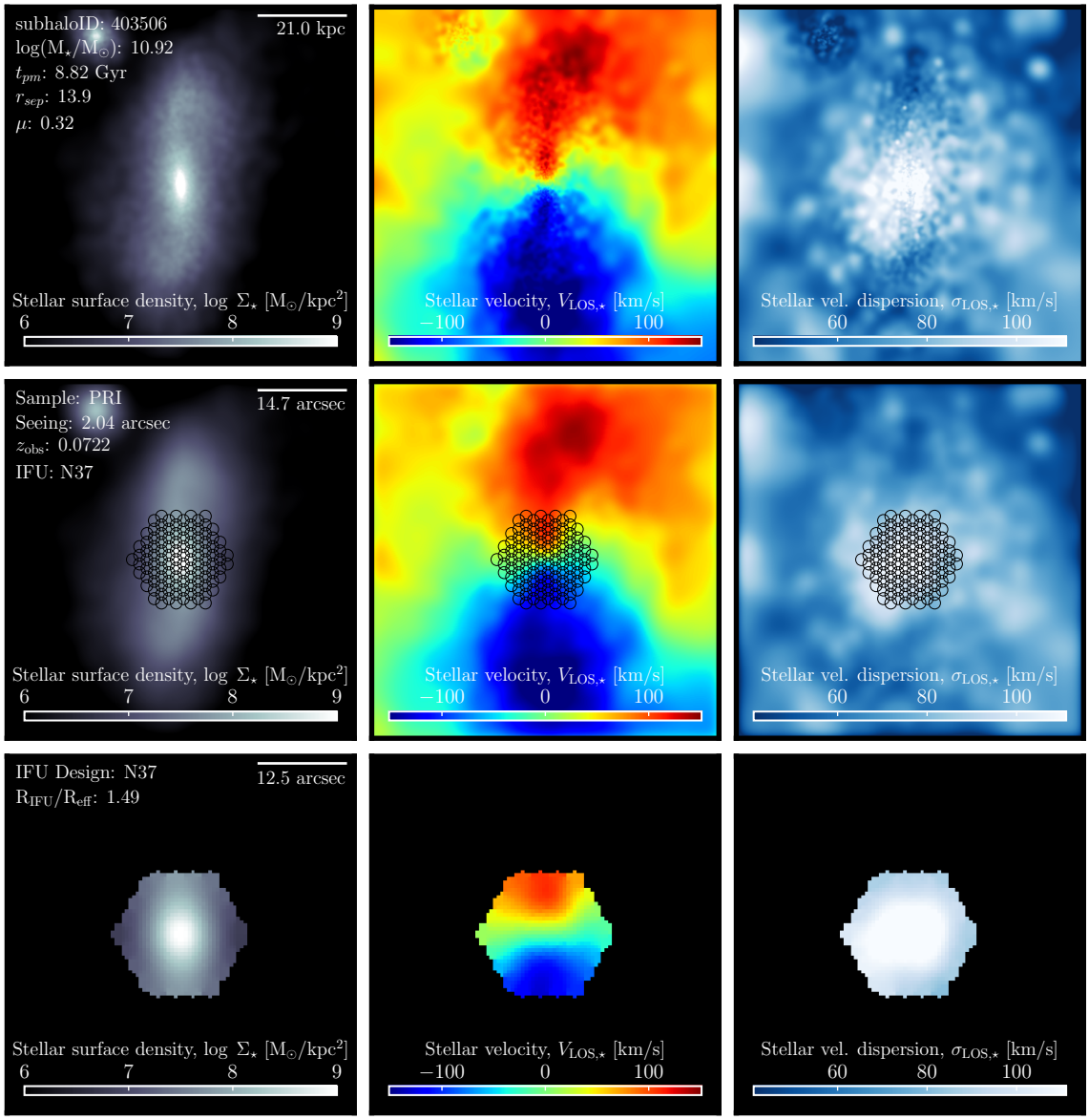

Figure G.10 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403506, camera: 3 observed for the PRI sample.

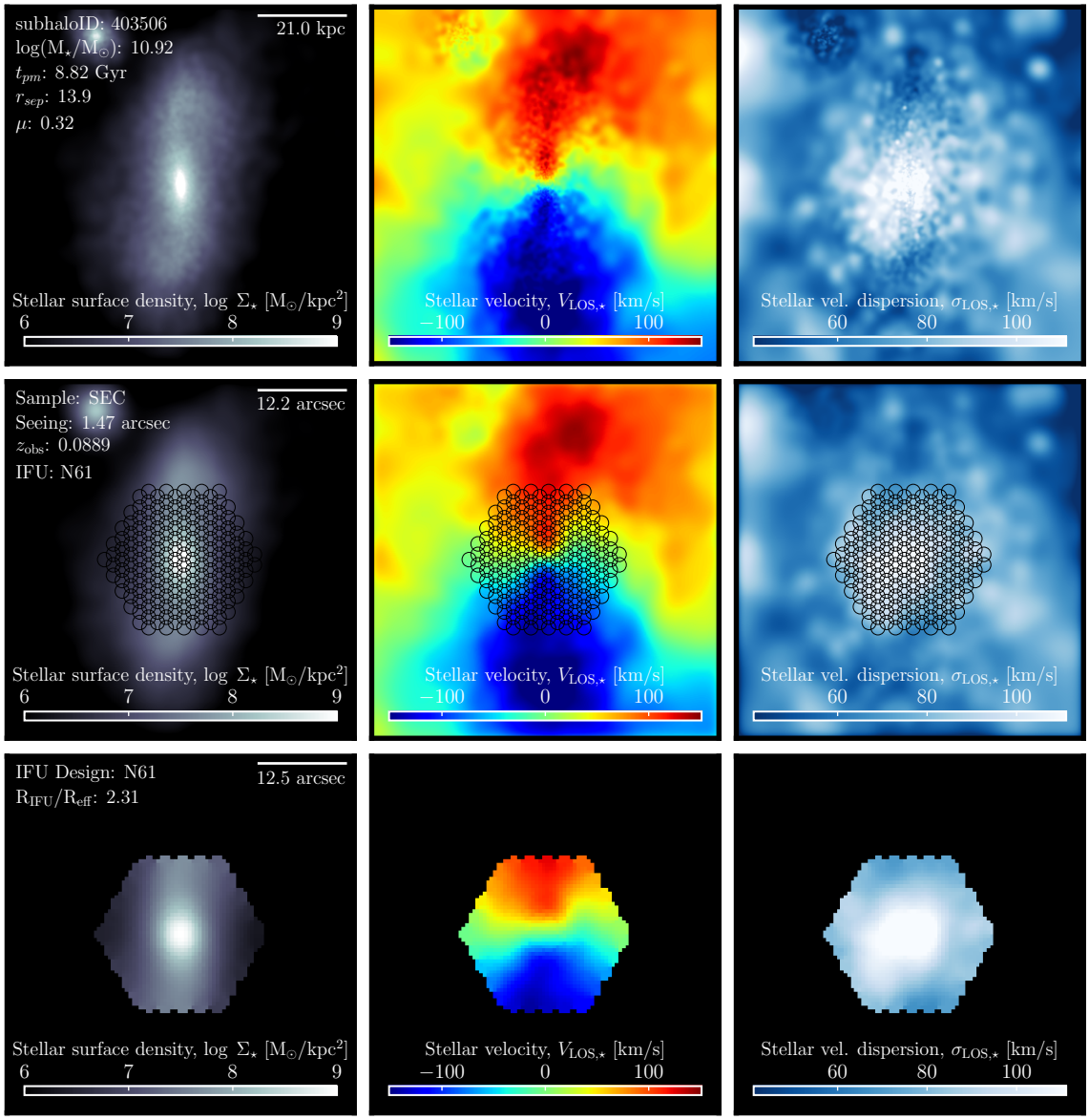

Figure G.11 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403506, camera: 3 observed for the SEC sample.

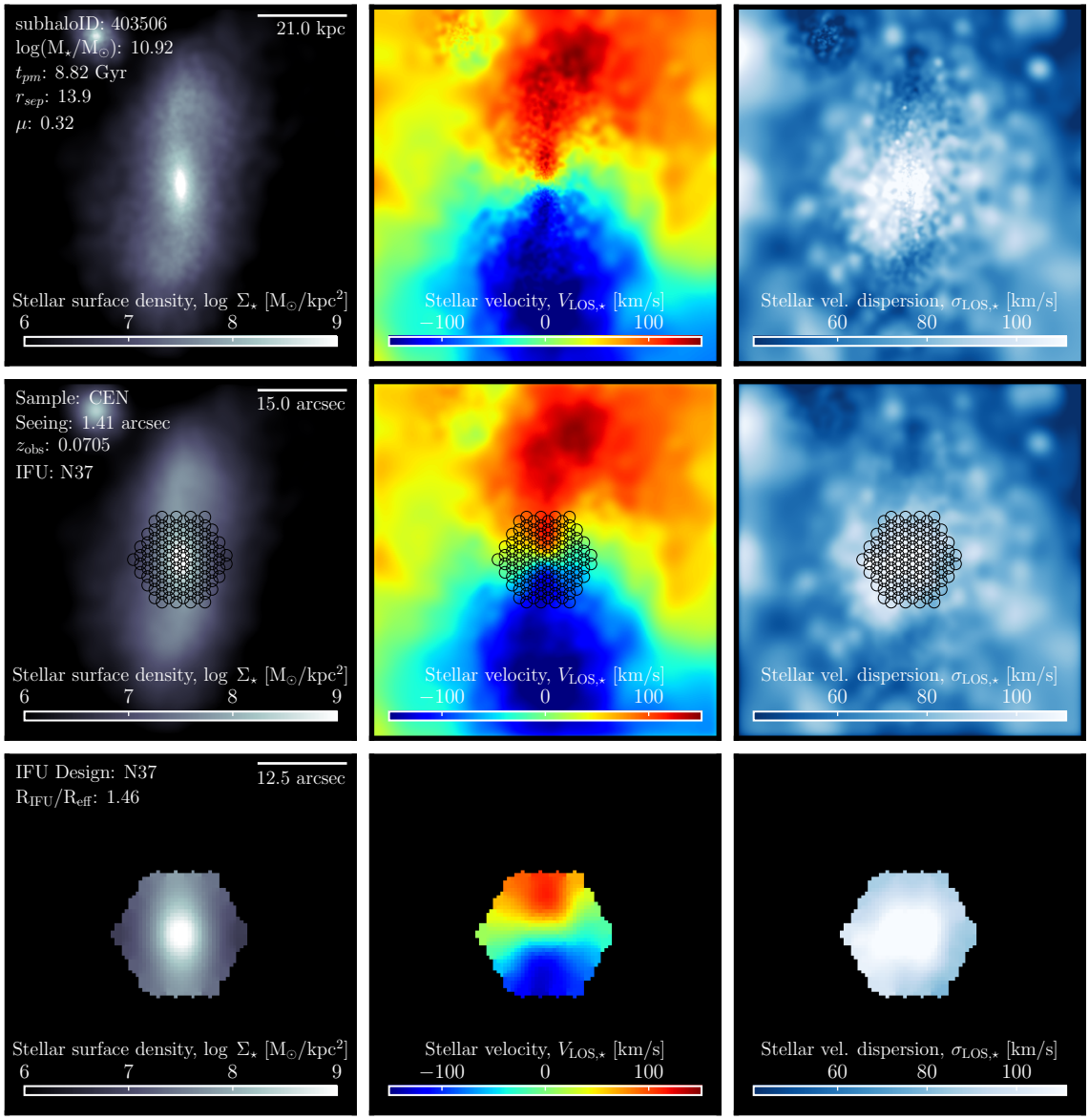

Figure G.12 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403506, camera: 3 observed for the CEN sample.

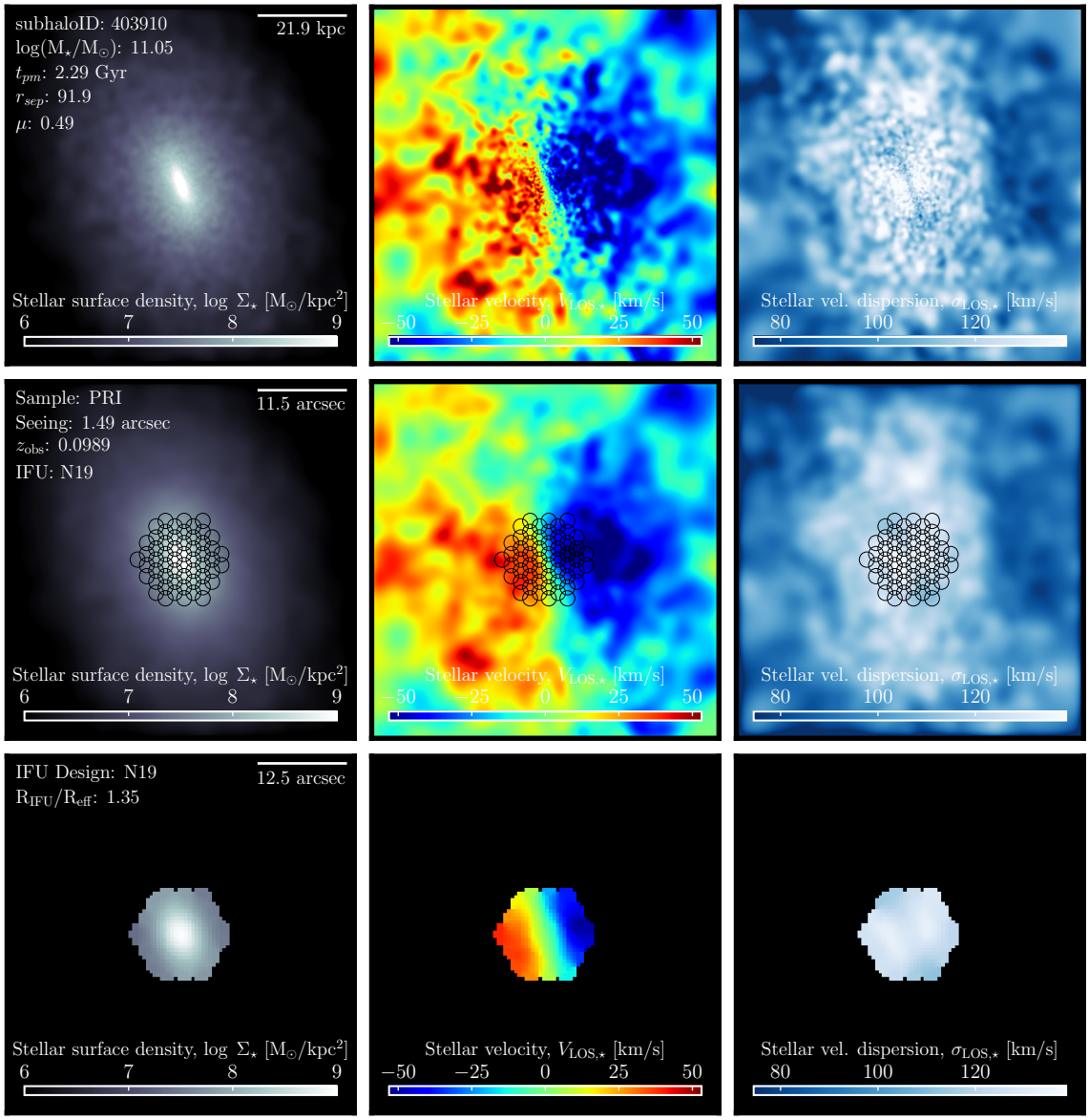

Figure G.13 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403910, camera: 1 observed for the PRI sample.

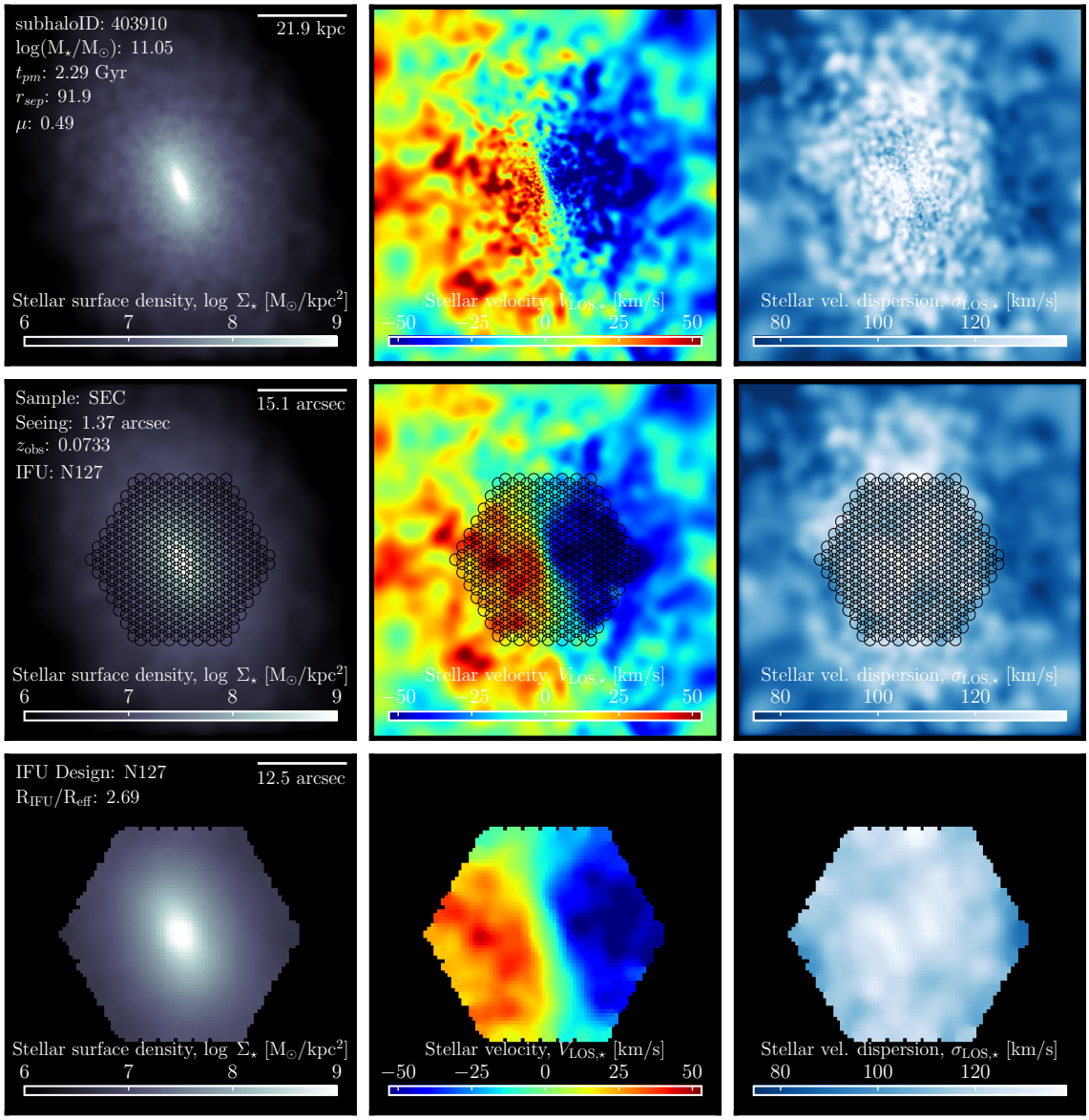

Figure G.14 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403910, camera: 1 observed for the SEC sample.

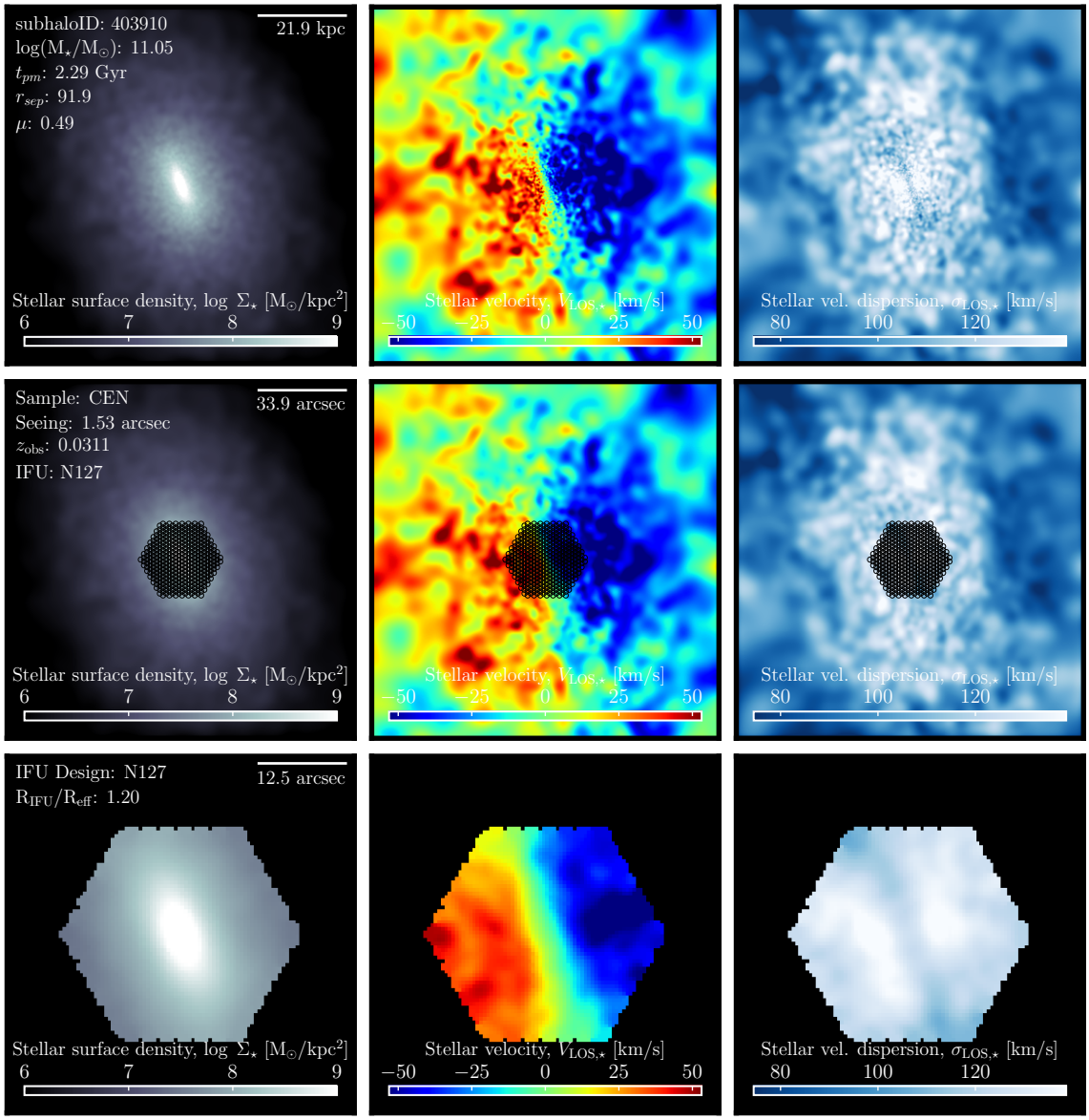

Figure G.15 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403910, camera: 1 observed for the CEN sample.

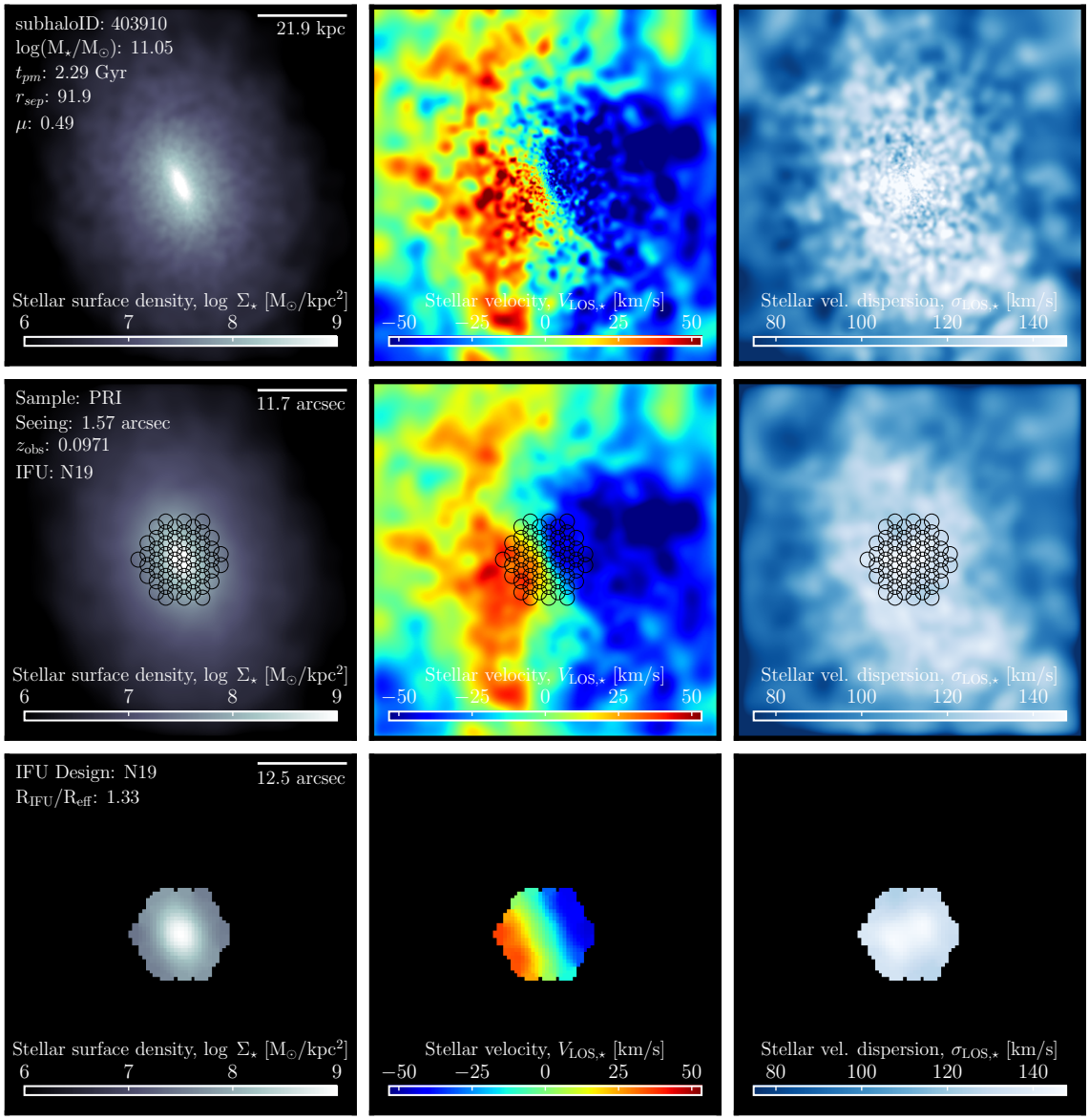

Figure G.16 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403910, camera: 3 observed for the PRI sample.

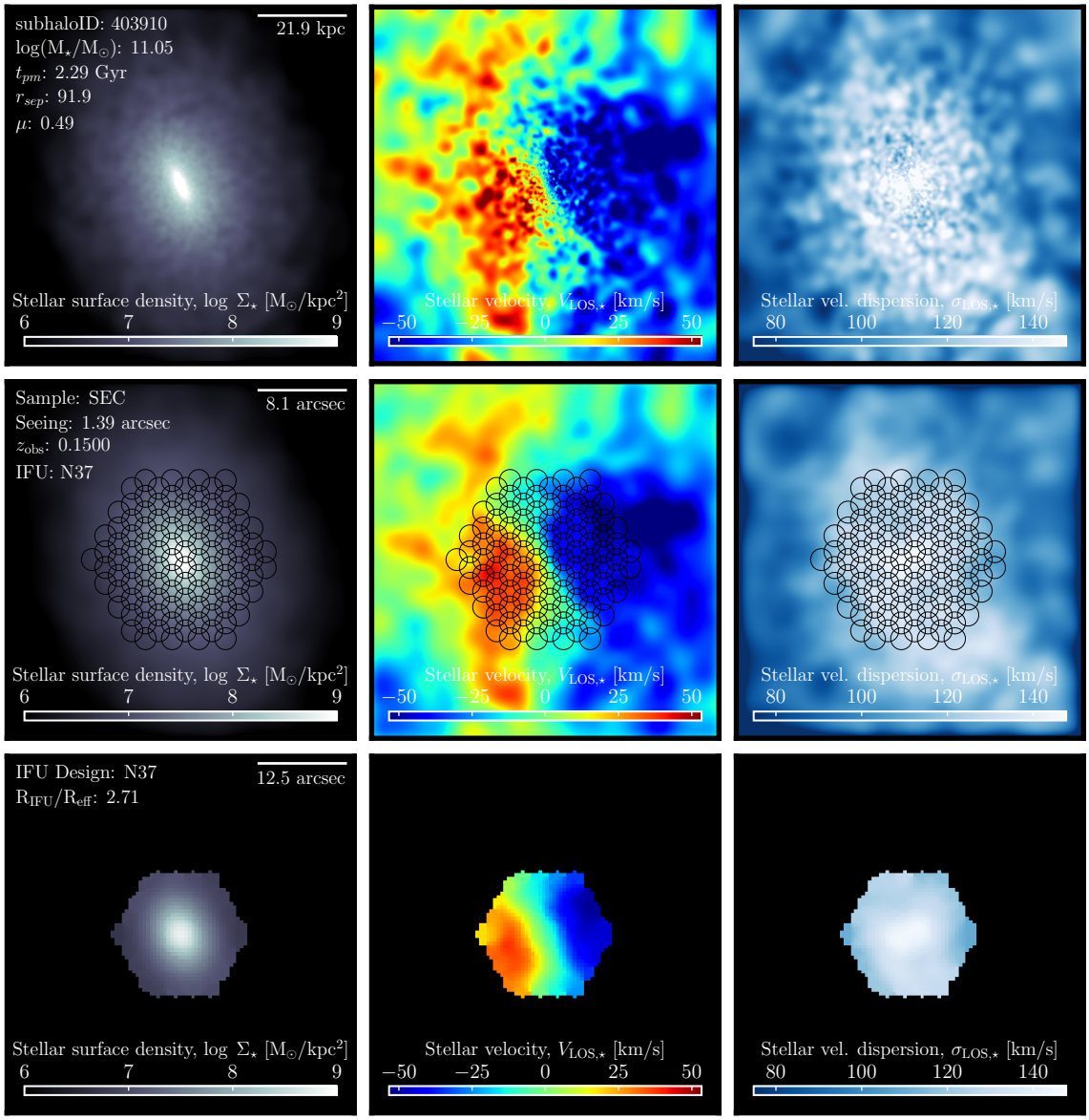

Figure G.17 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403910, camera: 3 observed for the SEC sample.

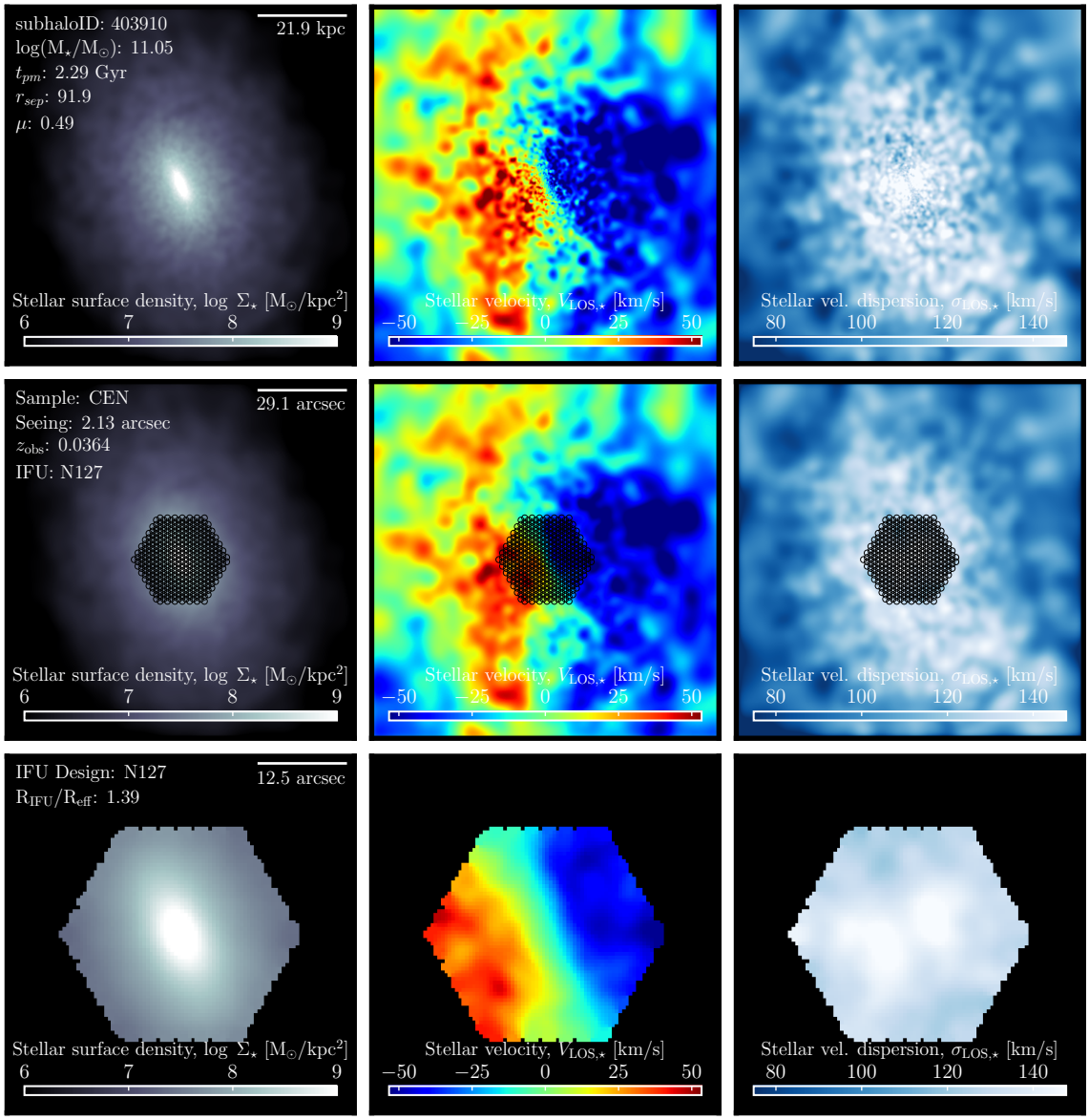

Figure G.18 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 403910, camera: 3 observed for the CEN sample.

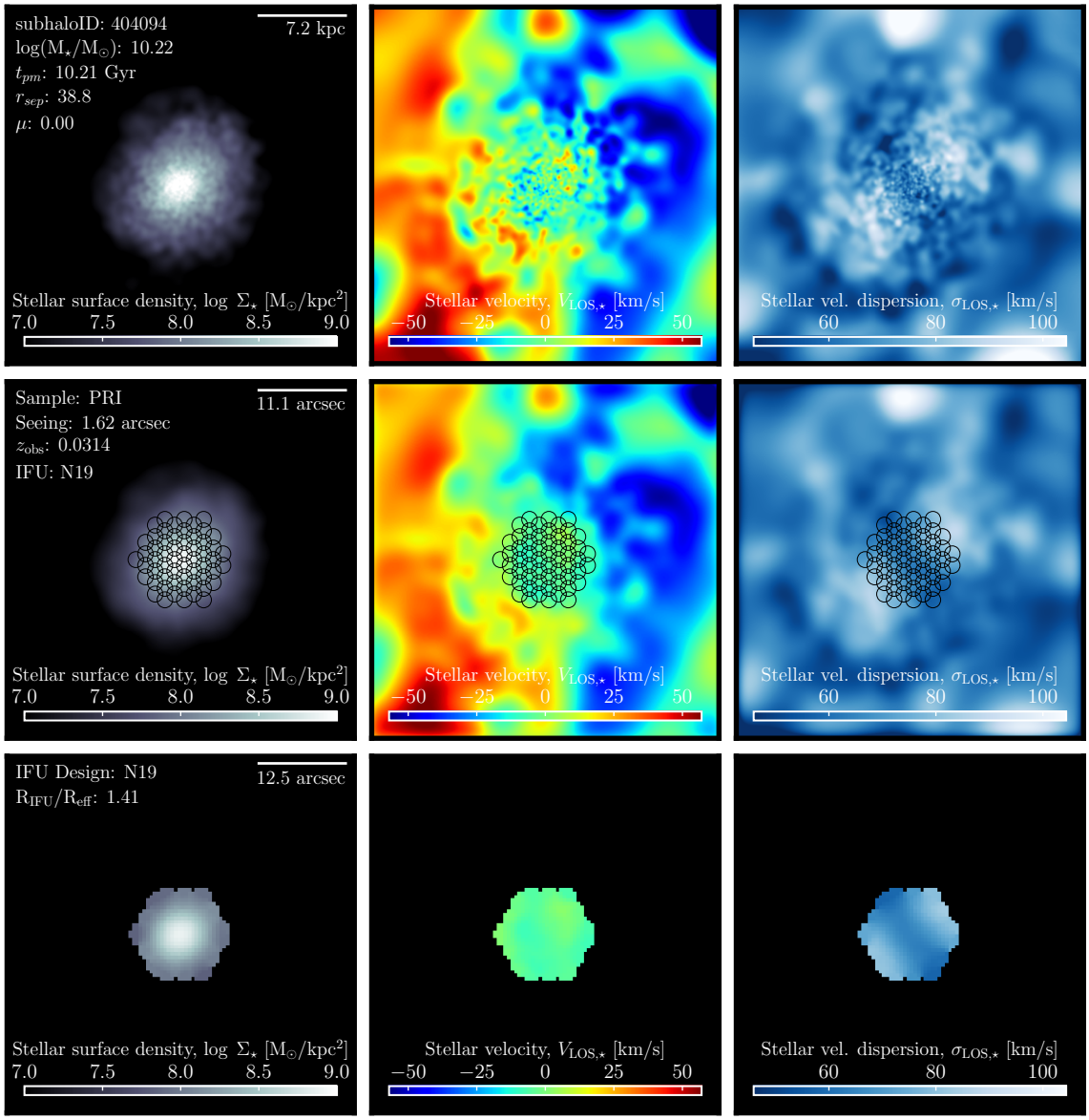

Figure G.19 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404094, camera: 1 observed for the PRI sample.

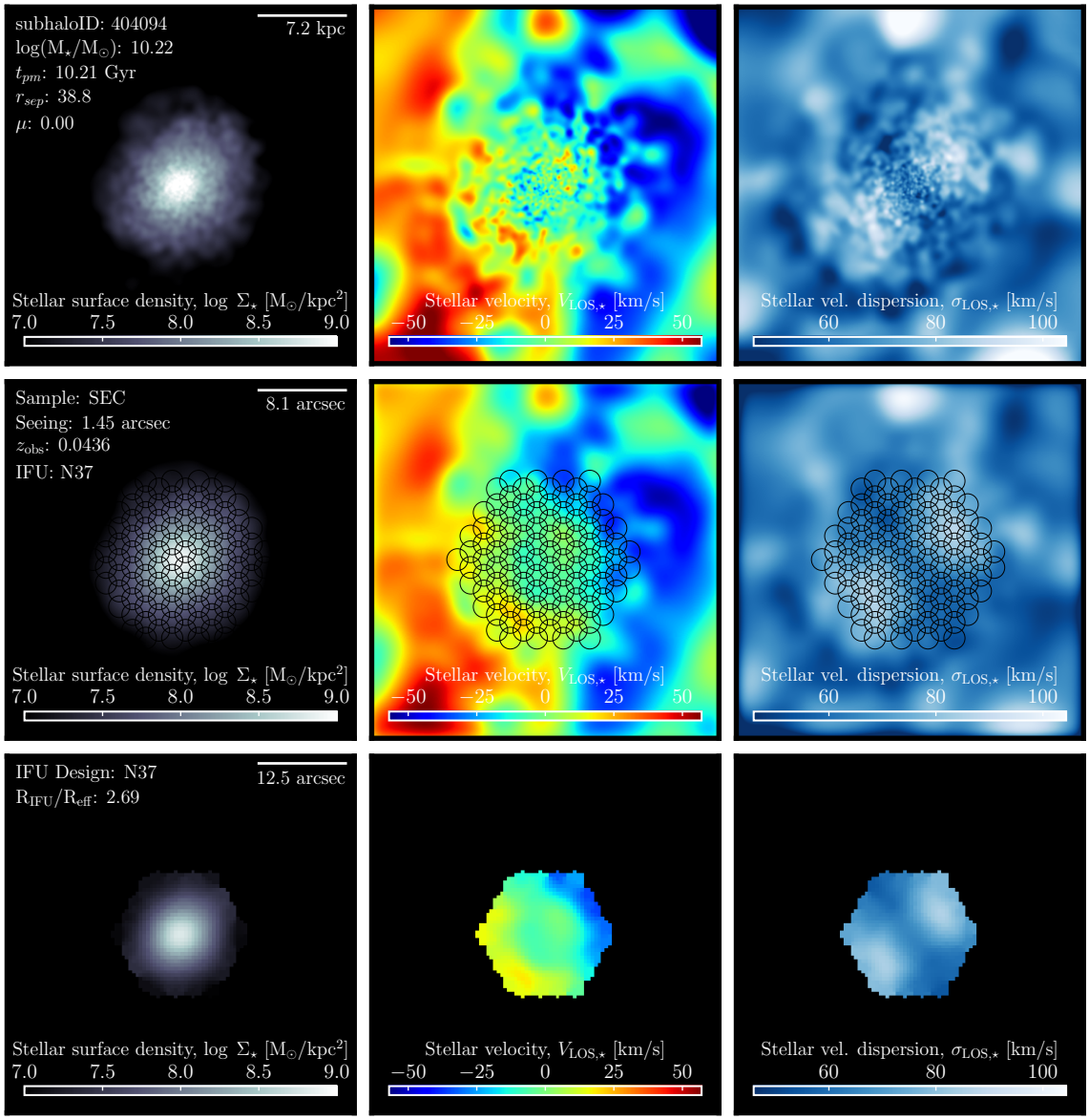

Figure G.20 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404094, camera: 1 observed for the SEC sample.

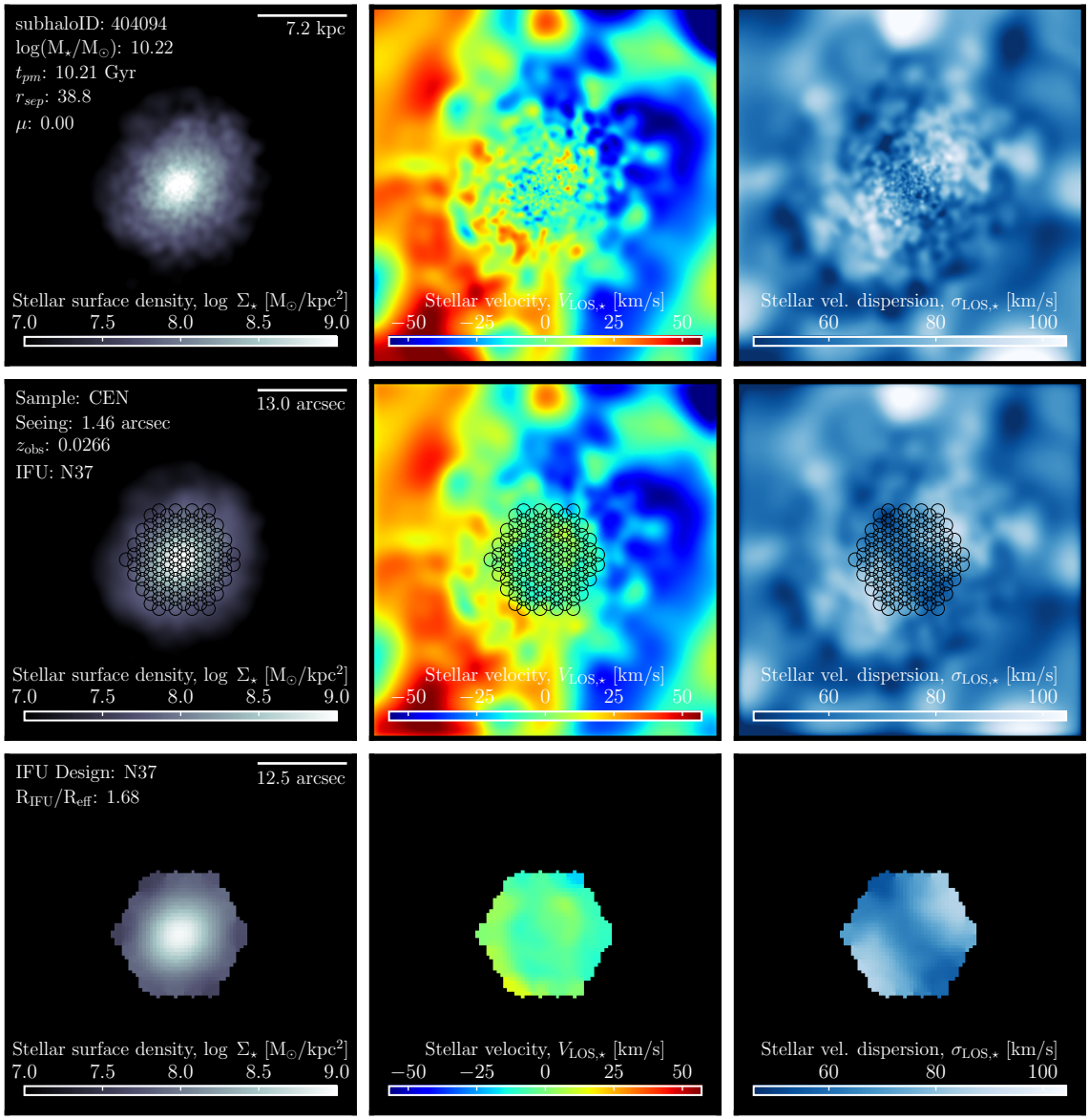

Figure G.21 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404094, camera: 1 observed for the CEN sample.

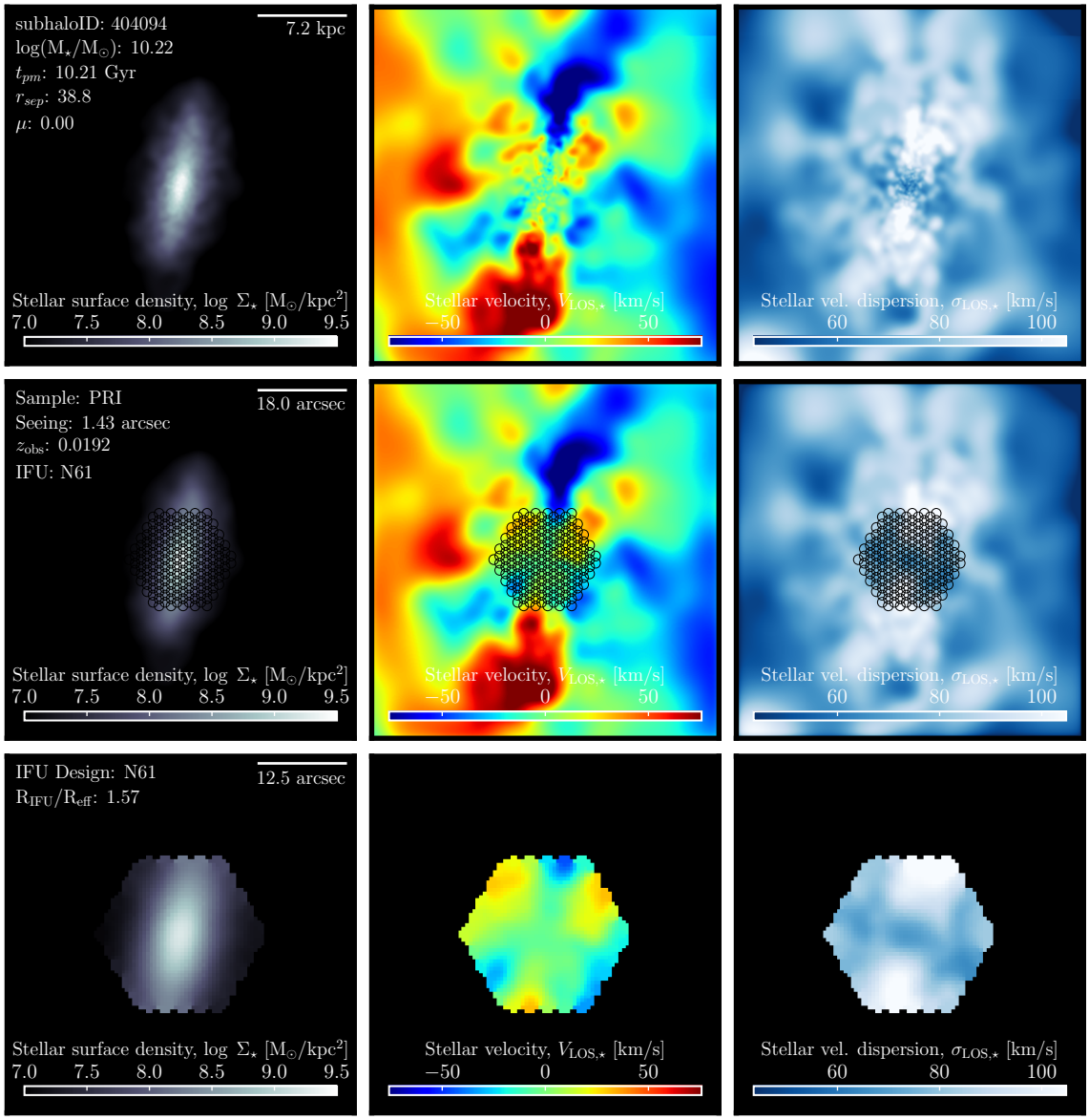

Figure G.22 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404094, camera: 3 observed for the PRI sample.

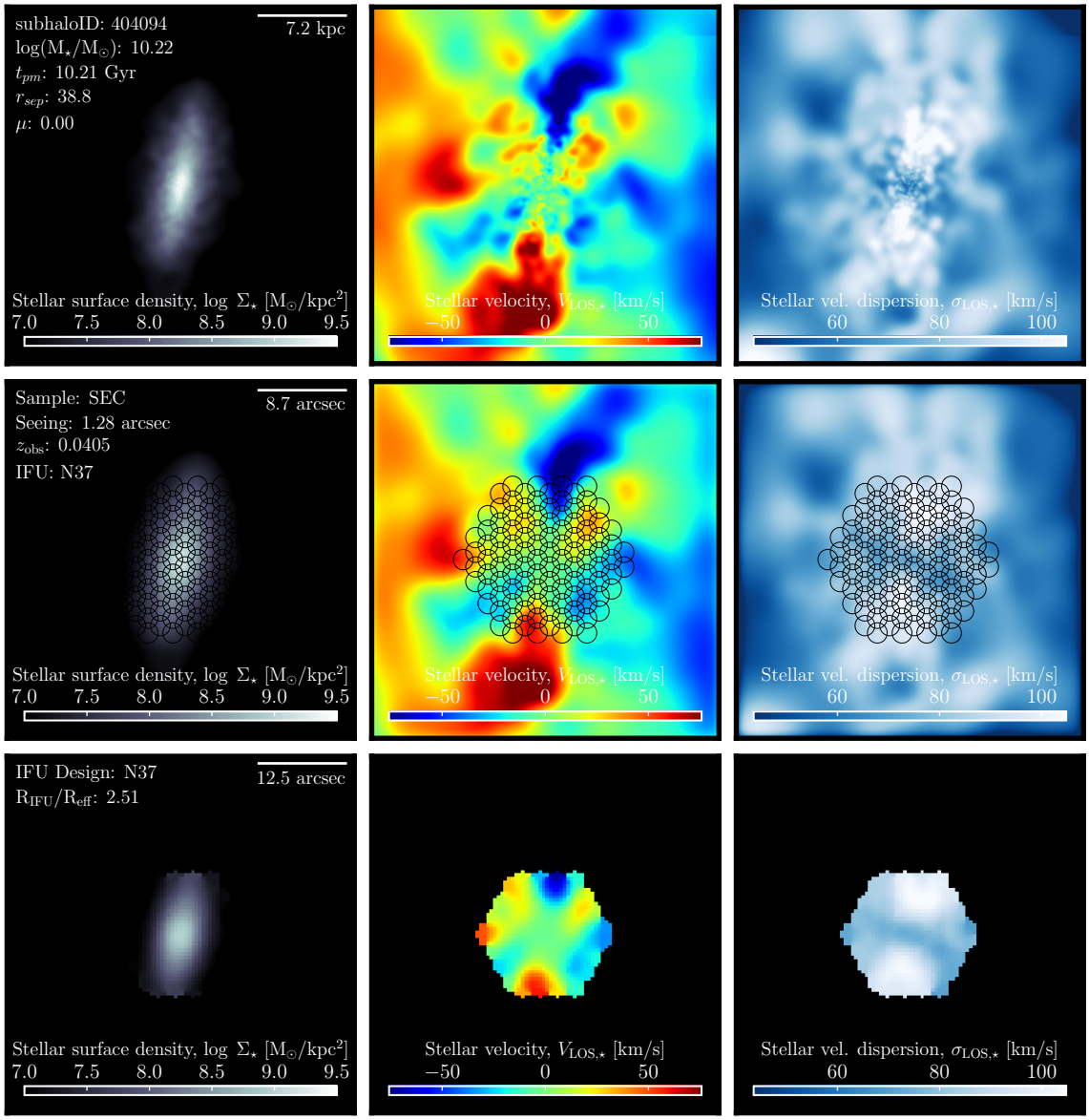

Figure G.23 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404094, camera: 3 observed for the SEC sample.

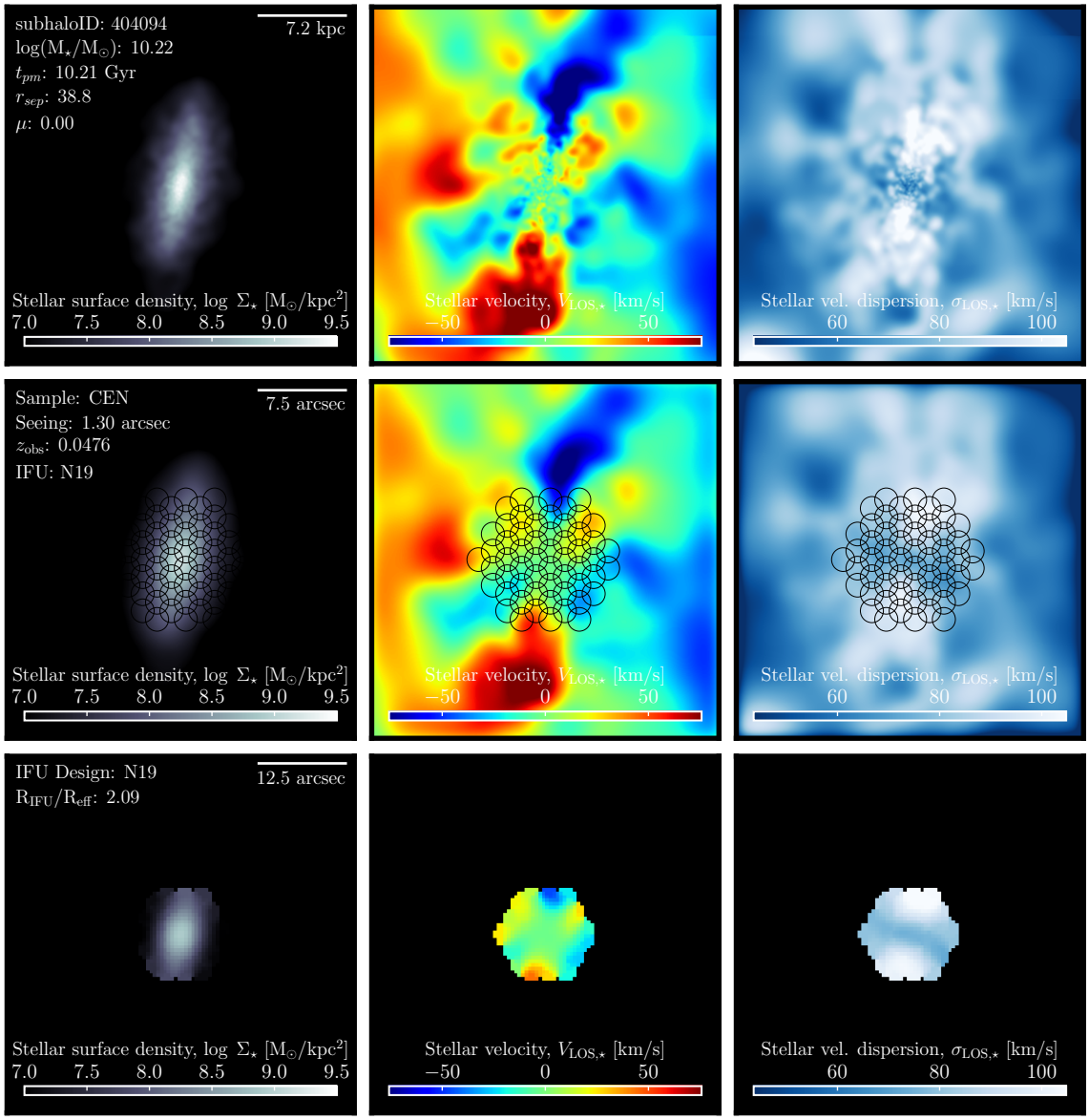

Figure G.24 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404094, camera: 3 observed for the CEN sample.

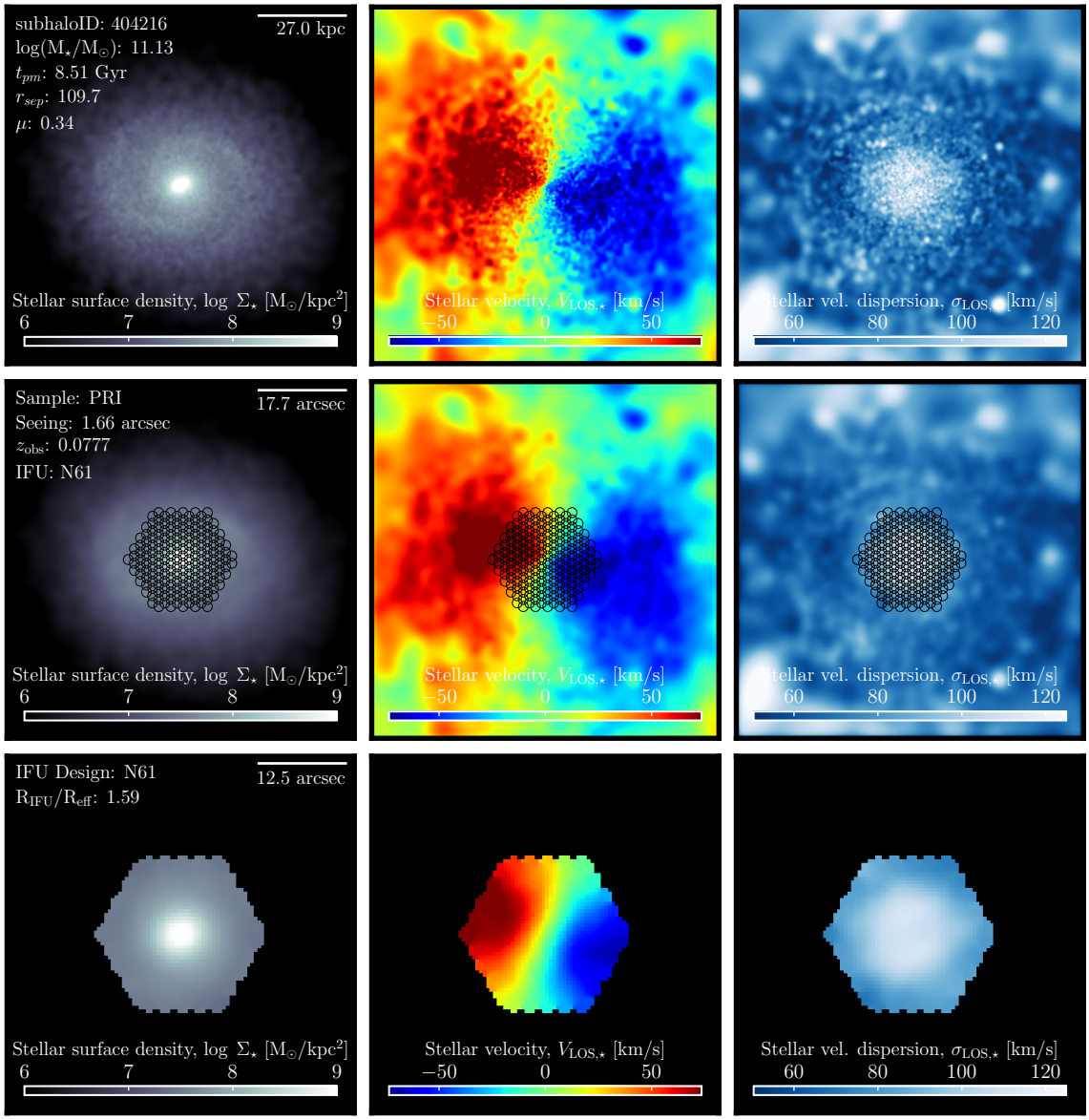

Figure G.25 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404216, camera: 1 observed for the PRI sample.

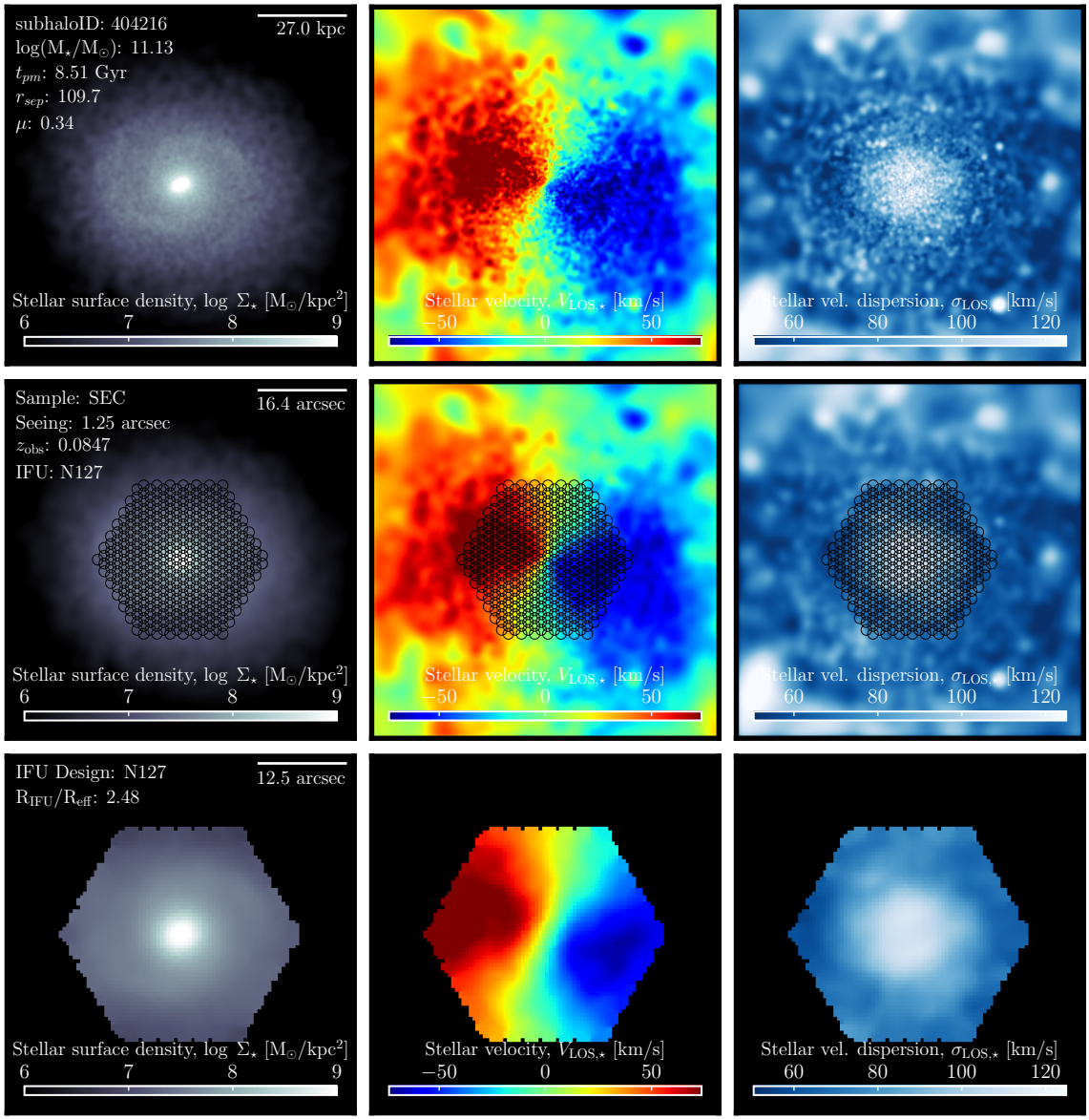

Figure G.26 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404216, camera: 1 observed for the SEC sample.

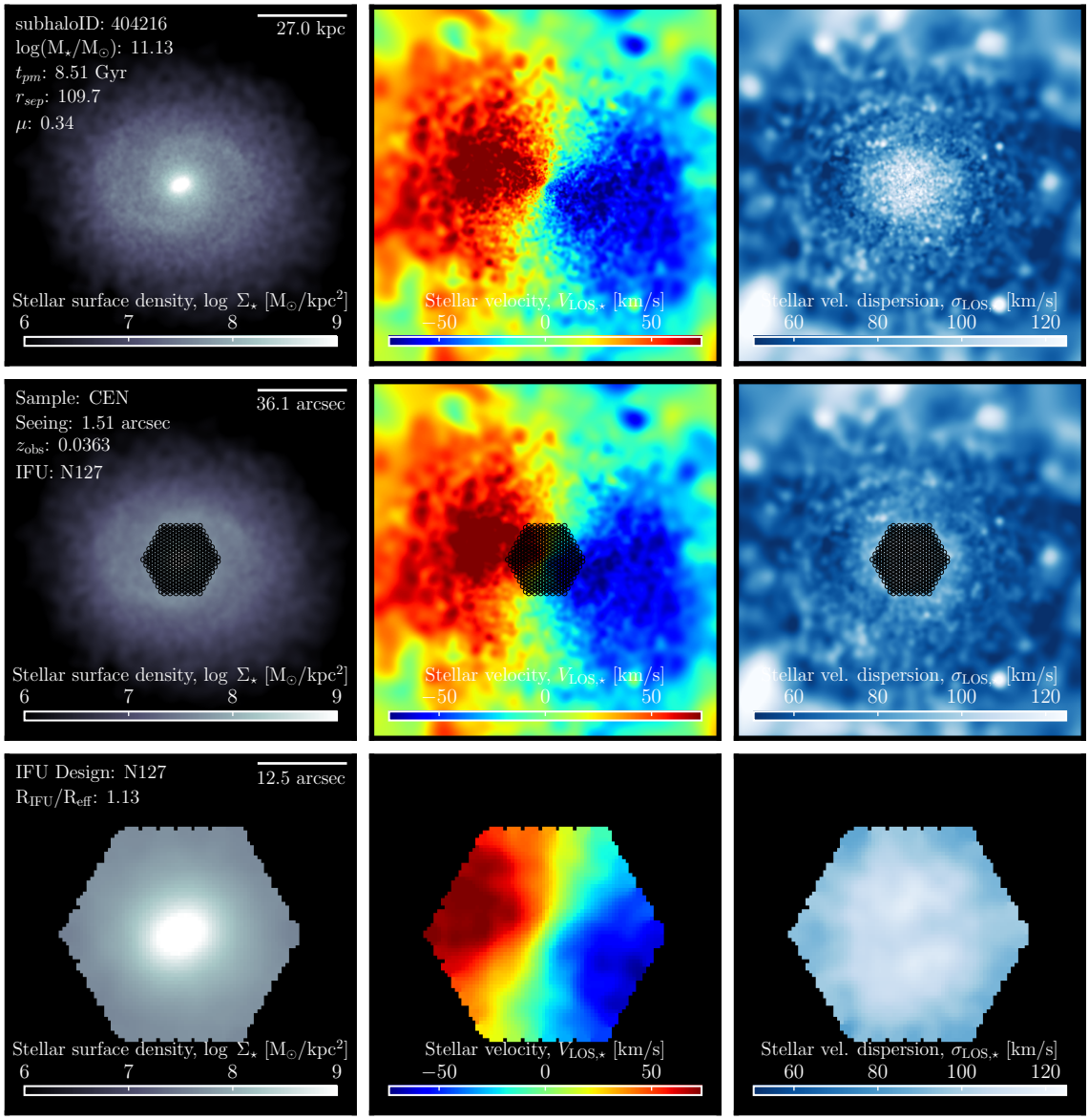

Figure G.27 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404216, camera: 1 observed for the CEN sample.

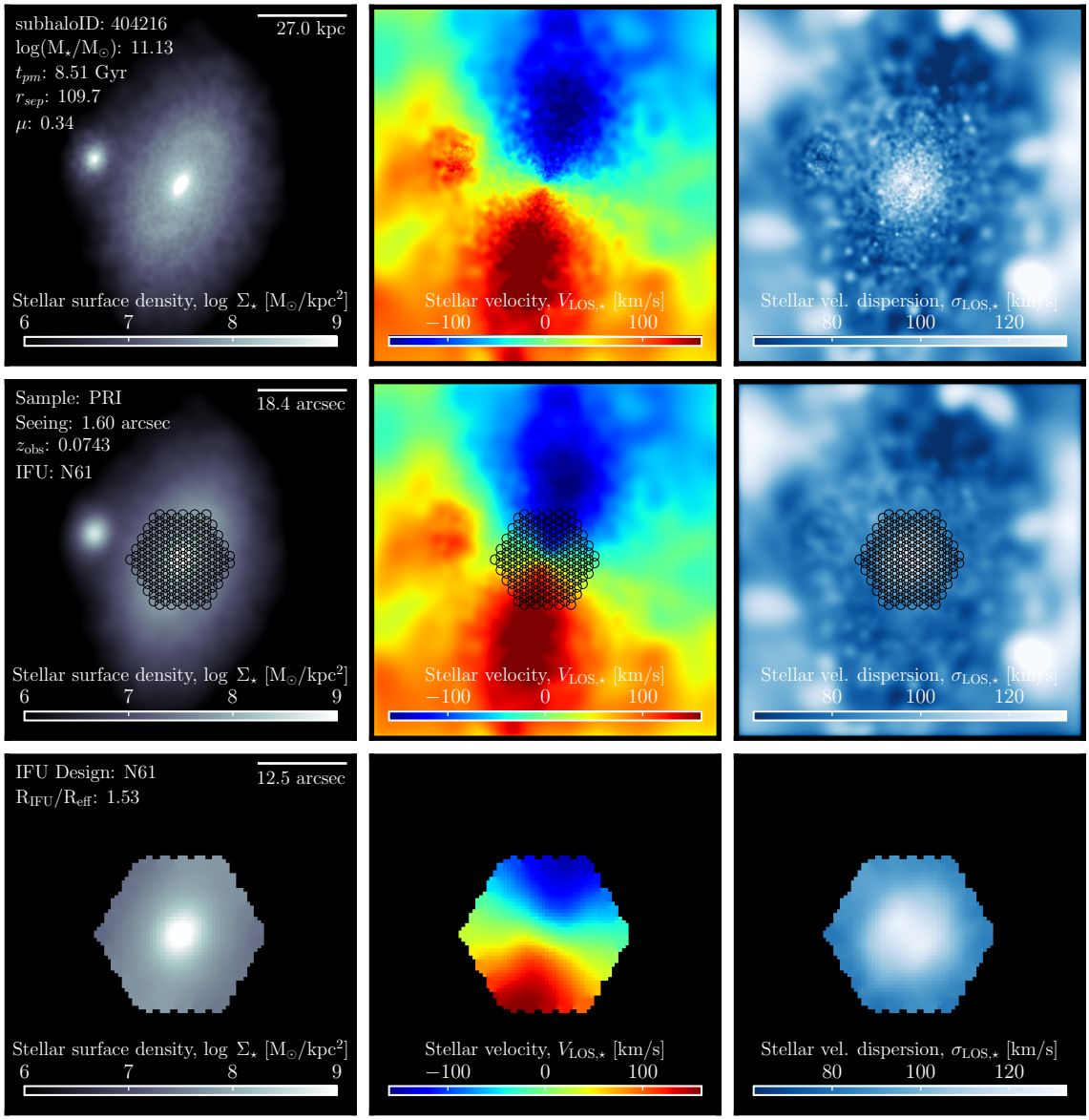

Figure G.28 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404216, camera: 3 observed for the PRI sample.

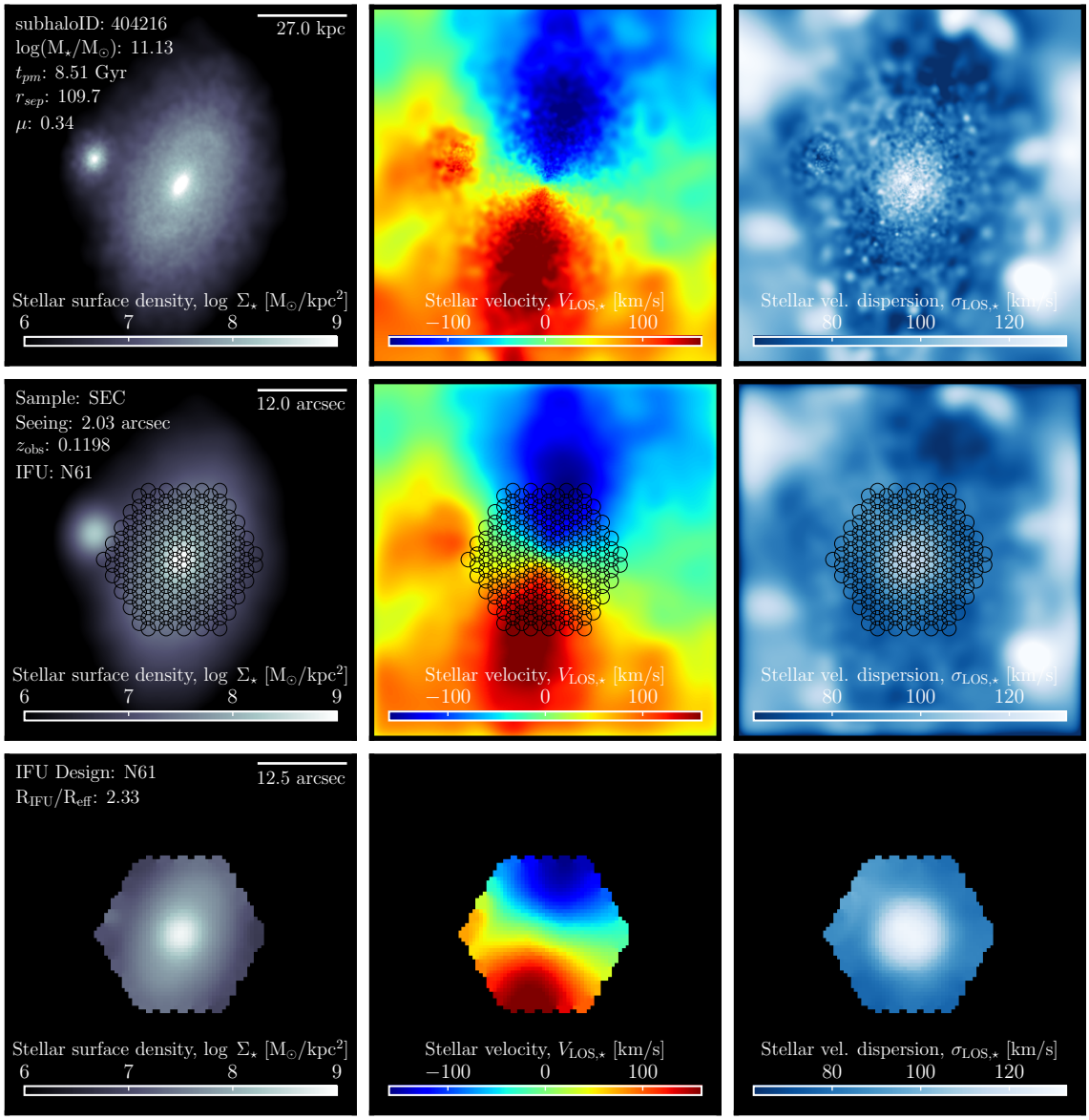

Figure G.29 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404216, camera: 3 observed for the SEC sample.

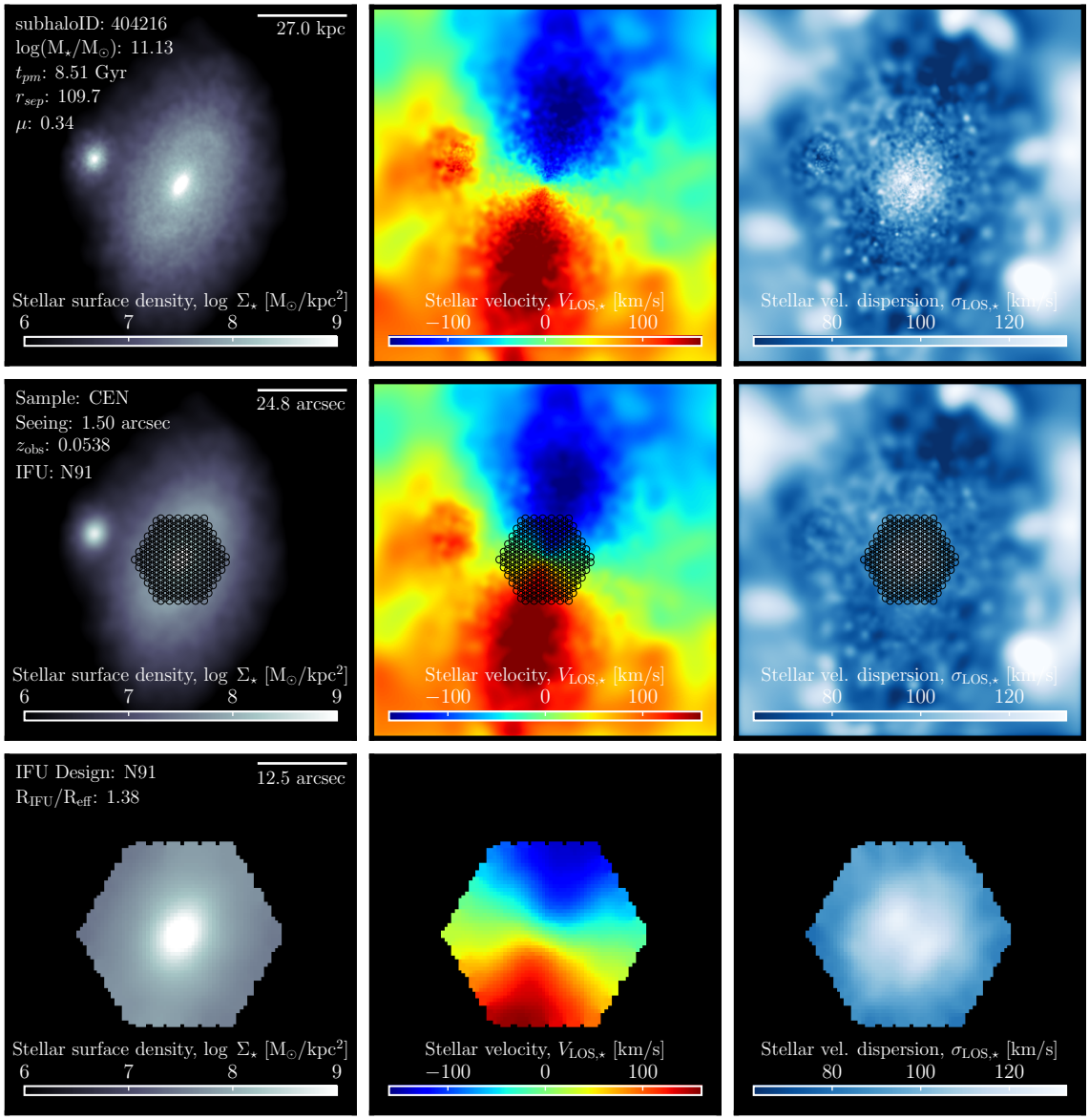

Figure G.30 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404216, camera: 3 observed for the CEN sample.

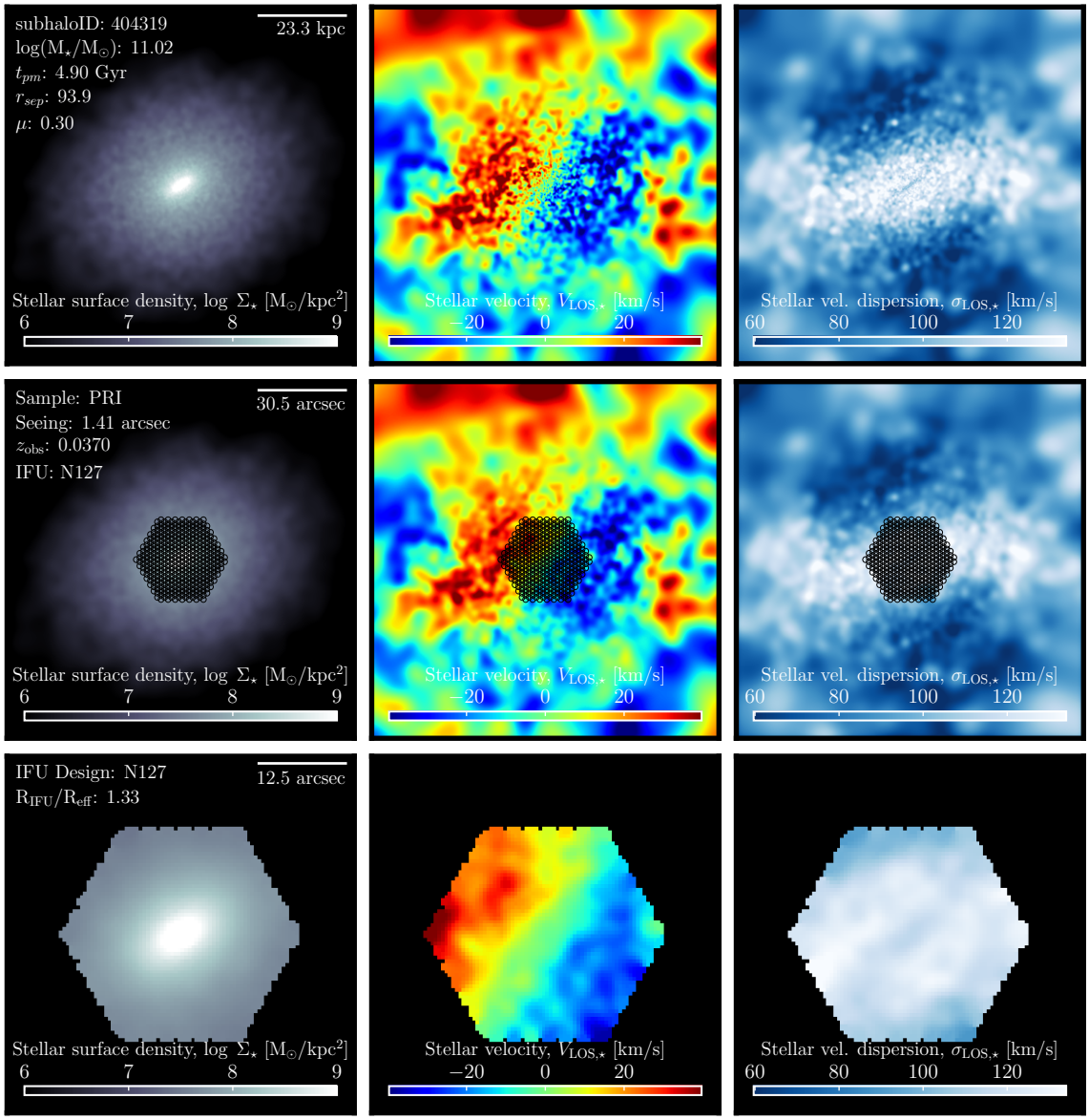

Figure G.31 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404319, camera: 1 observed for the PRI sample.

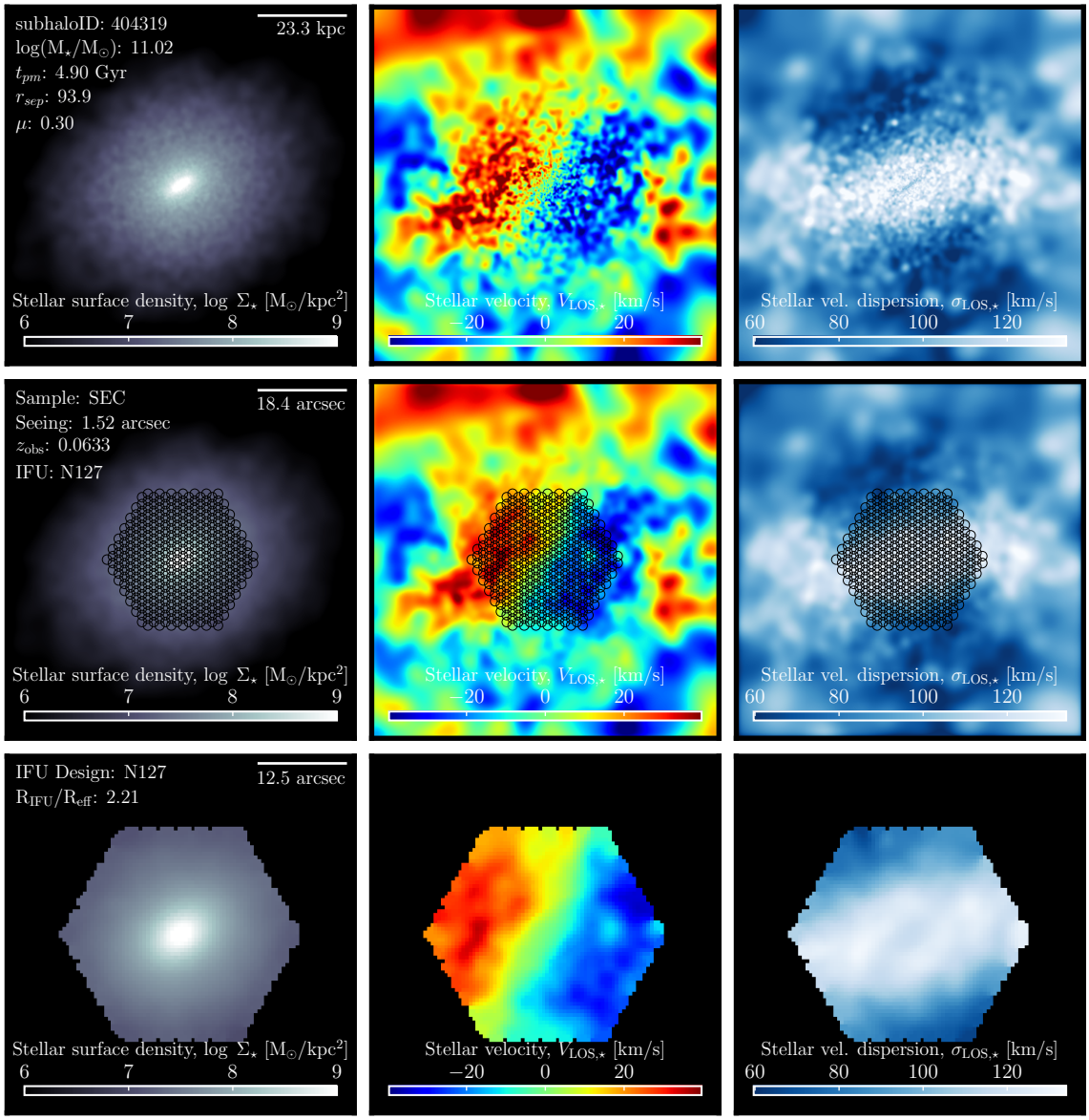

Figure G.32 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404319, camera: 1 observed for the SEC sample.

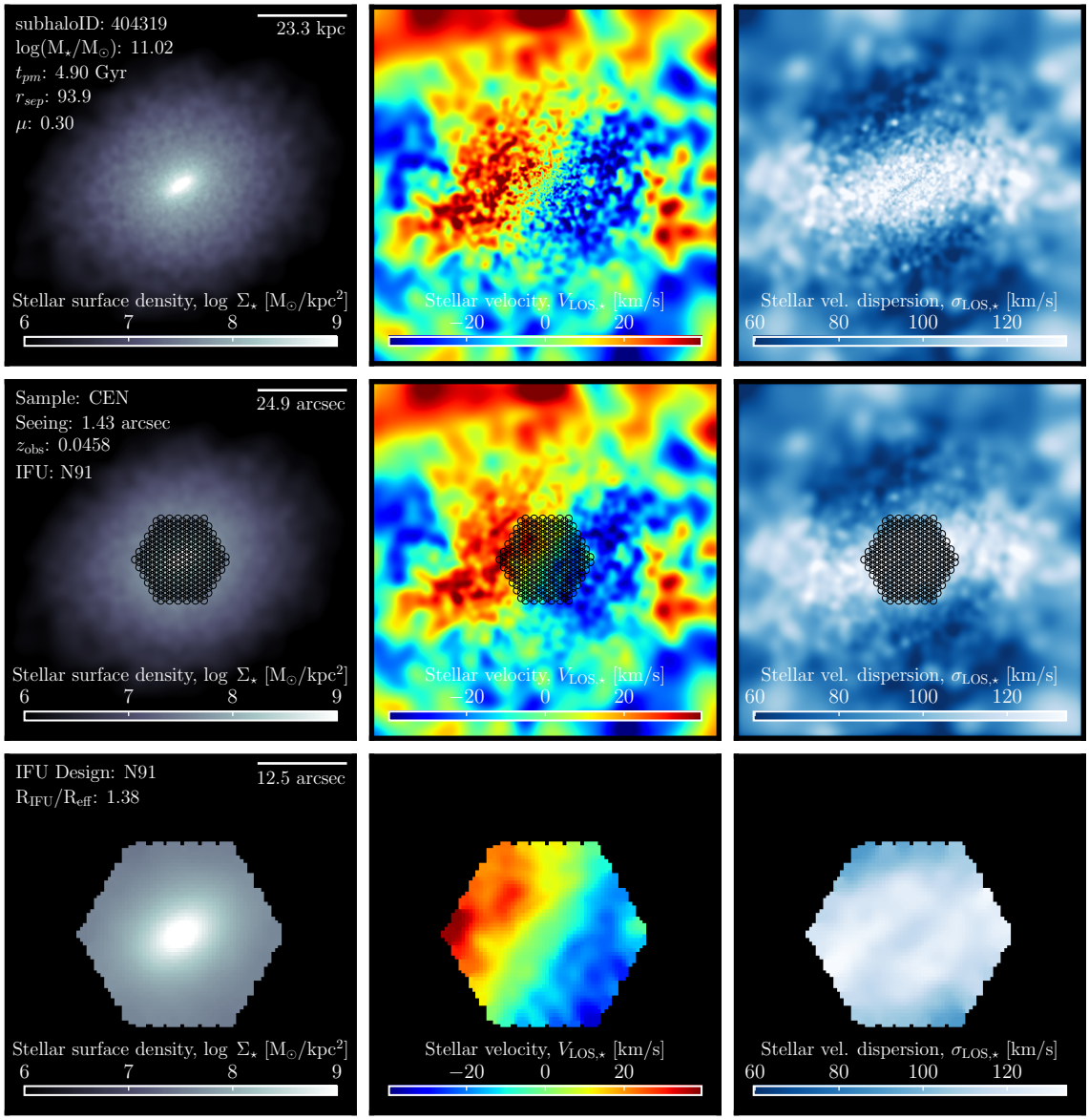

Figure G.33 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404319, camera: 1 observed for the CEN sample.

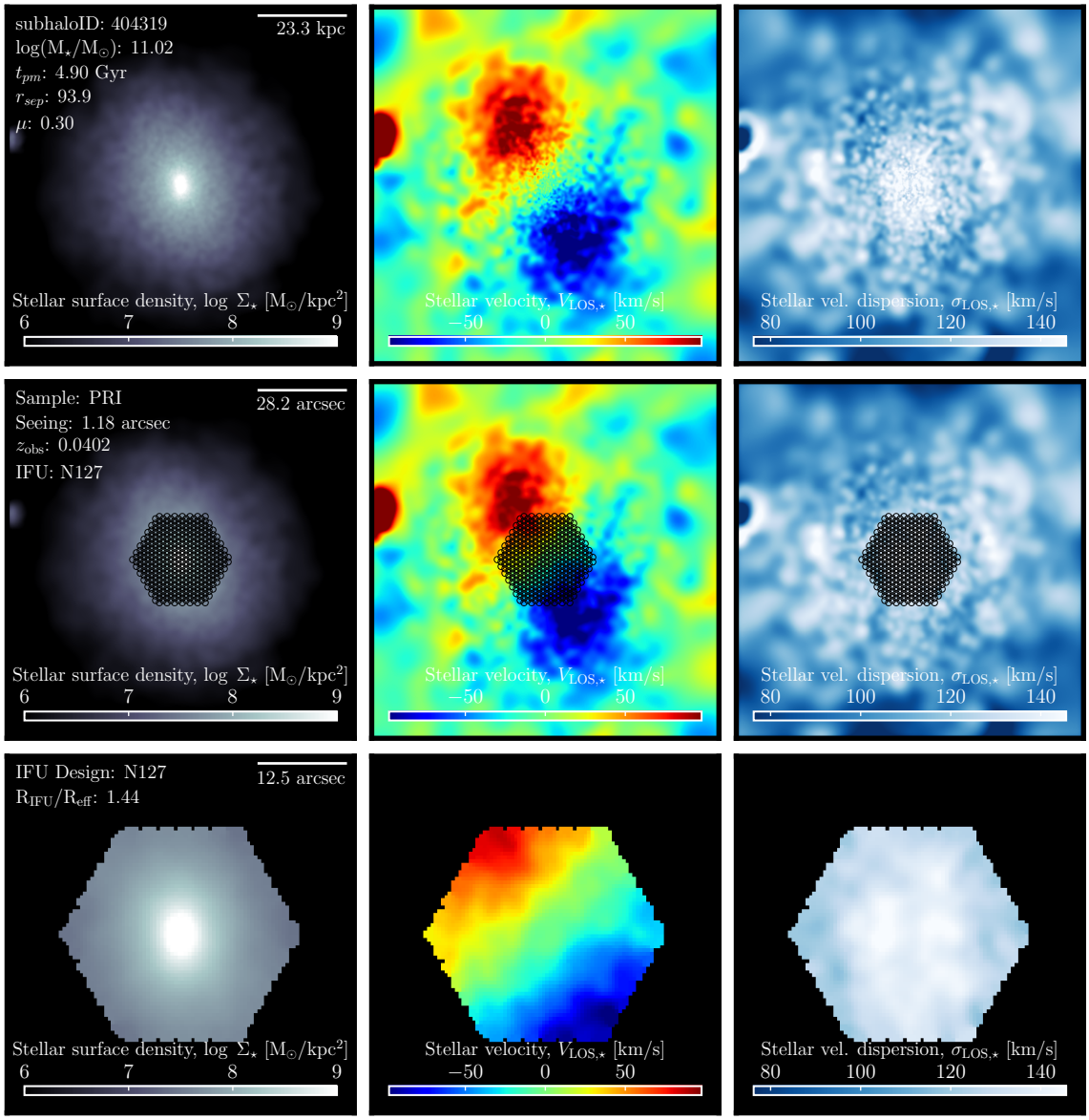

Figure G.34 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404319, camera: 3 observed for the PRI sample.

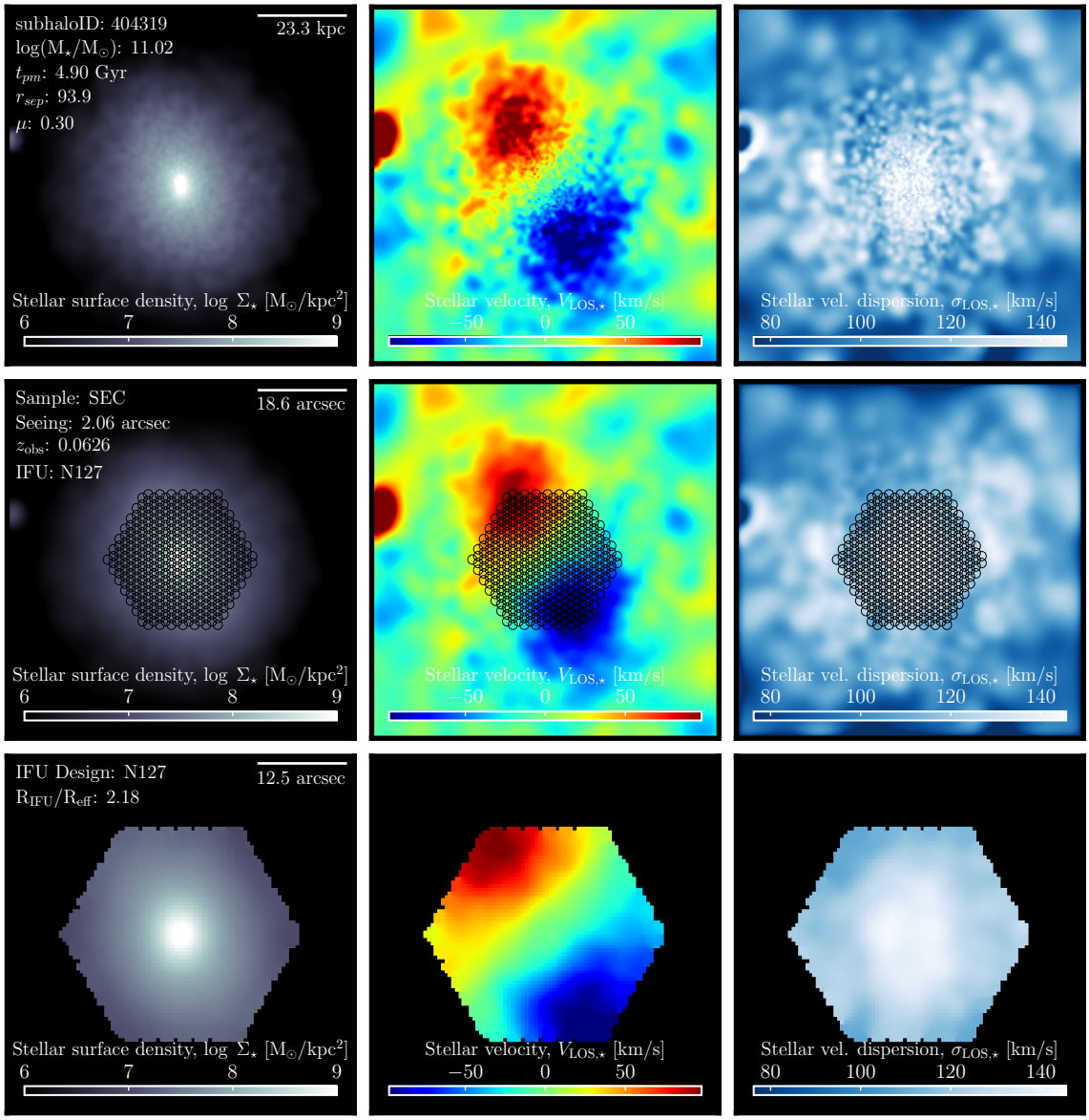

Figure G.35 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404319, camera: 3 observed for the SEC sample.

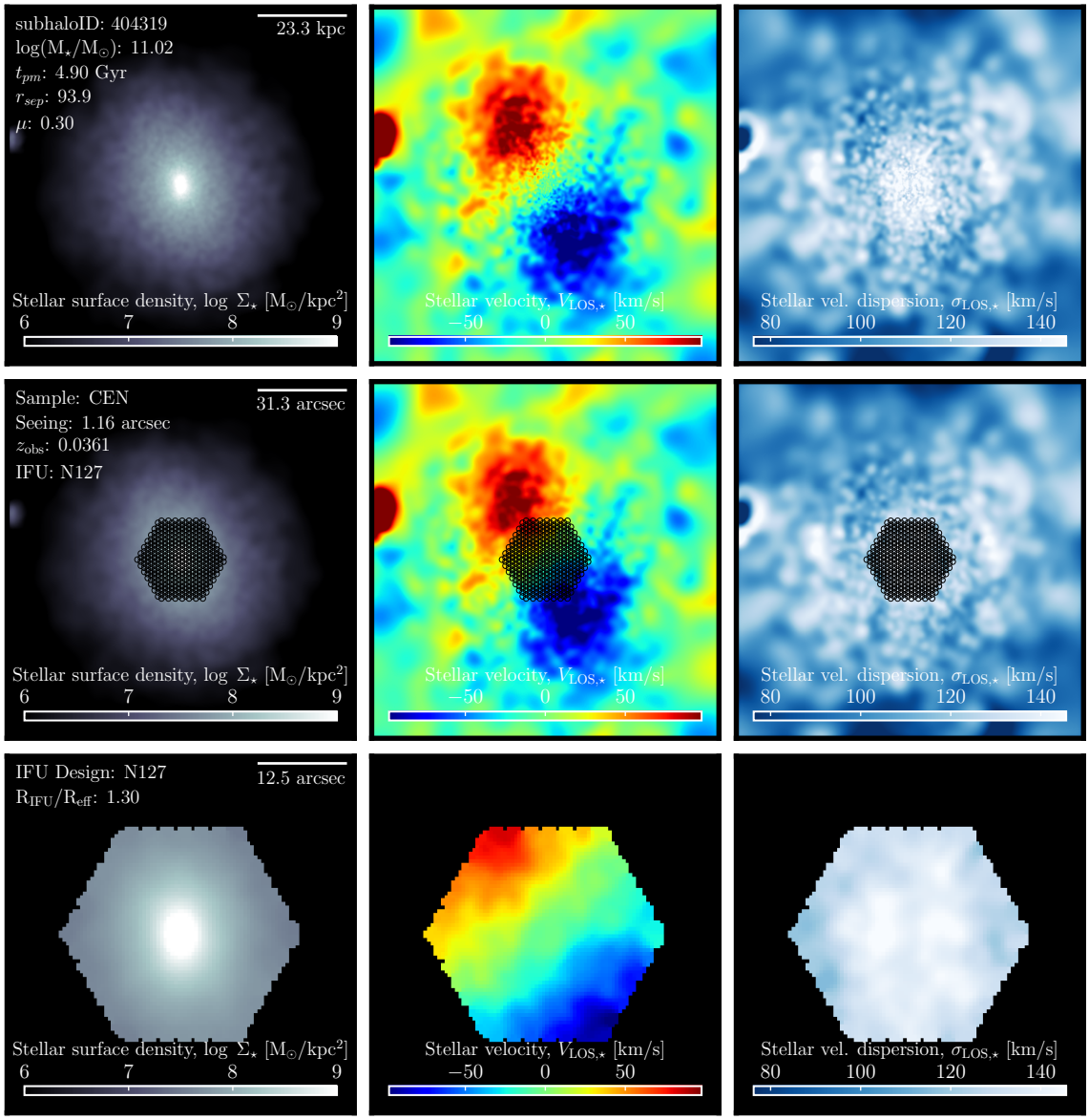

Figure G.36 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404319, camera: 3 observed for the CEN sample.

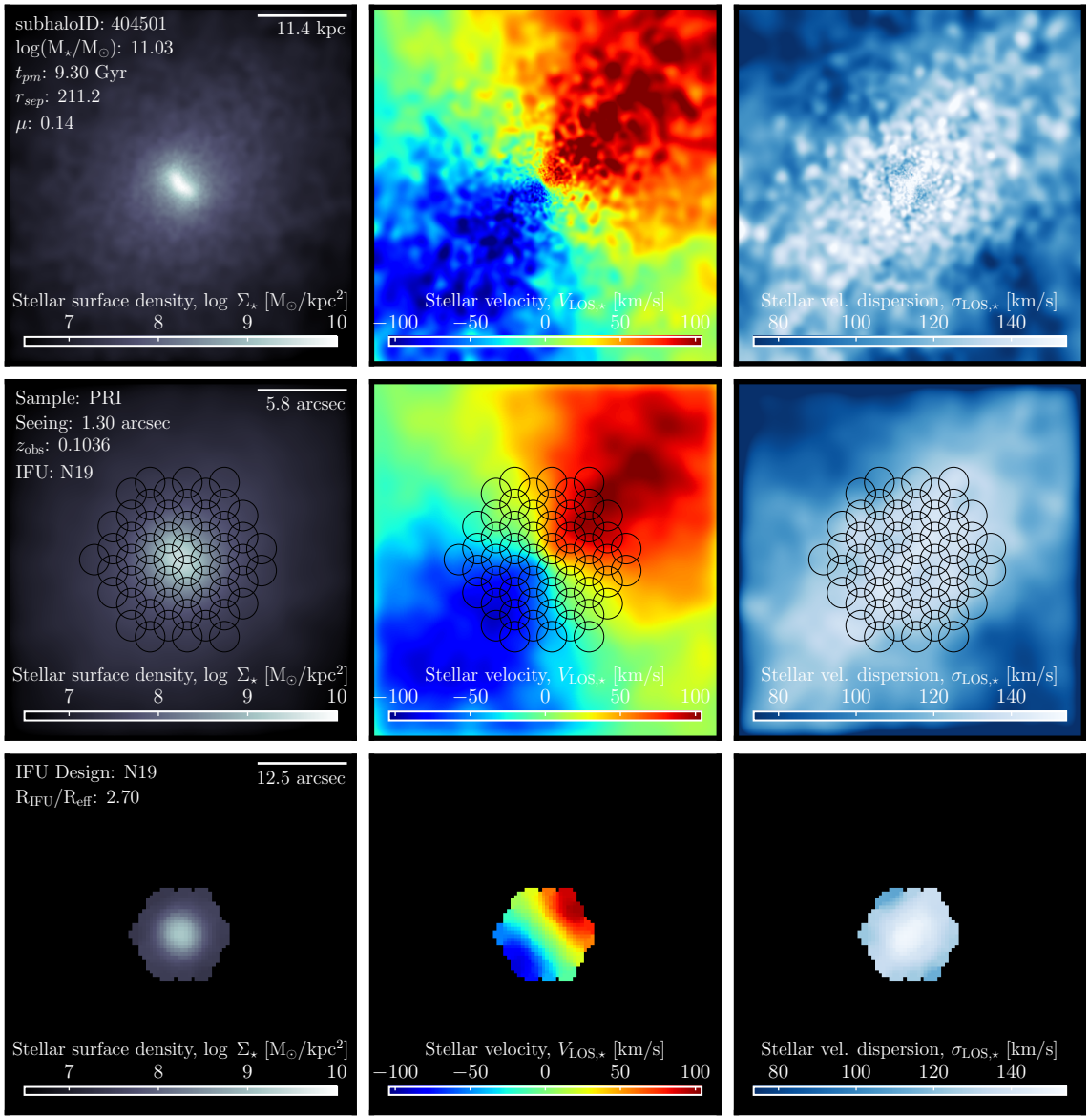

Figure G.37 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404501, camera: 1 observed for the PRI sample.

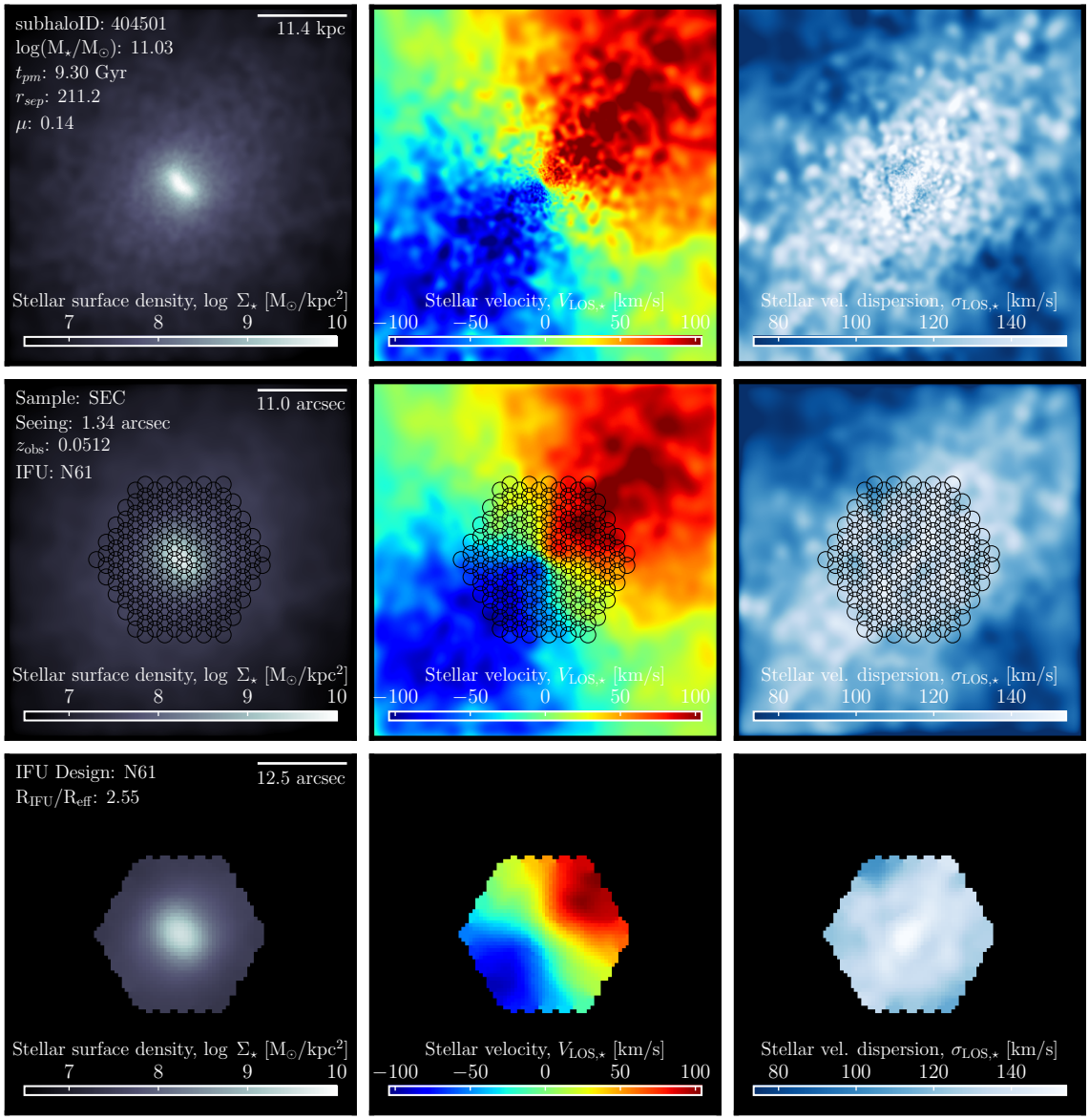

Figure G.38 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404501, camera: 1 observed for the SEC sample.

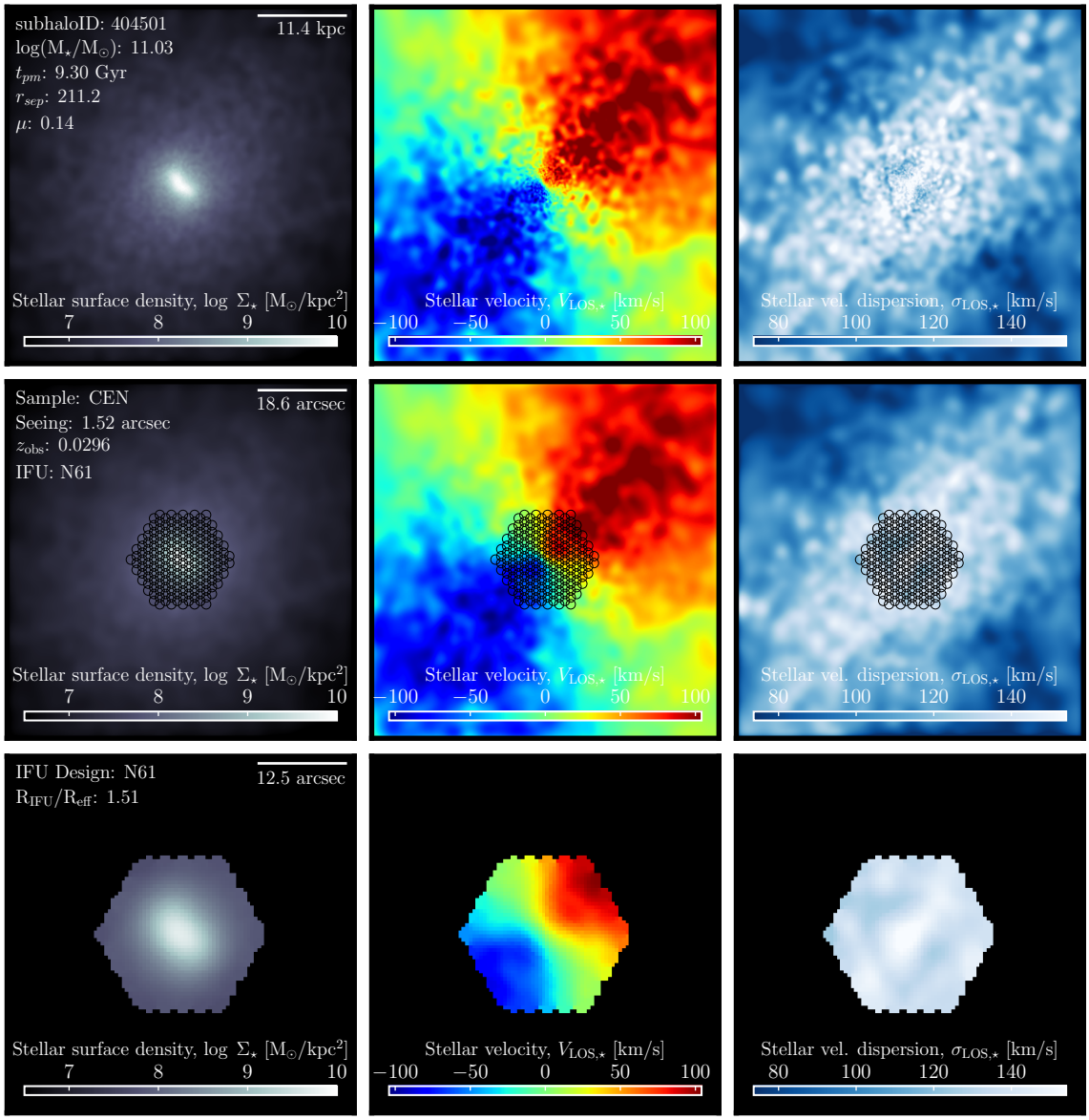

Figure G.39 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404501, camera: 1 observed for the CEN sample.

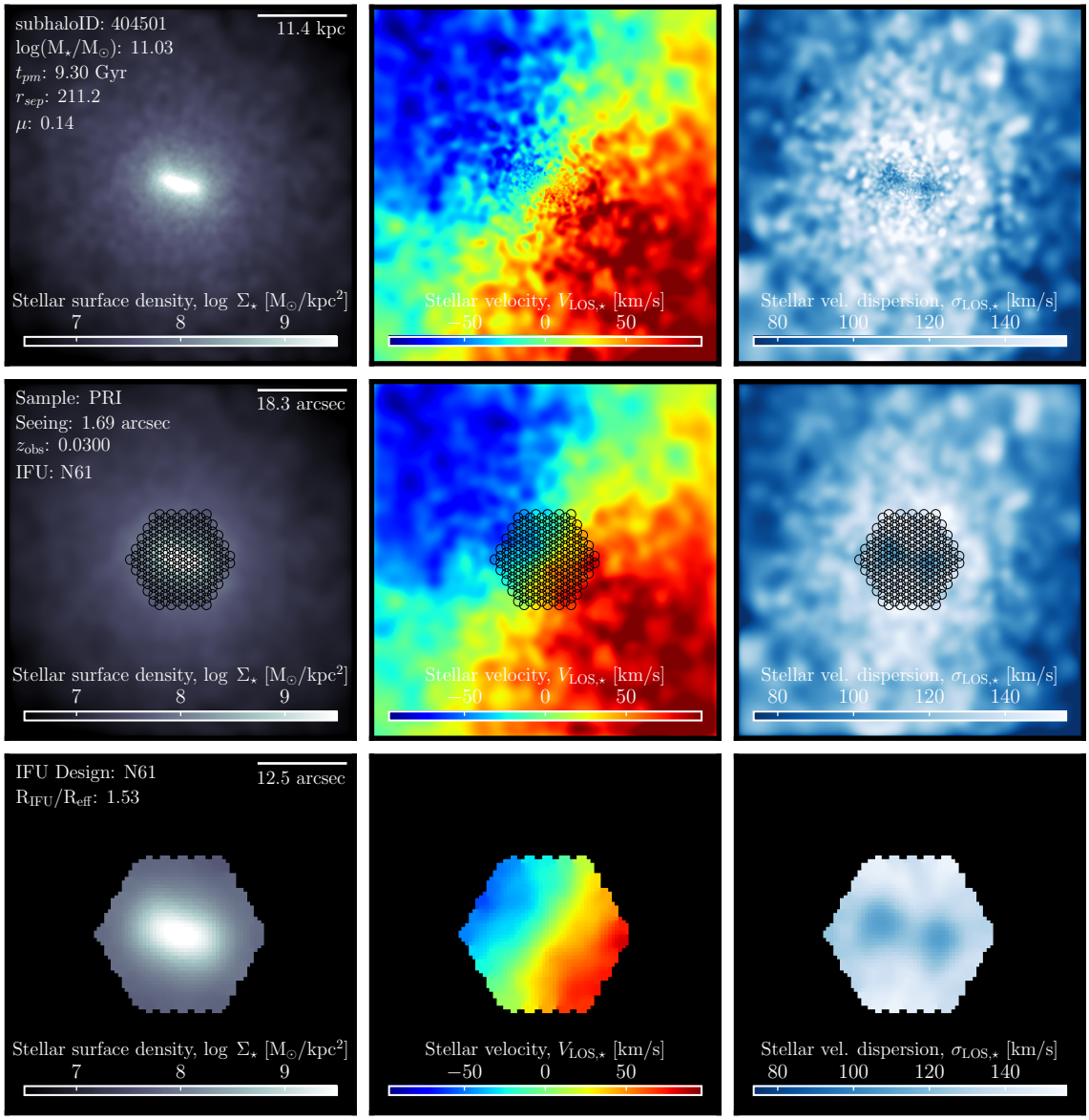

Figure G.40 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404501, camera: 3 observed for the PRI sample.

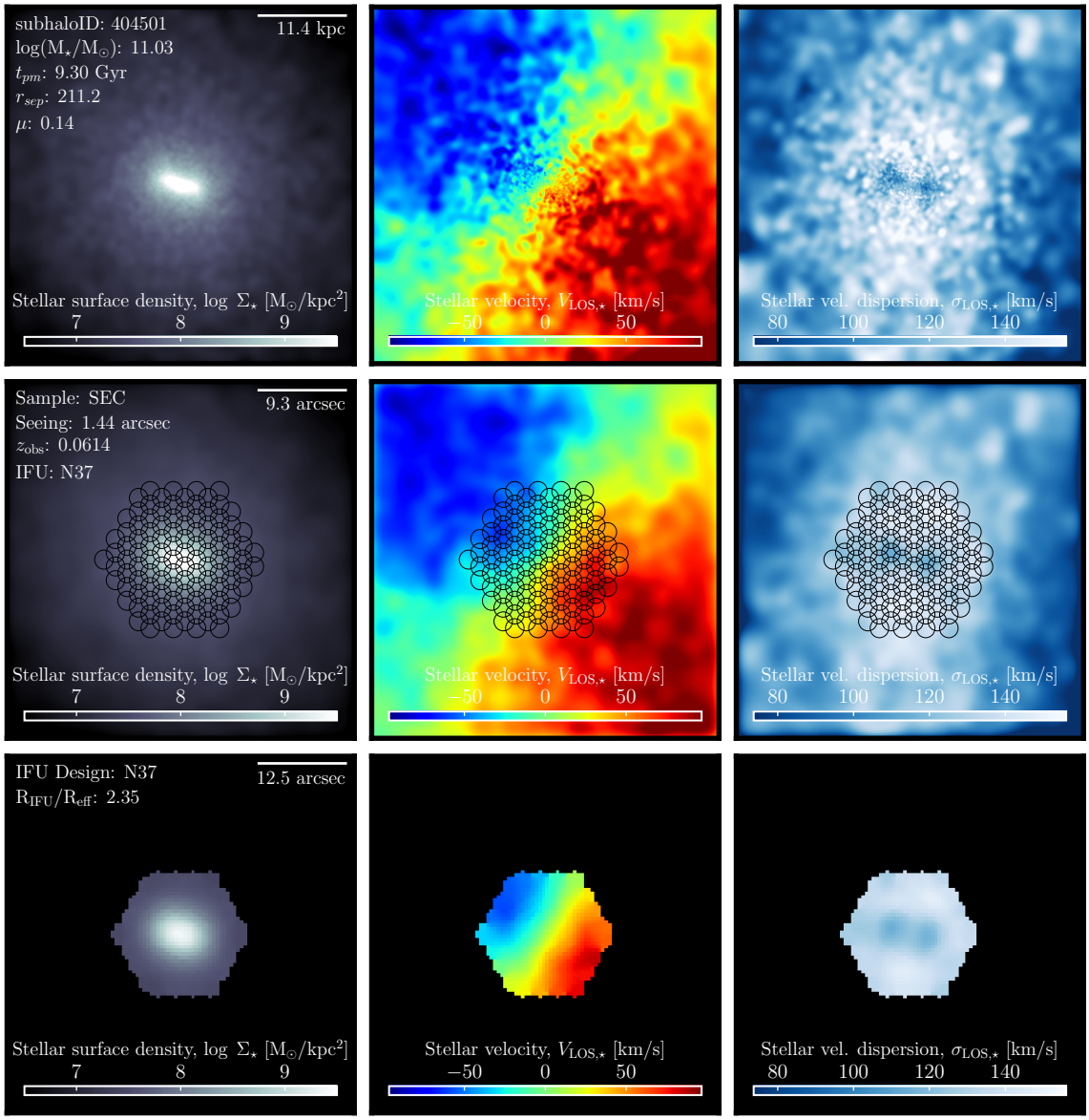

Figure G.41 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404501, camera: 3 observed for the SEC sample.

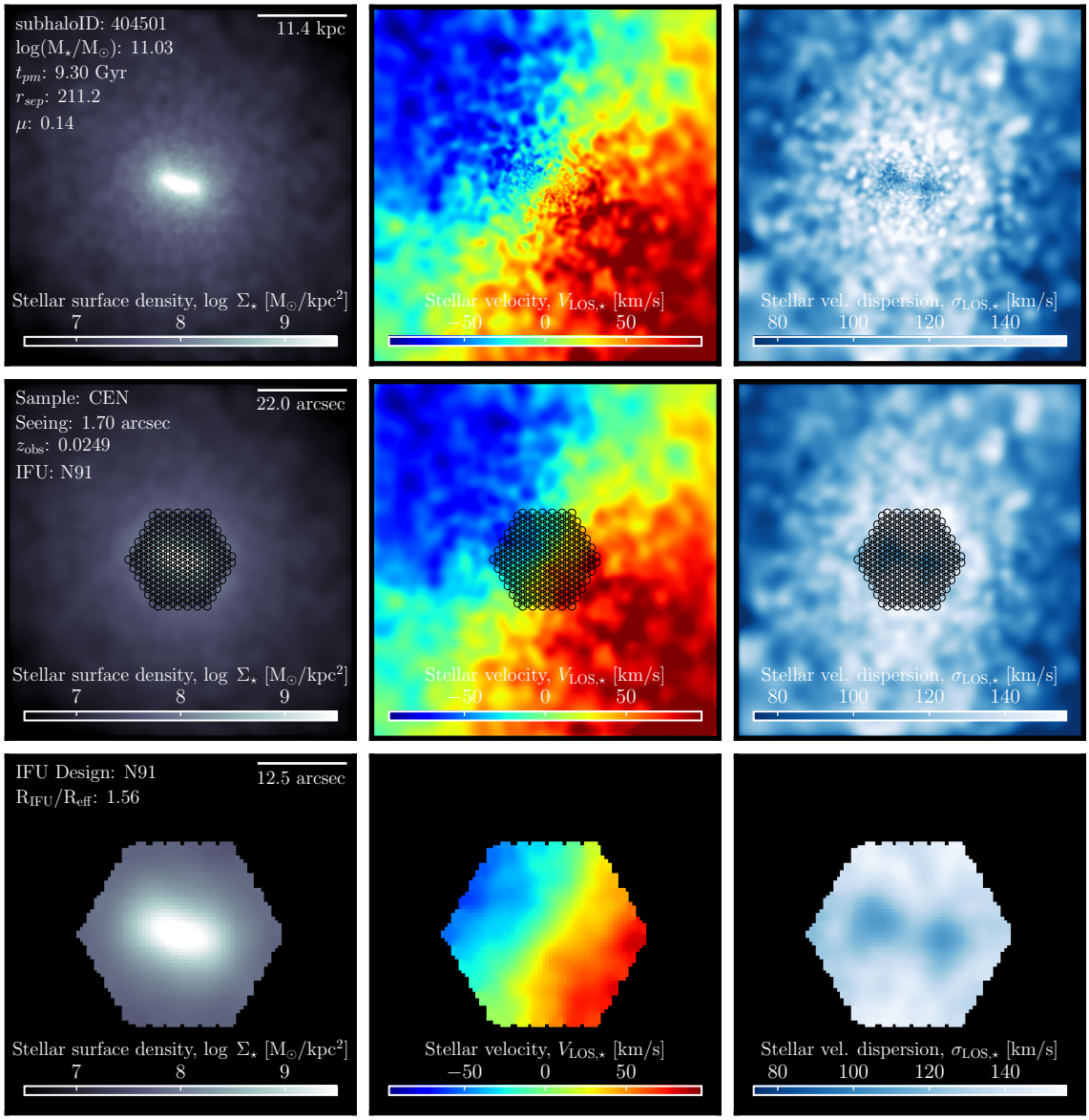

Figure G.42 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 404501, camera: 3 observed for the CEN sample.

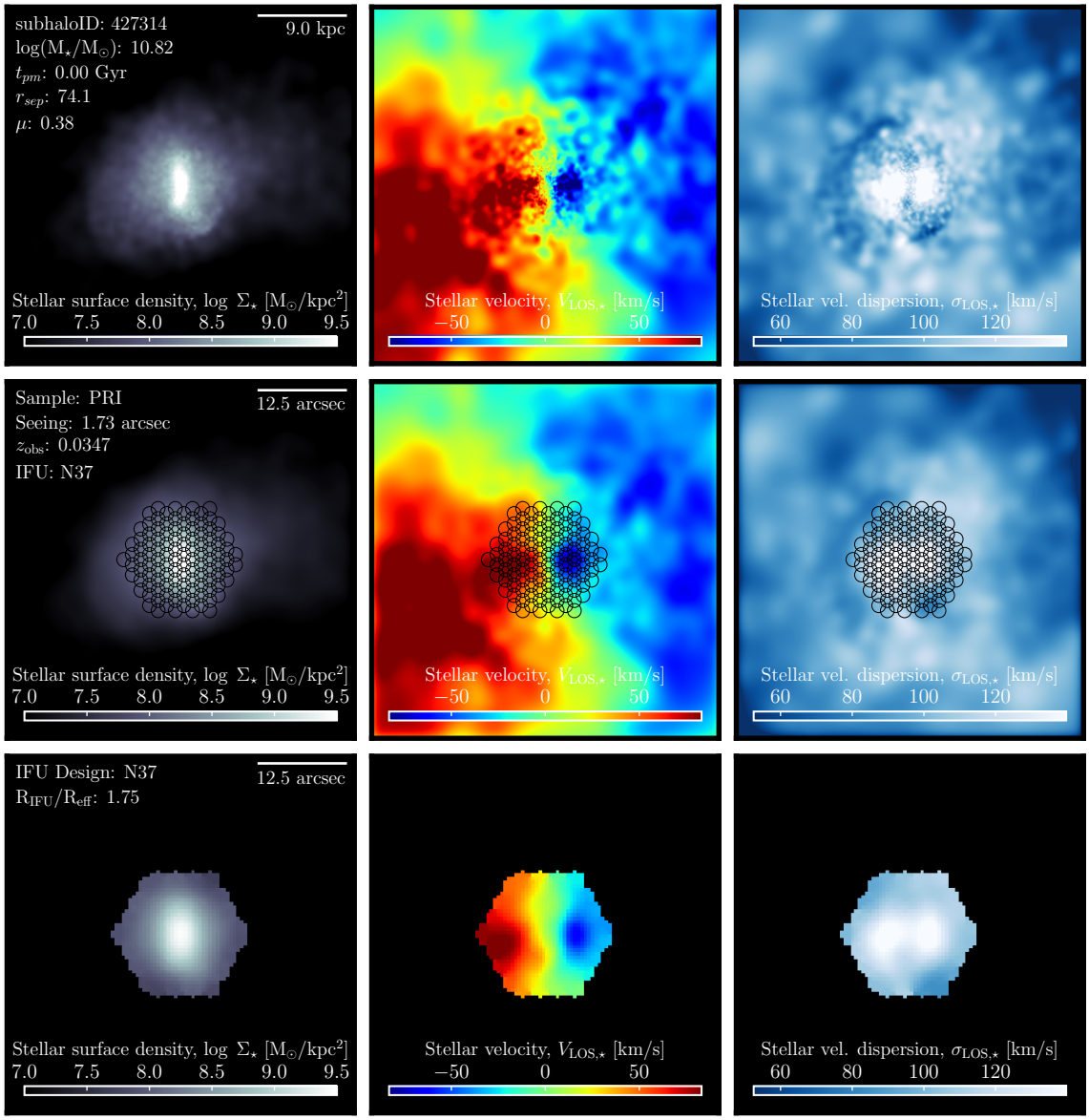

Figure G.43 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 427314, camera: 1 observed for the PRI sample.

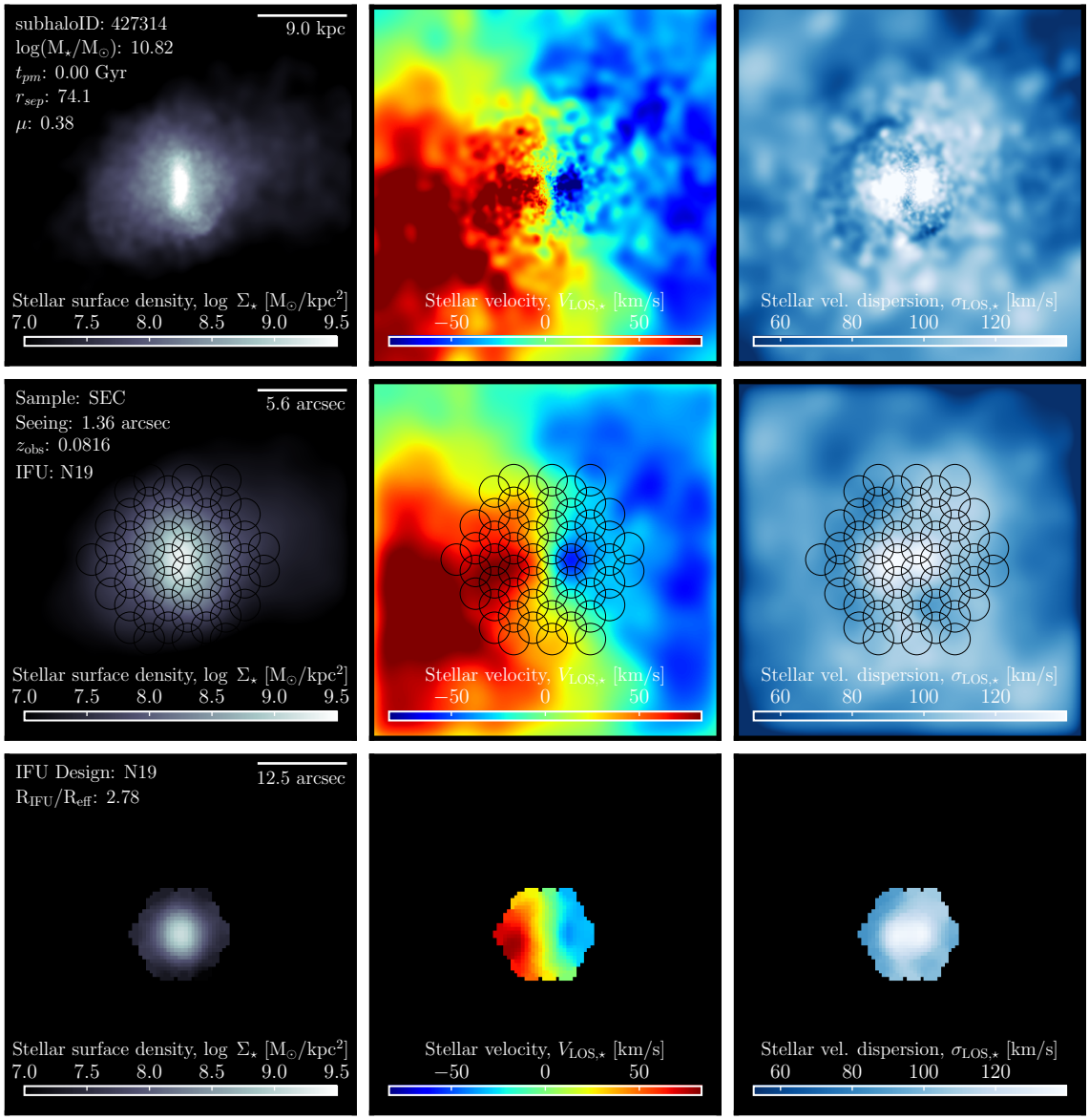

Figure G.44 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 427314, camera: 1 observed for the SEC sample.

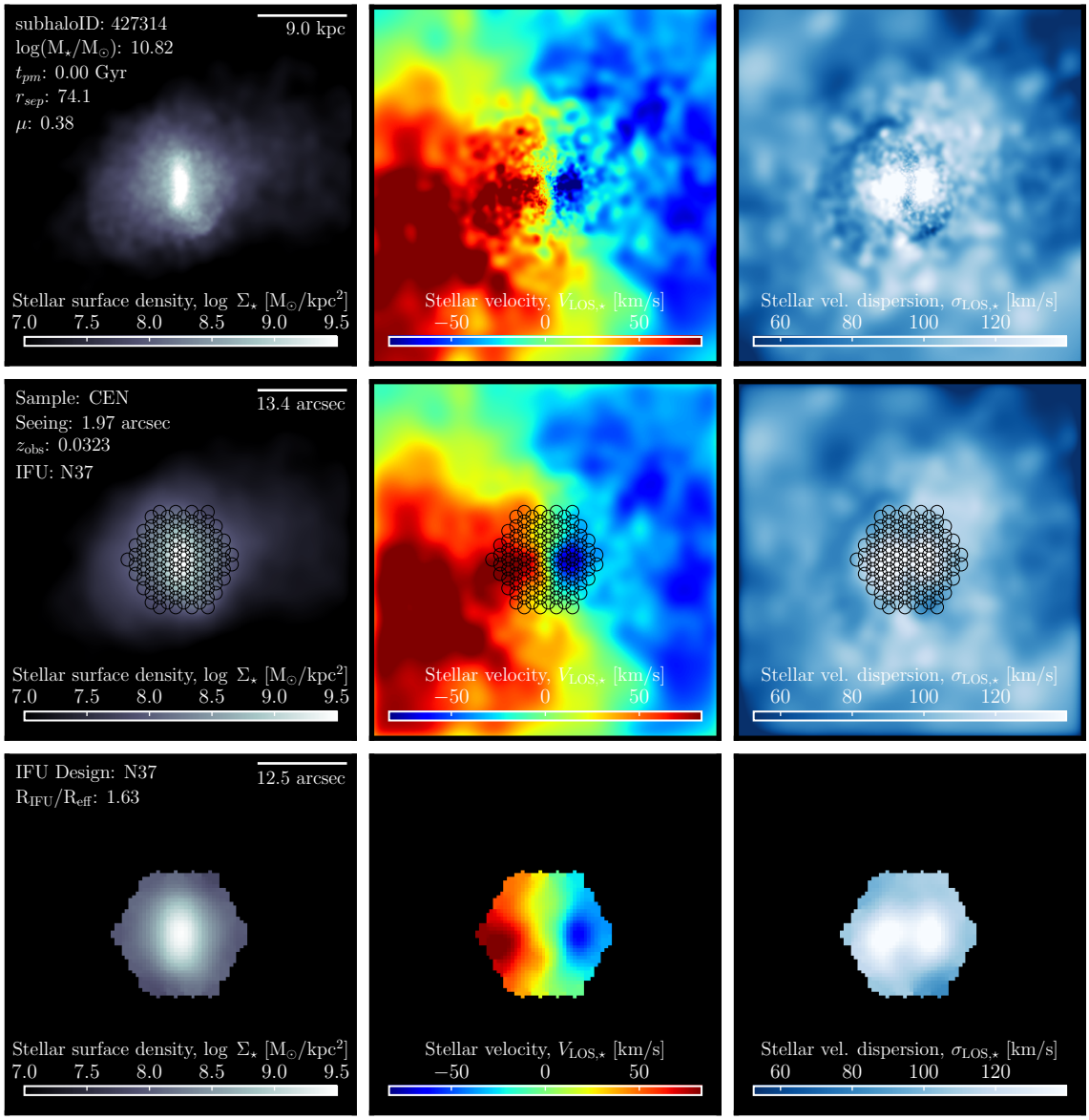

Figure G.45 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 427314, camera: 1 observed for the CEN sample.

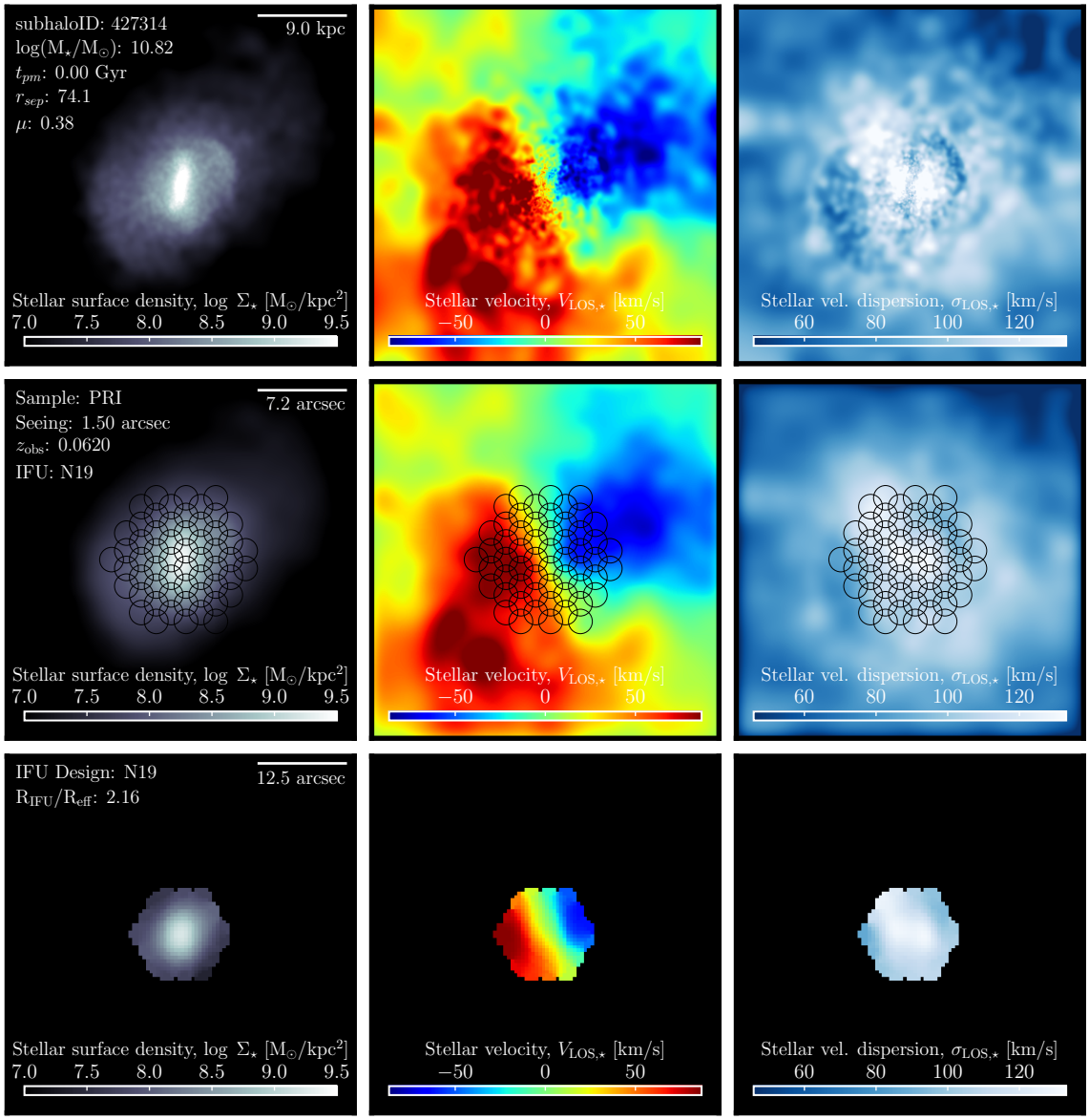

Figure G.46 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 427314, camera: 3 observed for the PRI sample.

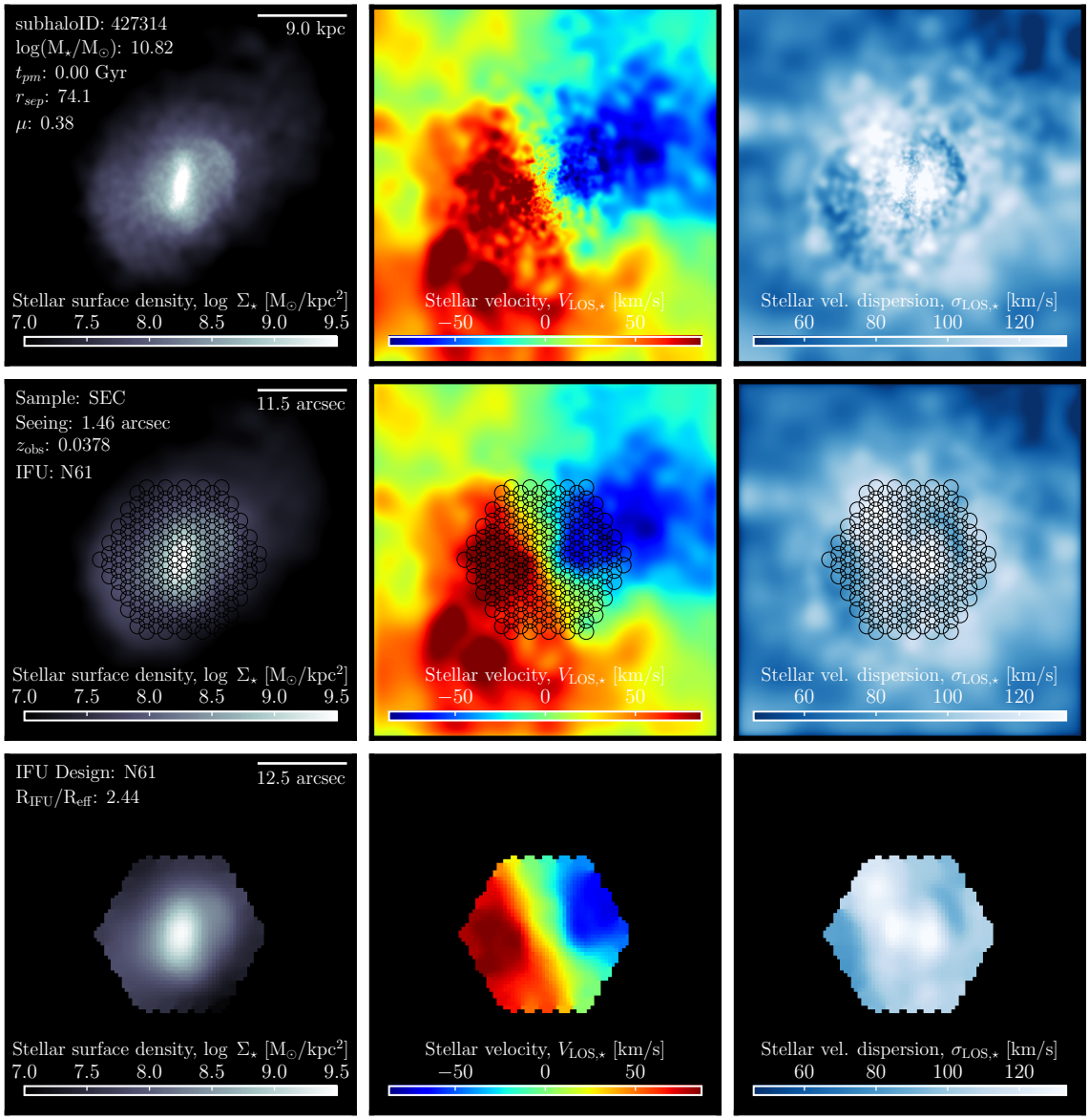

Figure G.47 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 427314, camera: 3 observed for the SEC sample.

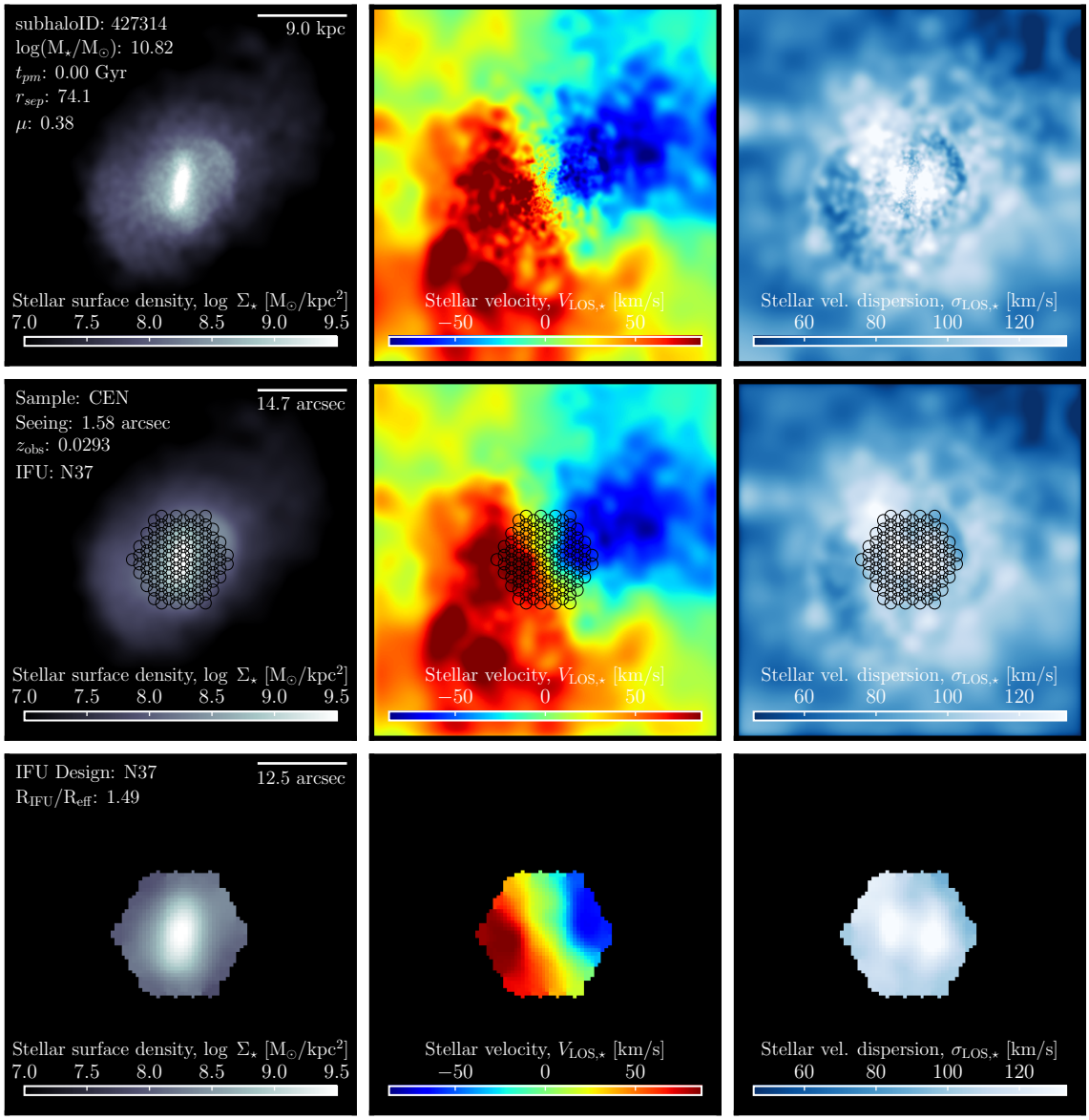

Figure G.48 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 427314, camera: 3 observed for the CEN sample.

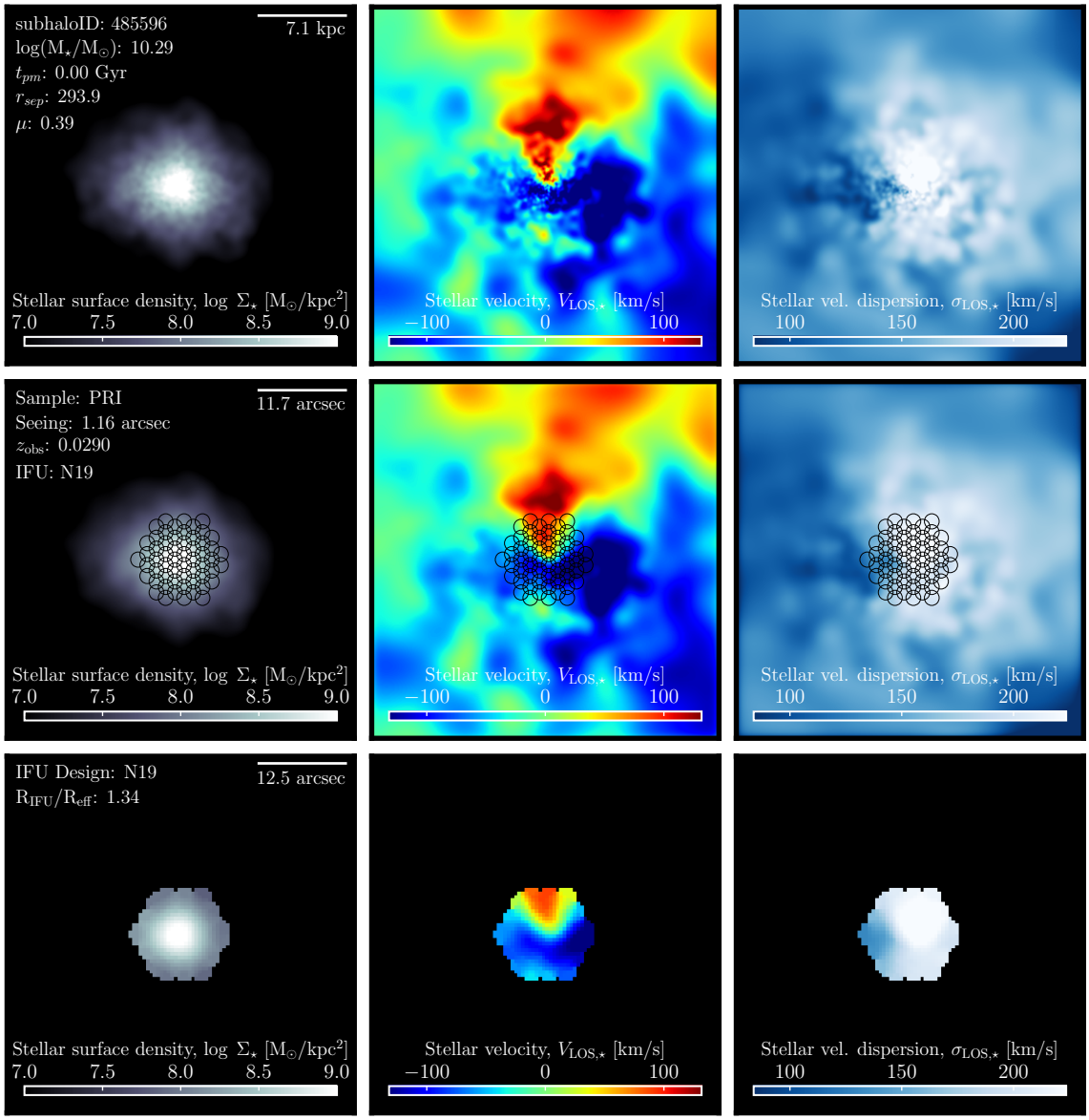

Figure G.49 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 485596, camera: 1 observed for the PRI sample.

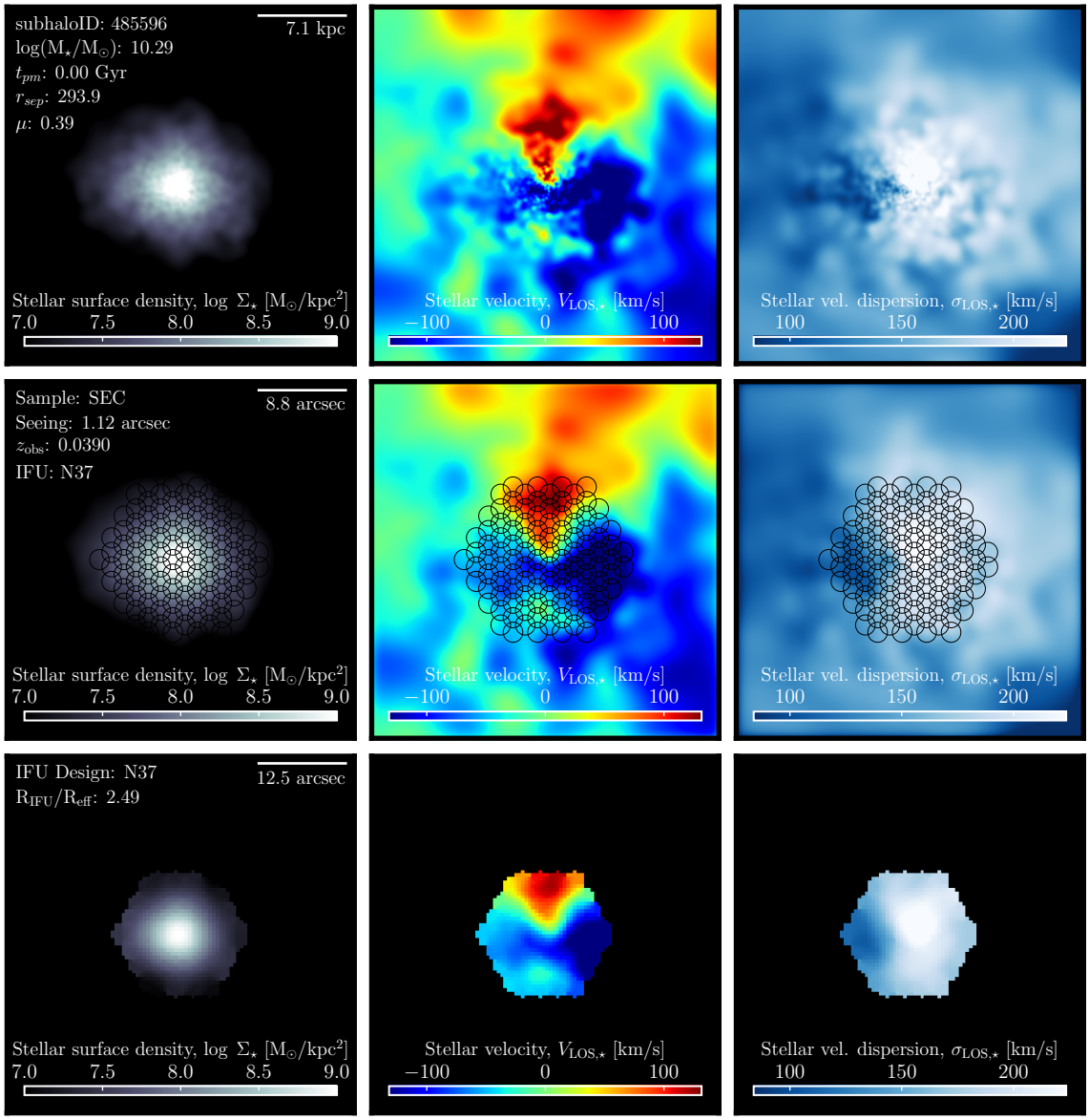

Figure G.50 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 485596, camera: 1 observed for the SEC sample.

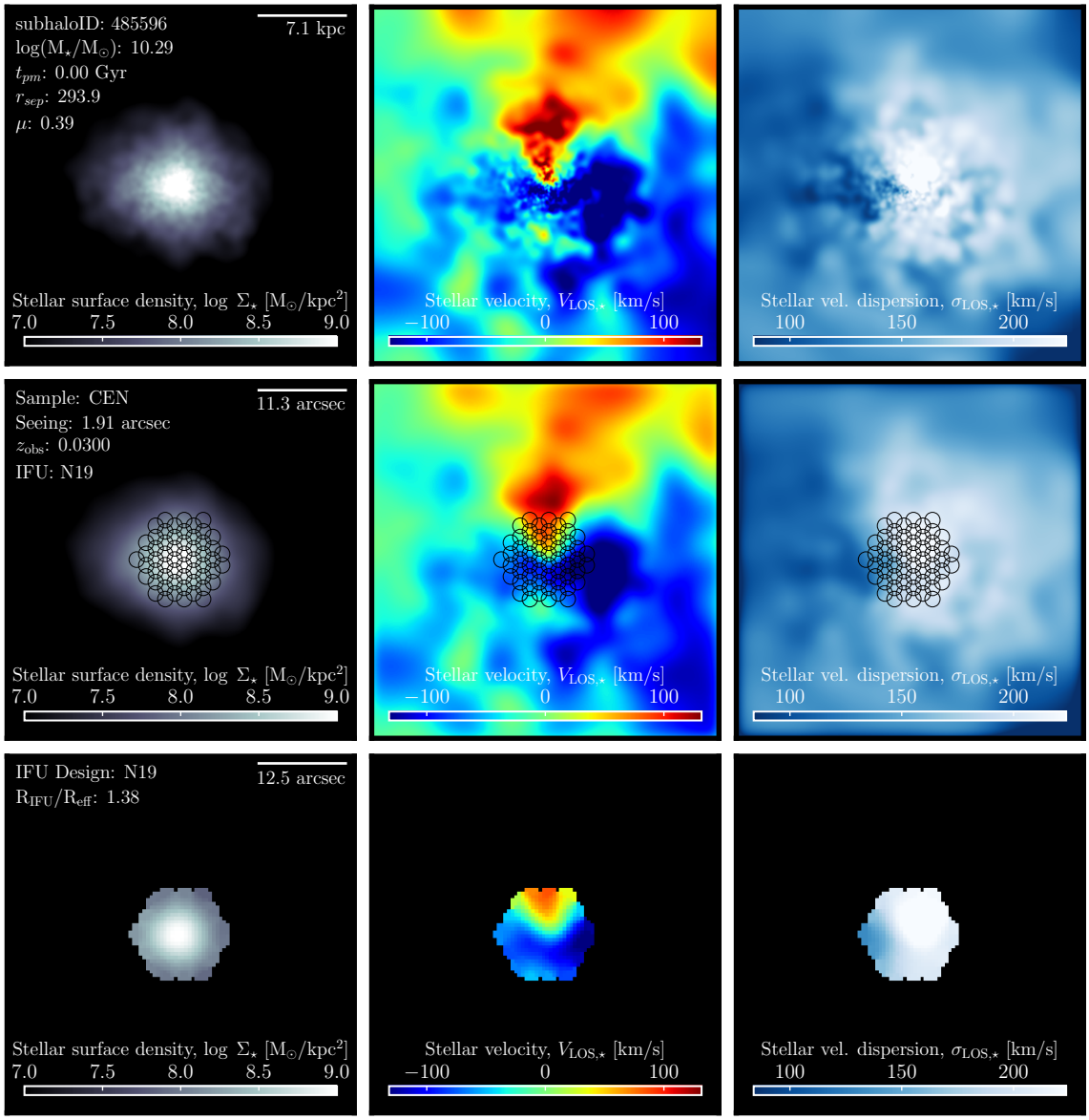

Figure G.51 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 485596, camera: 1 observed for the CEN sample.

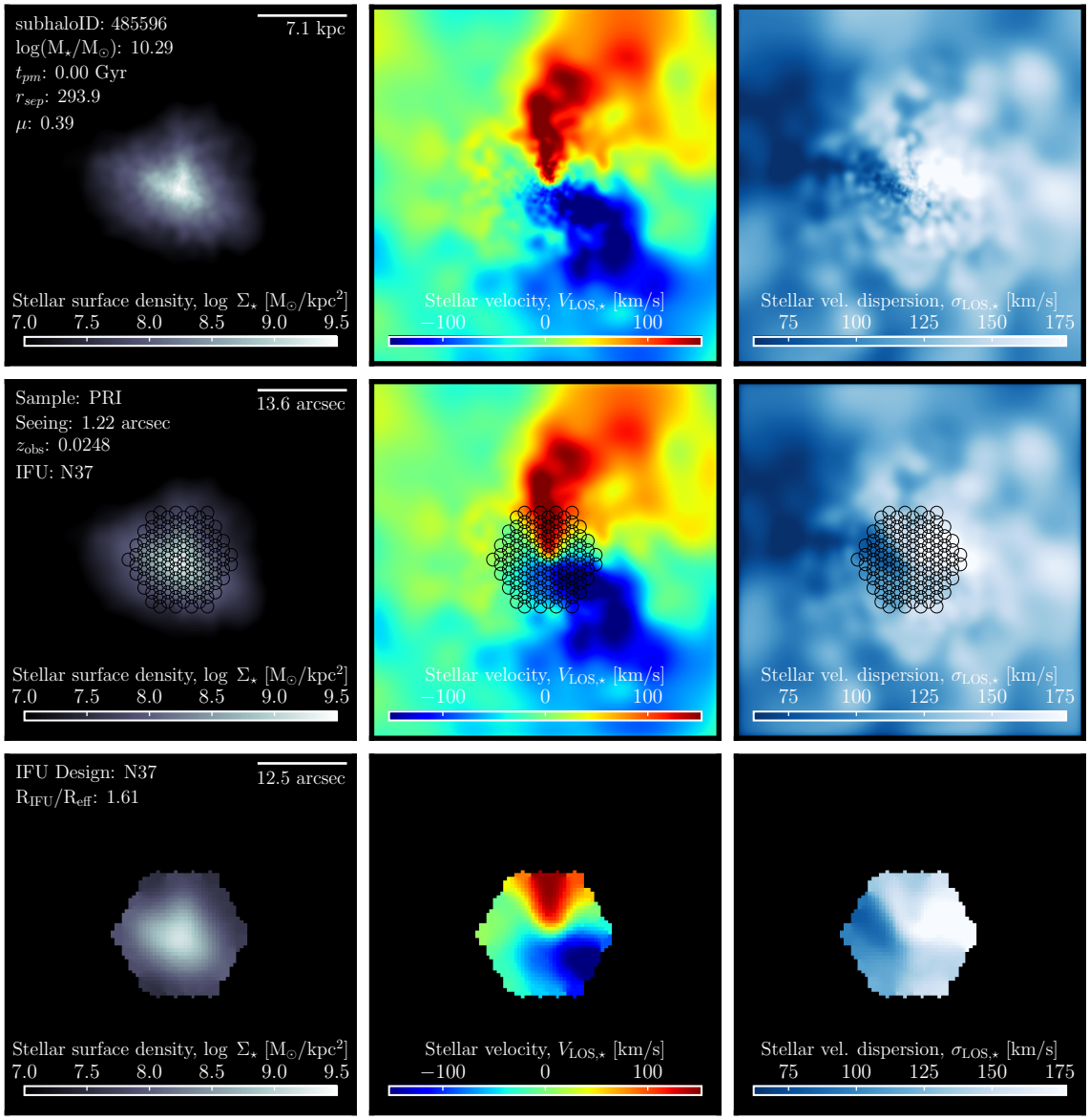

Figure G.52 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 485596, camera: 3 observed for the PRI sample.

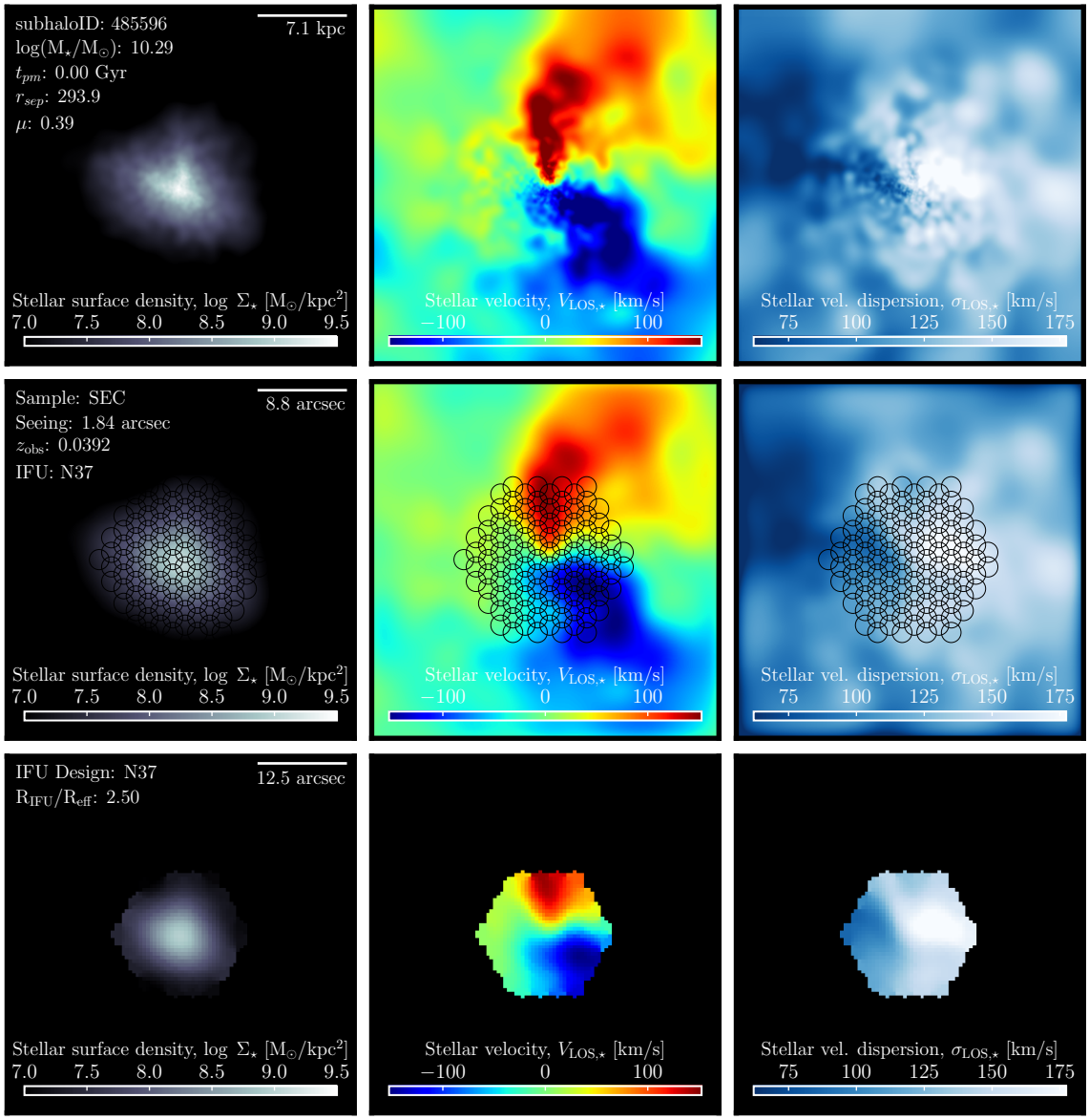

Figure G.53 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 485596, camera: 3 observed for the SEC sample.

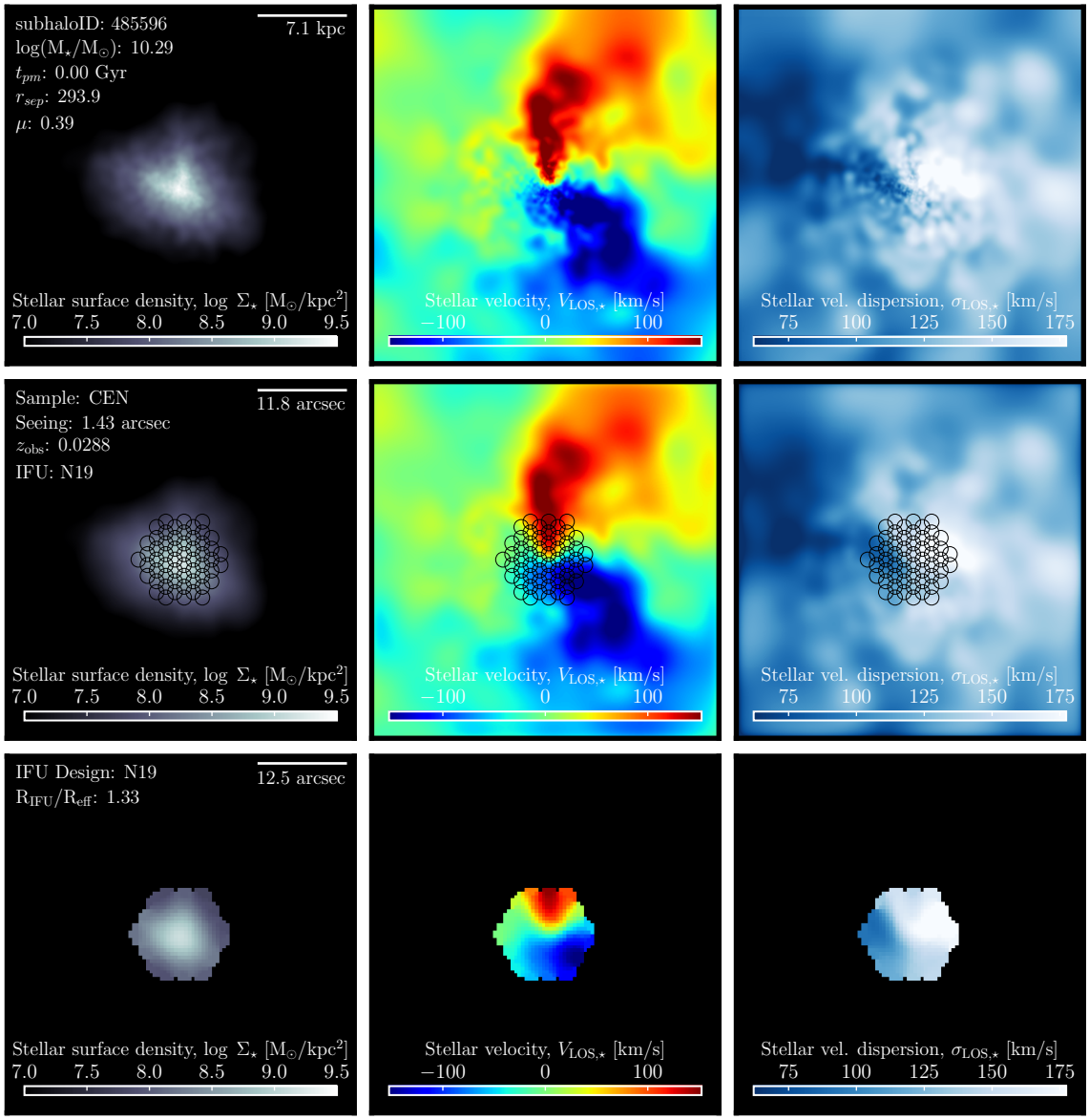

Figure G.54 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 485596, camera: 3 observed for the CEN sample.

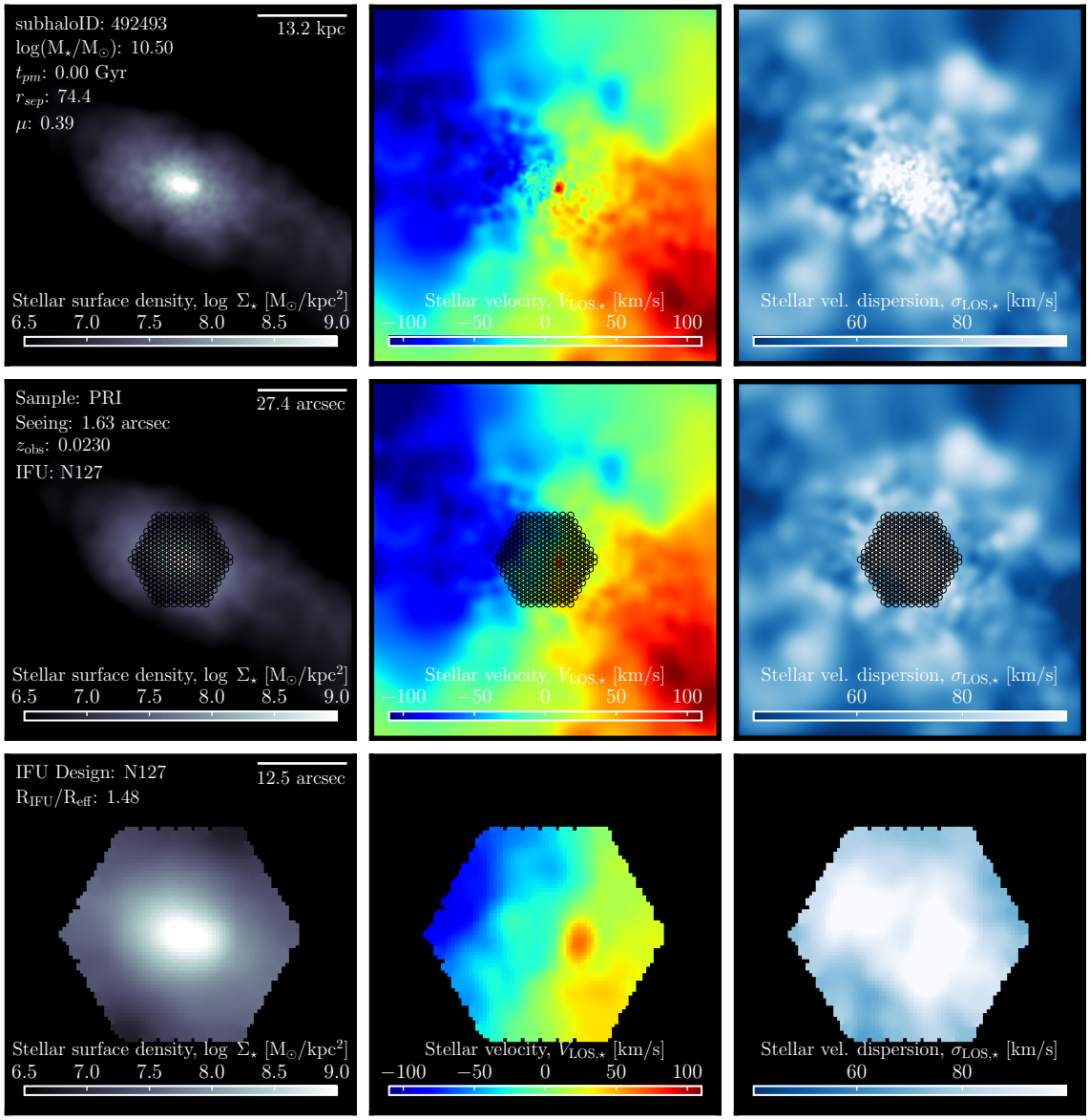

Figure G.55 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 492493, camera: 1 observed for the PRI sample.

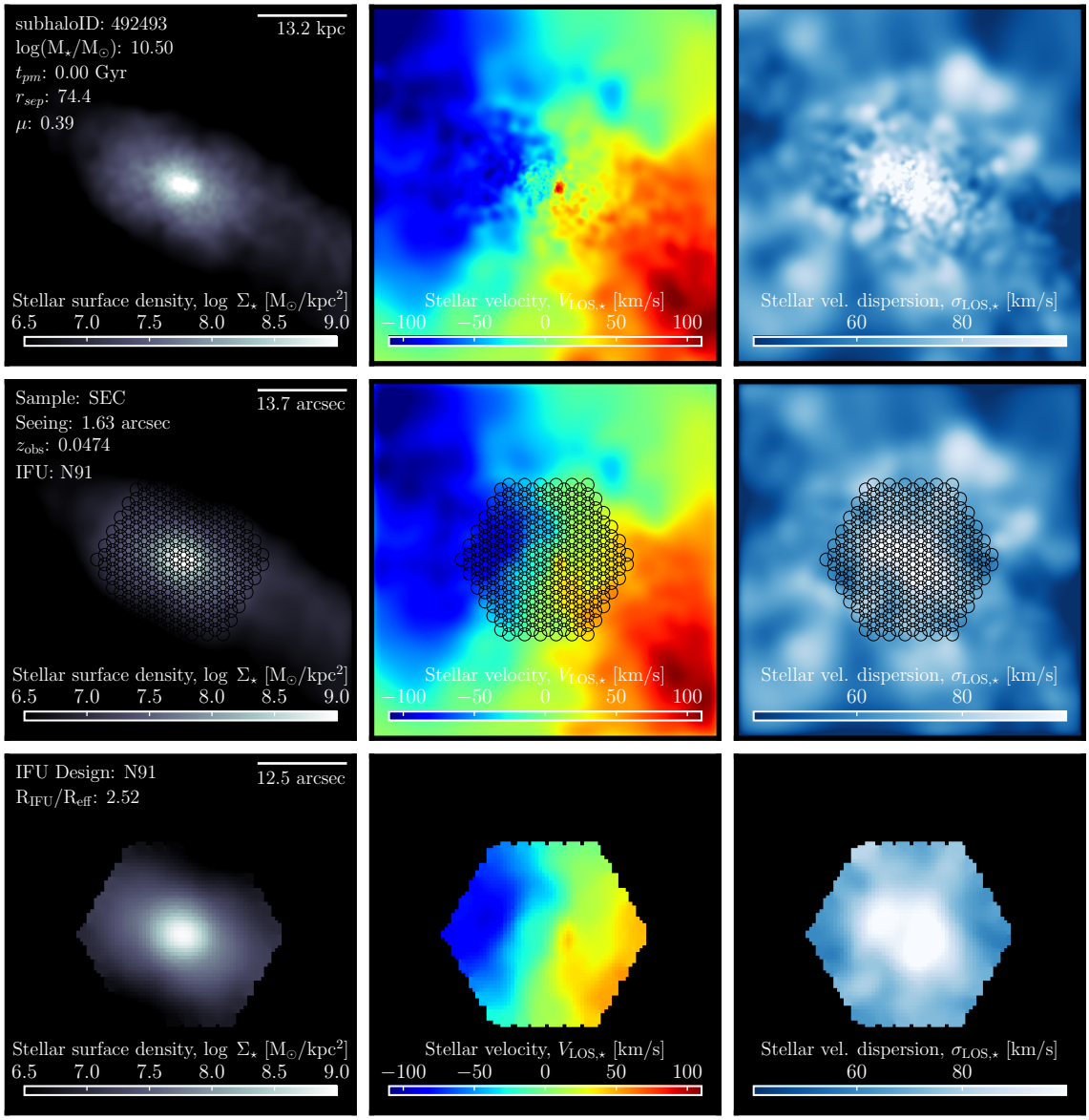

Figure G.56 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 492493, camera: 1 observed for the SEC sample.

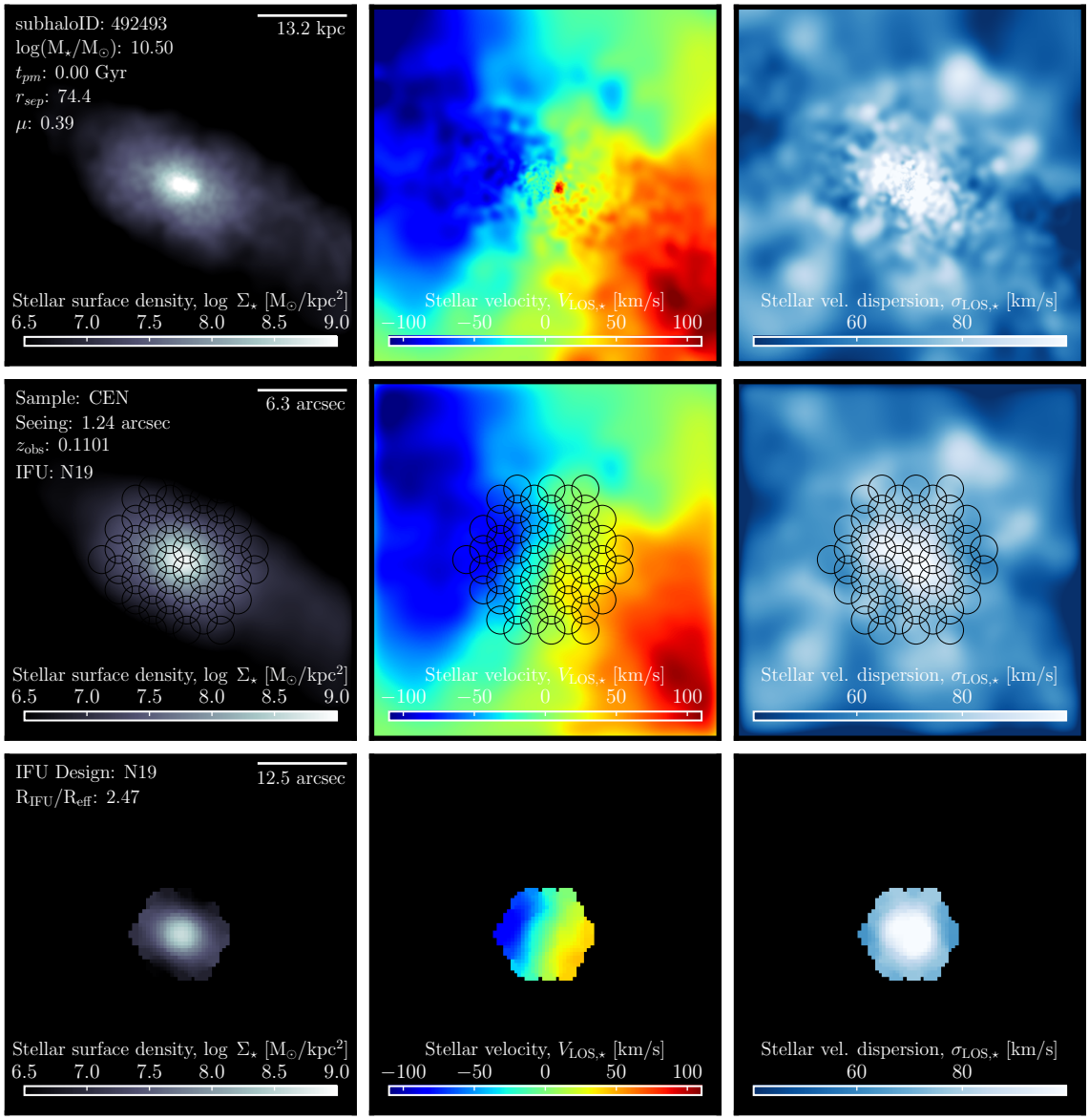

Figure G.57 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 492493, camera: 1 observed for the CEN sample.

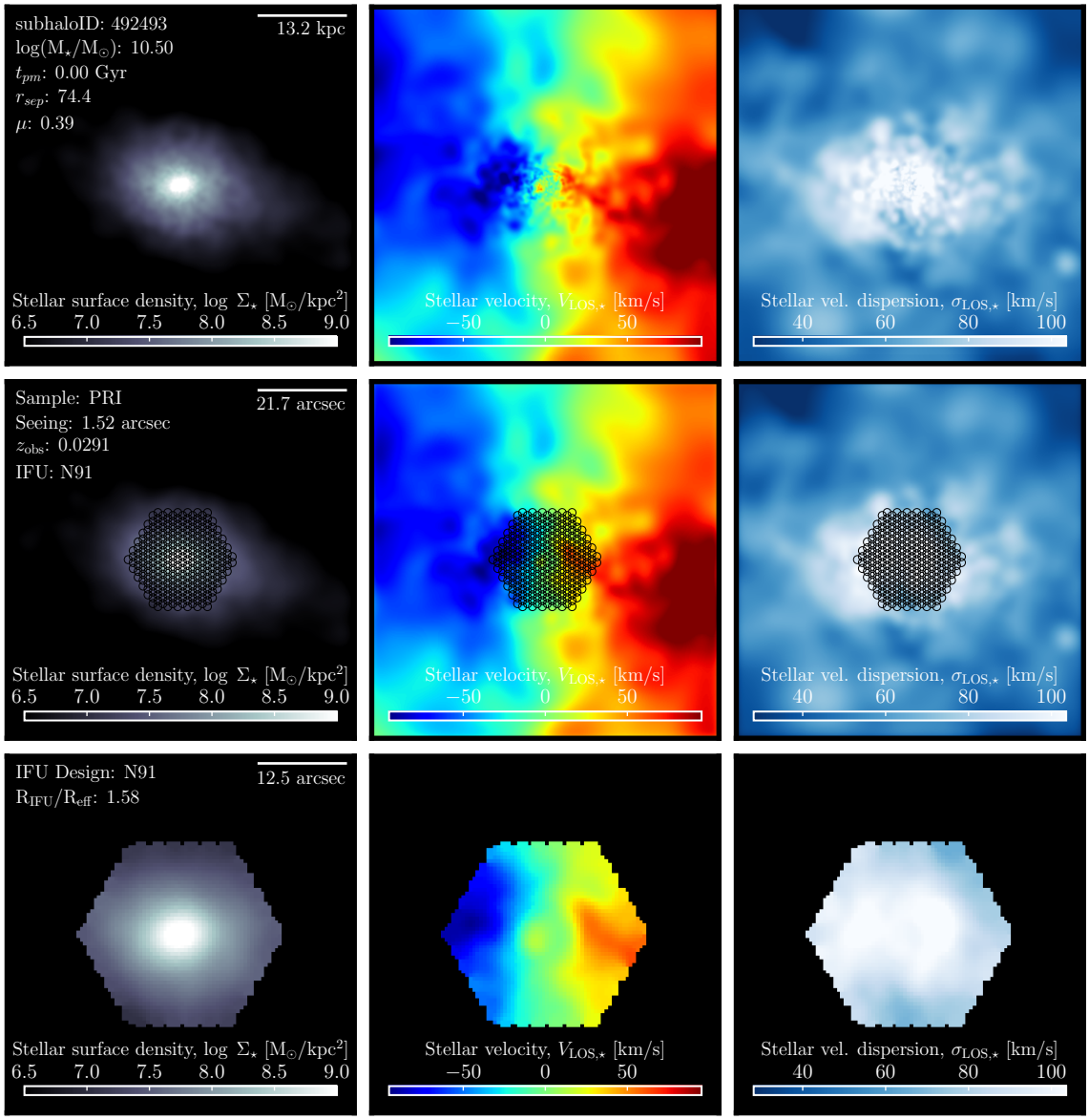

Figure G.58 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 492493, camera: 3 observed for the PRI sample.

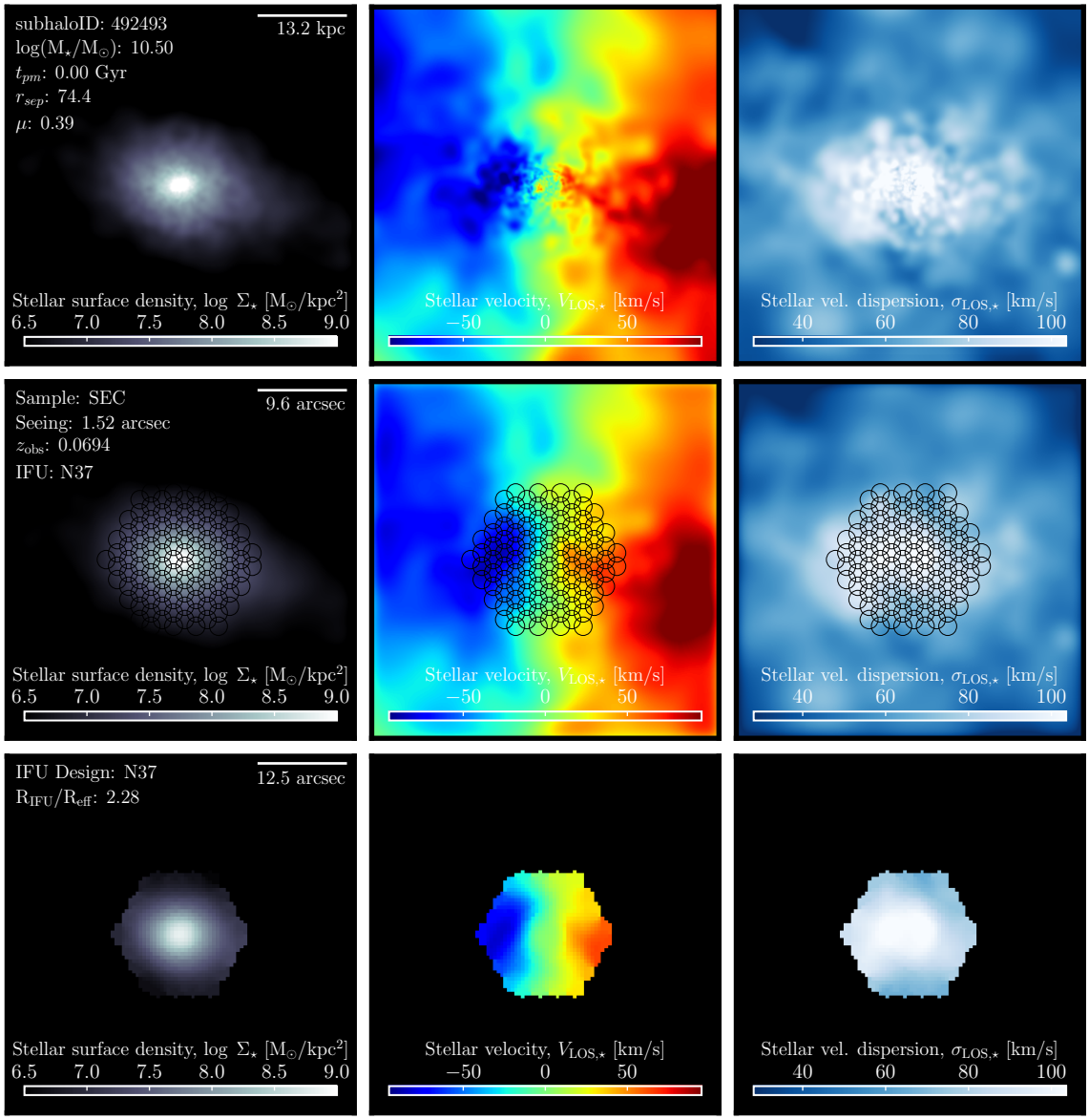

Figure G.59 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 492493, camera: 3 observed for the SEC sample.

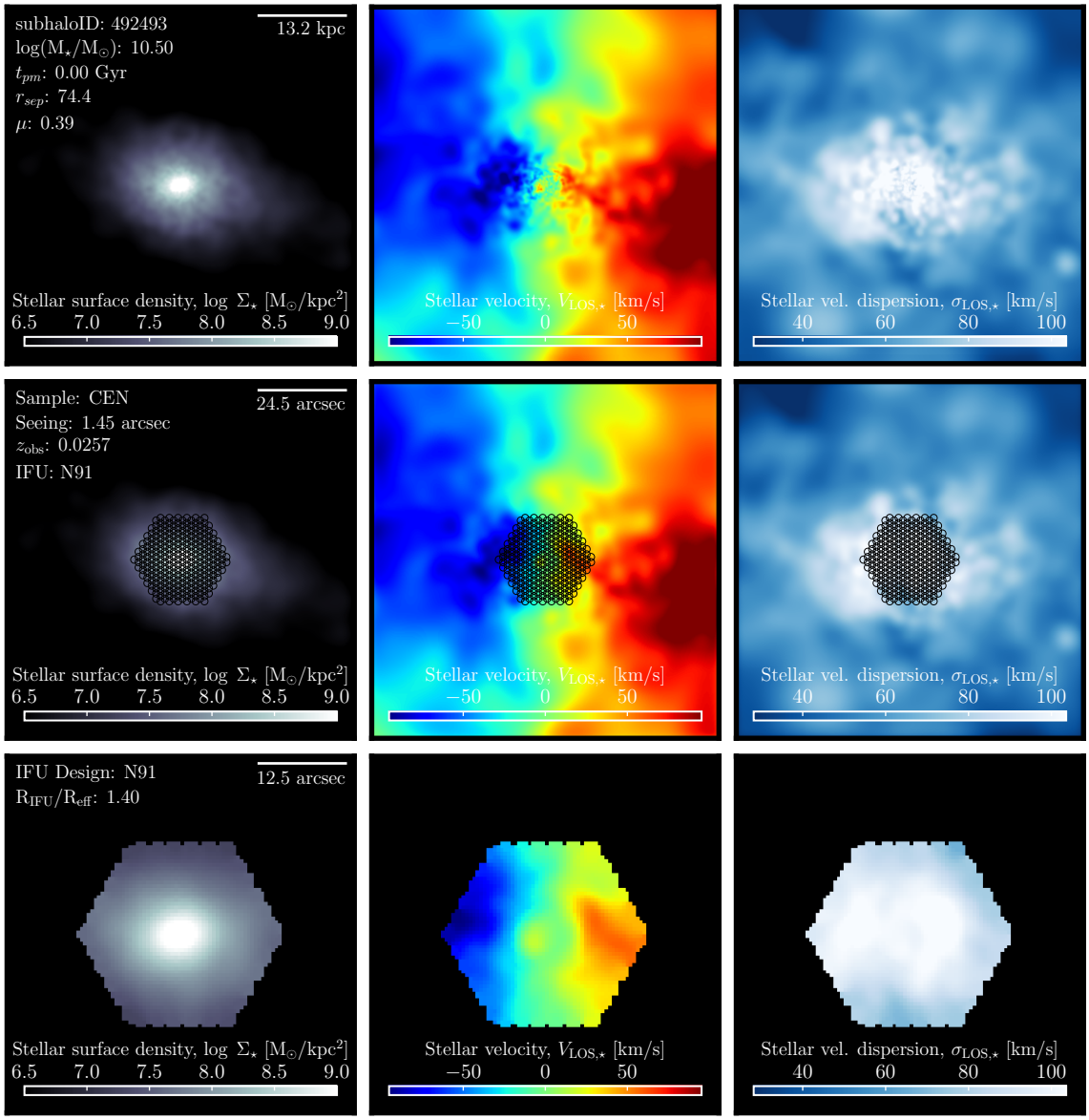

Figure G.60 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 492493, camera: 3 observed for the CEN sample.

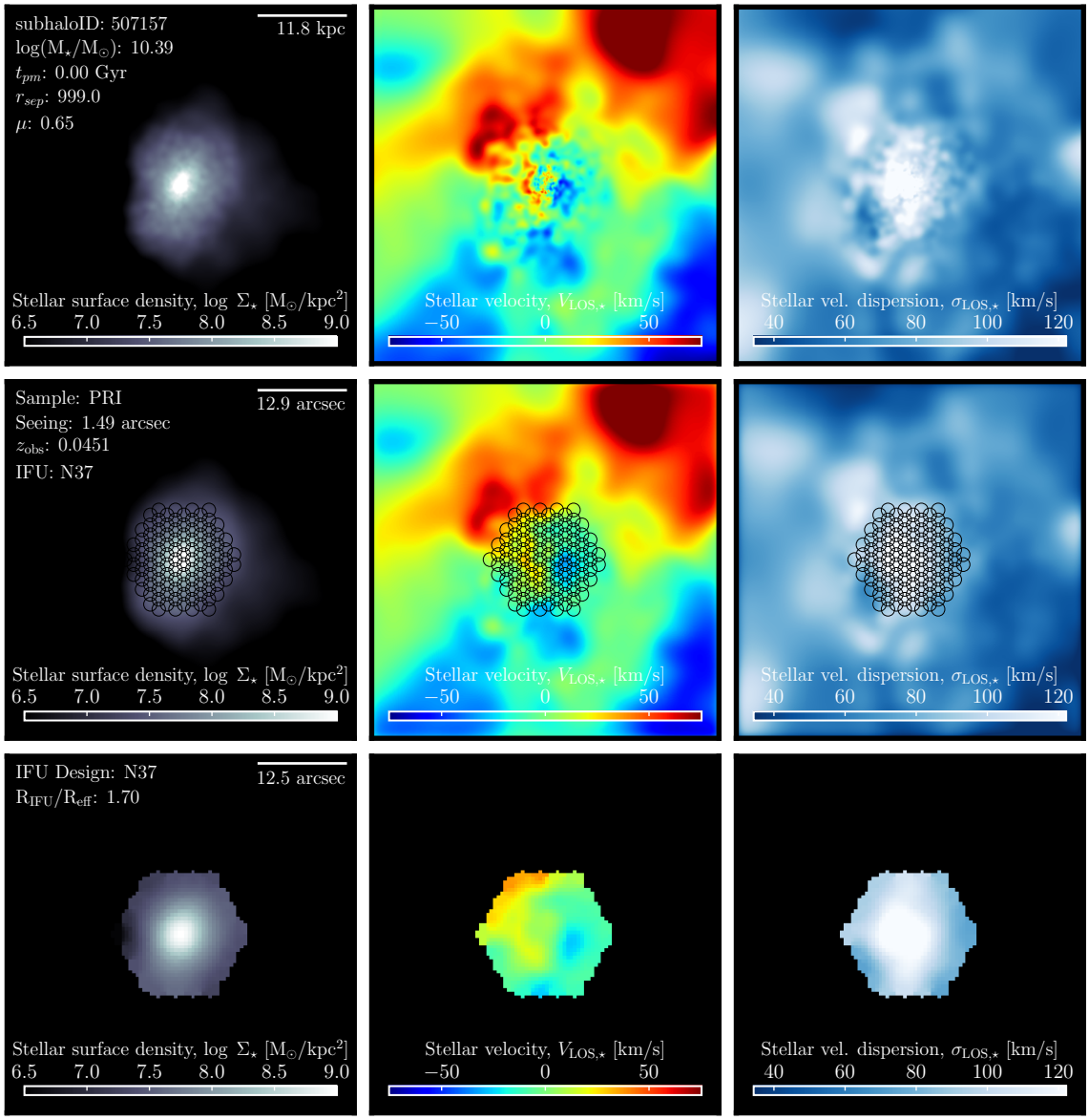

Figure G.61 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 507157, camera: 1 observed for the PRI sample.

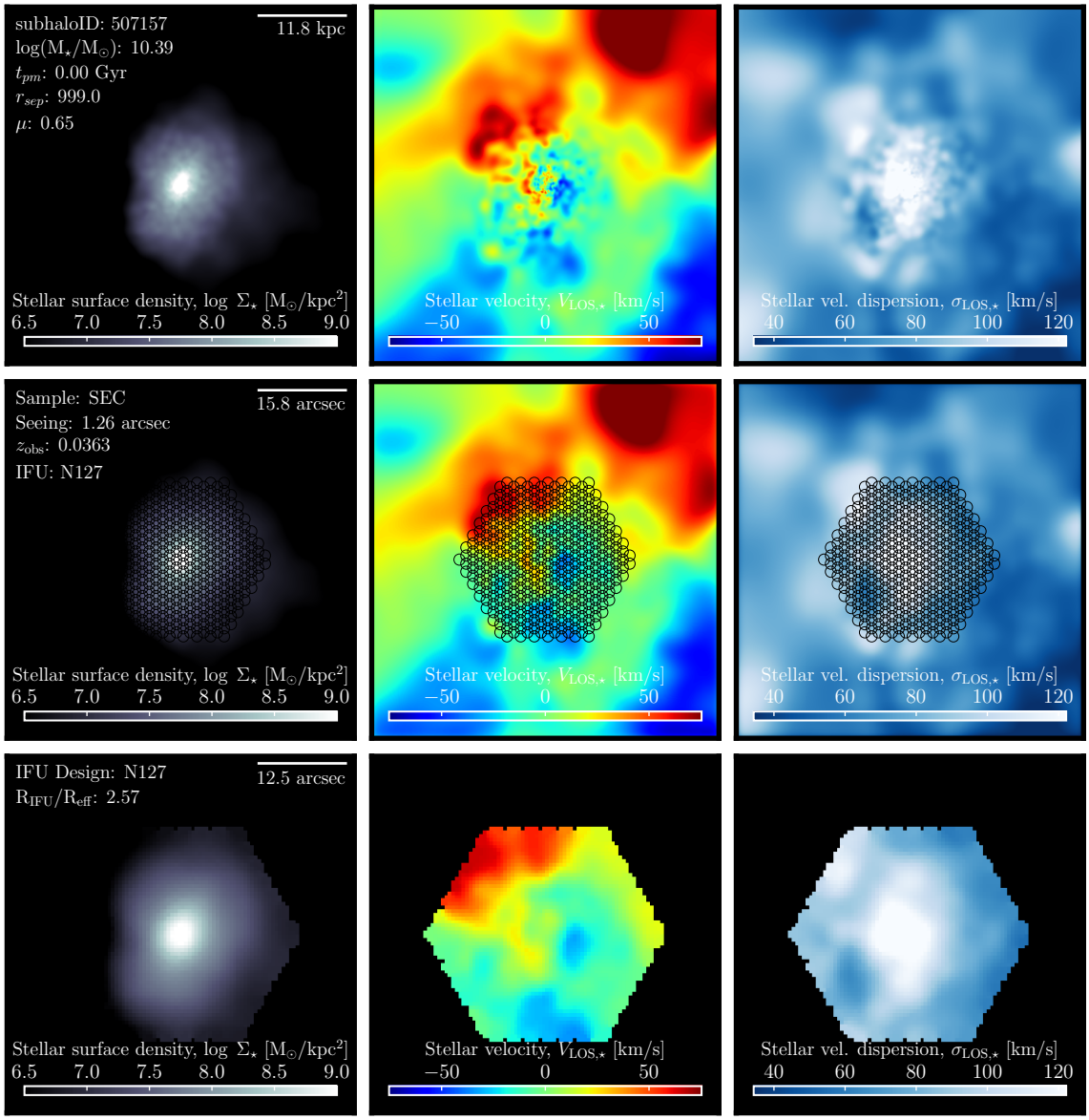

Figure G.62 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 507157, camera: 1 observed for the SEC sample.

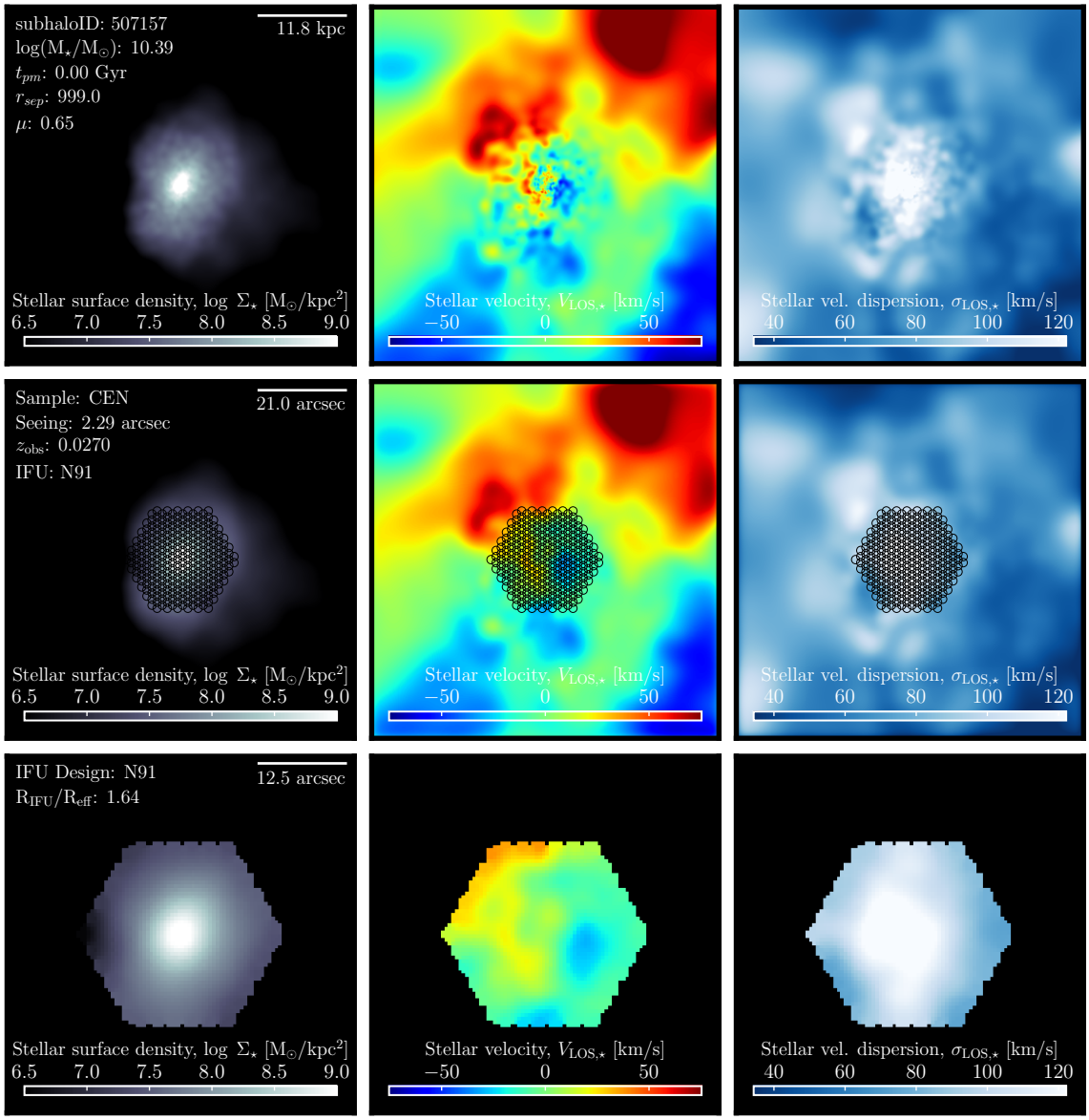

Figure G.63 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 507157, camera: 1 observed for the CEN sample.

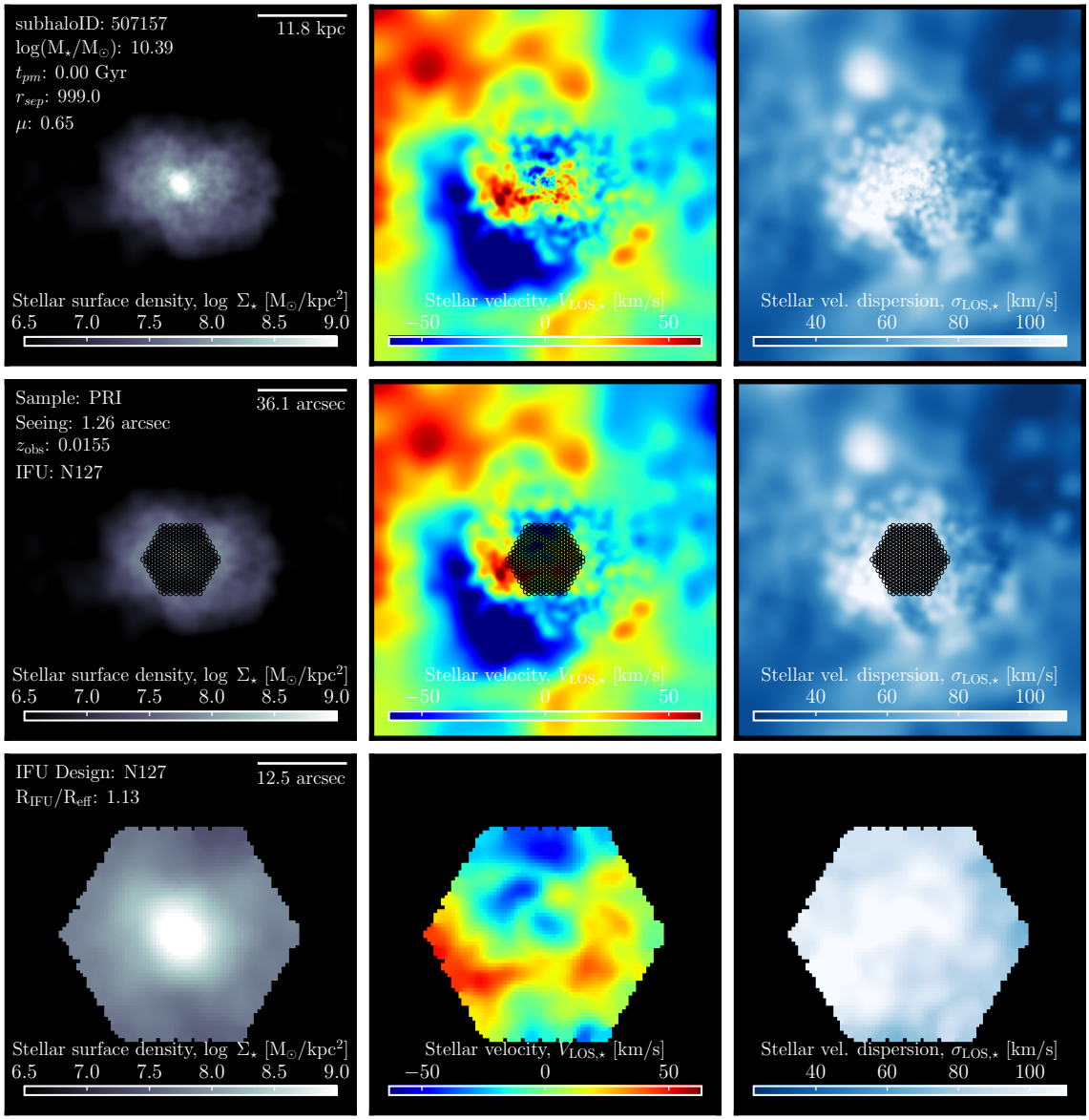

Figure G.64 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 507157, camera: 3 observed for the PRI sample.

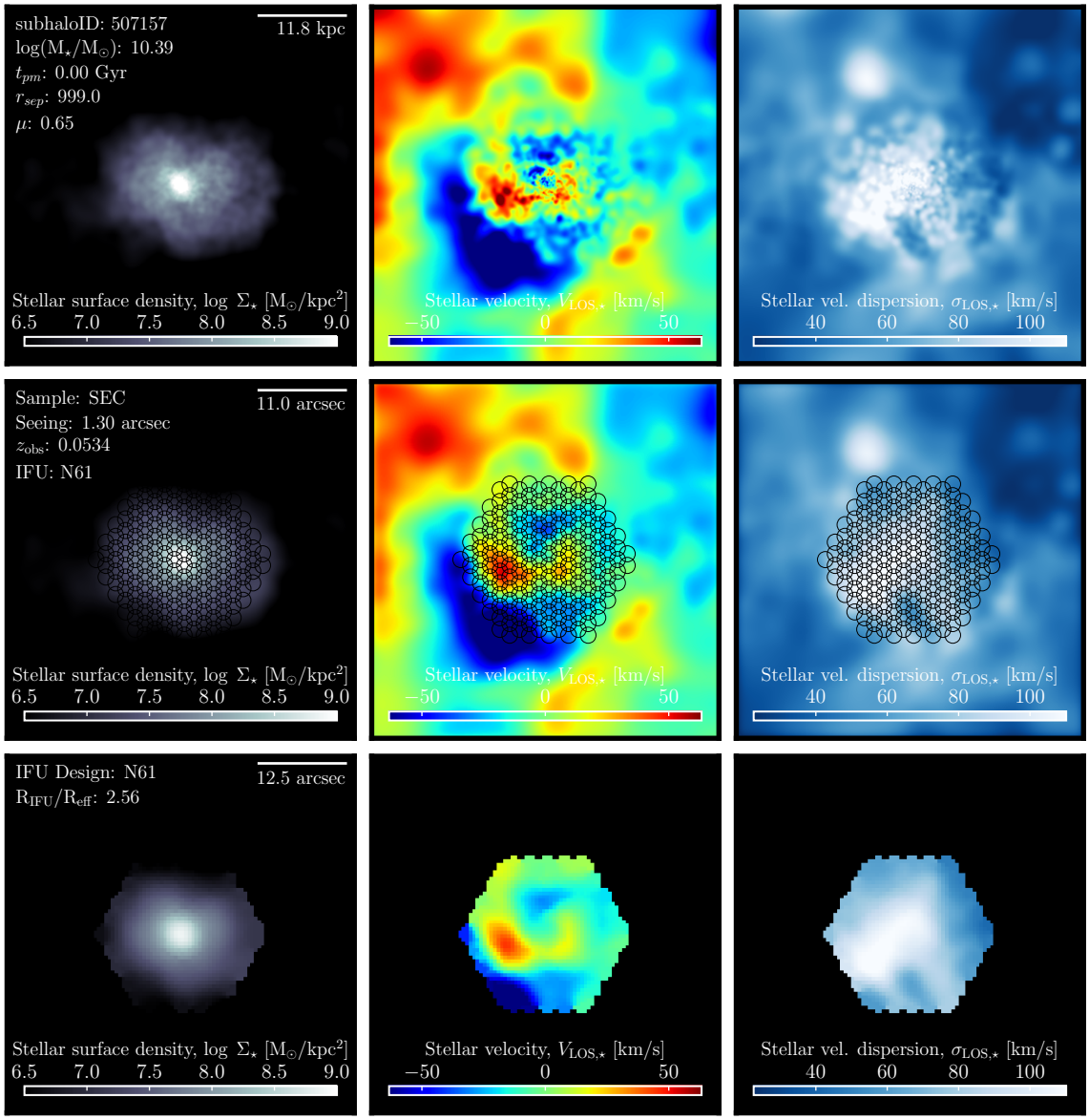

Figure G.65 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 507157, camera: 3 observed for the SEC sample.

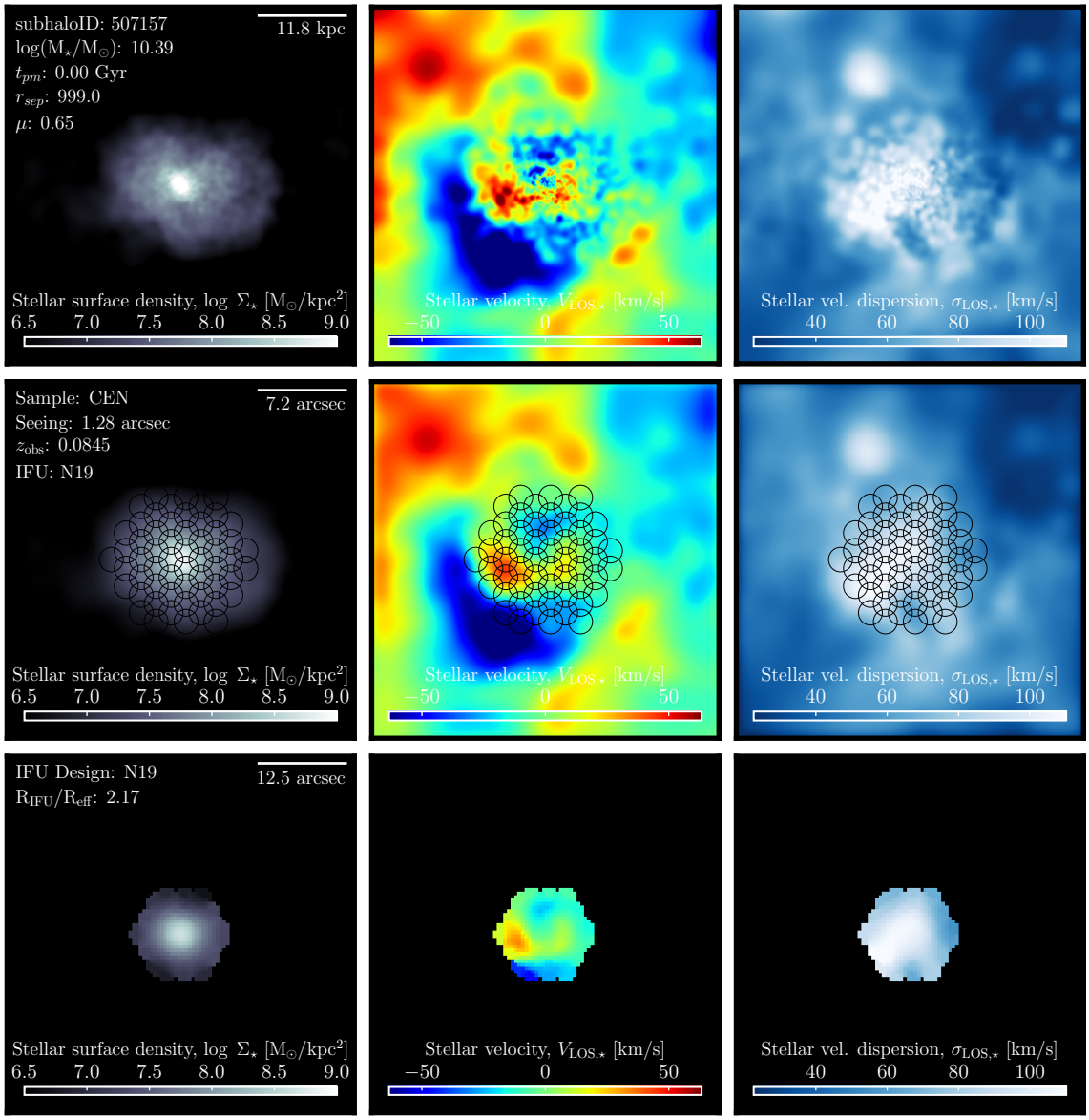

Figure G.66 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 507157, camera: 3 observed for the CEN sample.

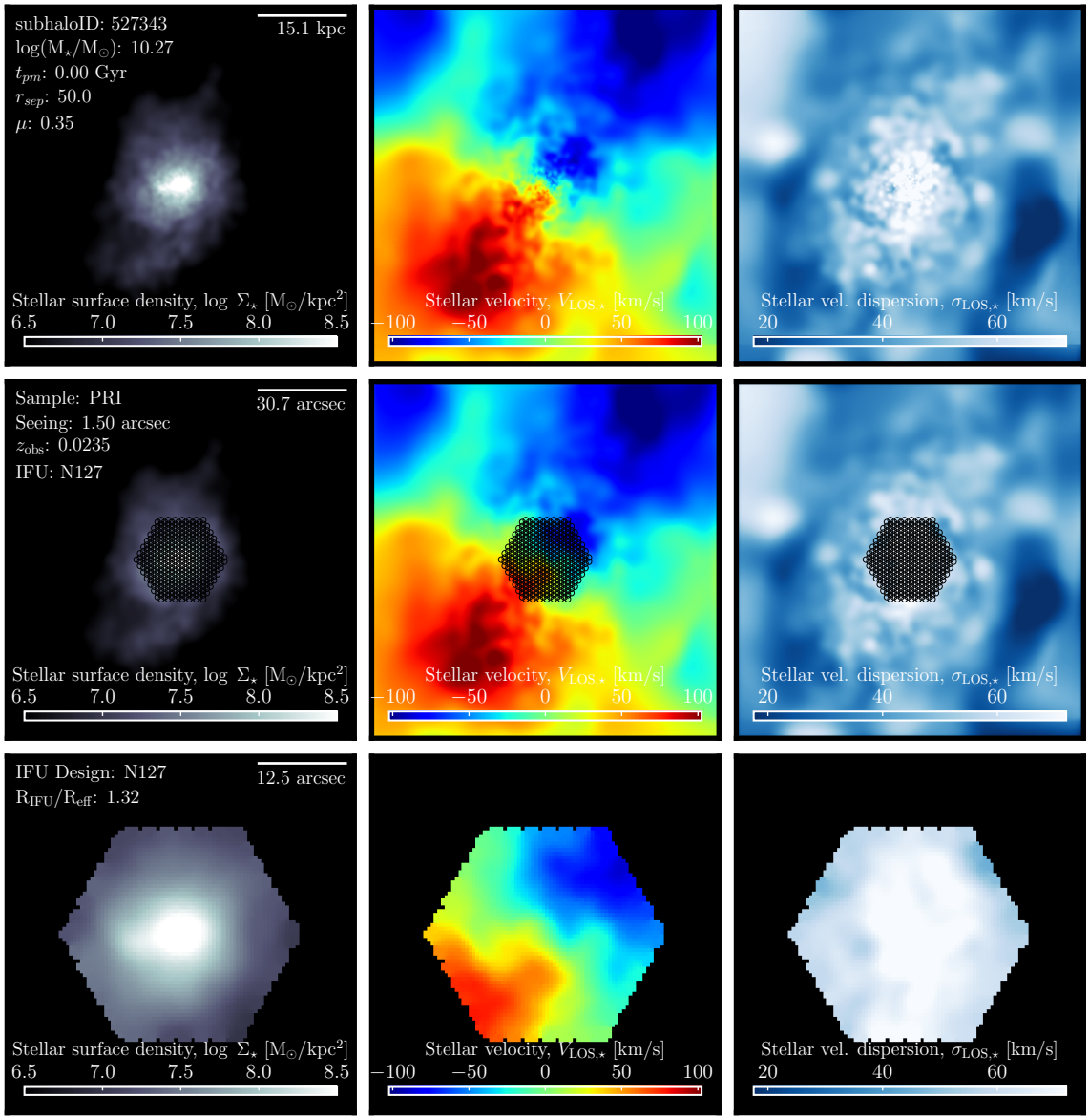

Figure G.67 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 527343, camera: 1 observed for the PRI sample.

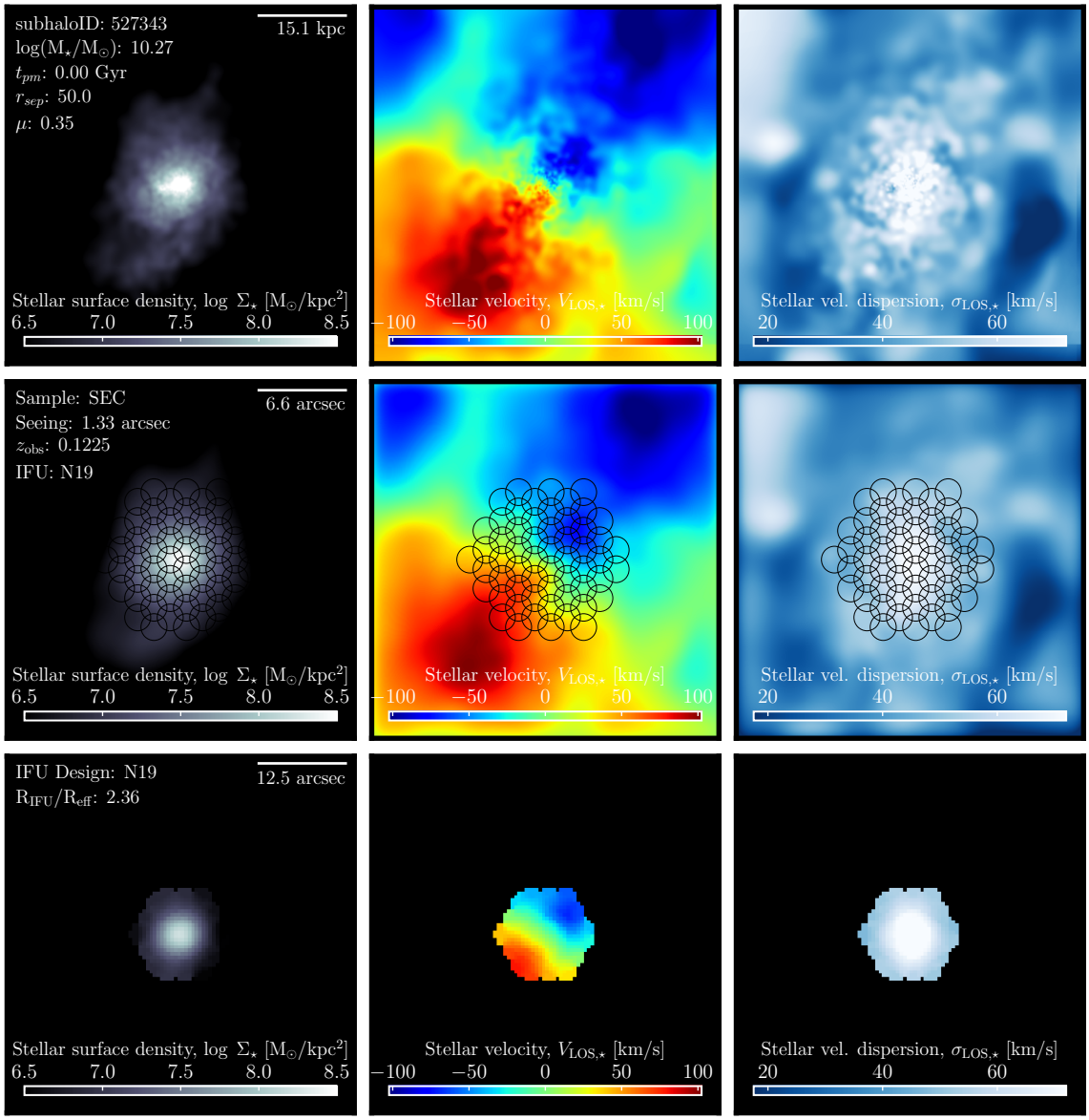

Figure G.68 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 527343, camera: 1 observed for the SEC sample.

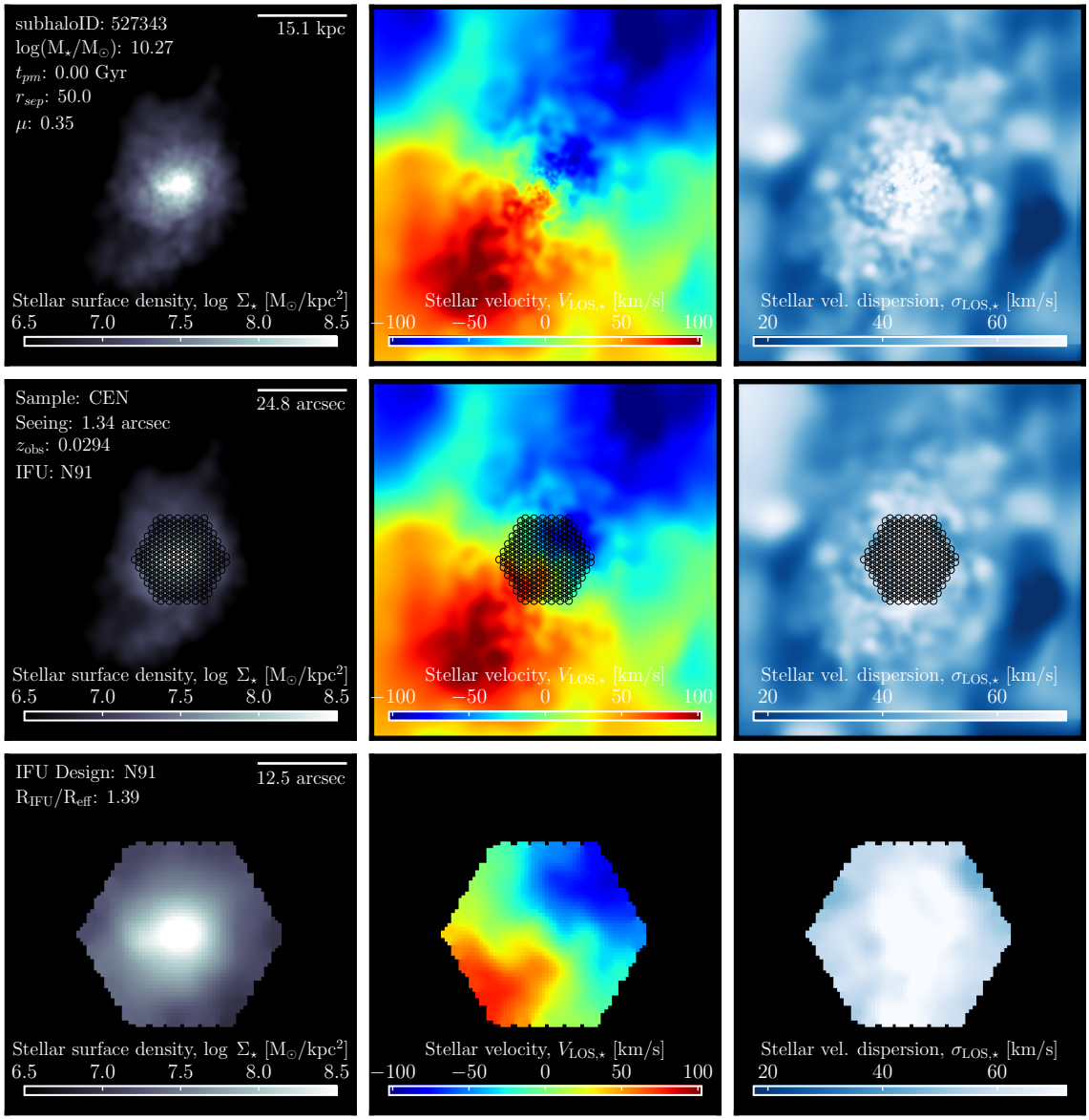

Figure G.69 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 527343, camera: 1 observed for the CEN sample.

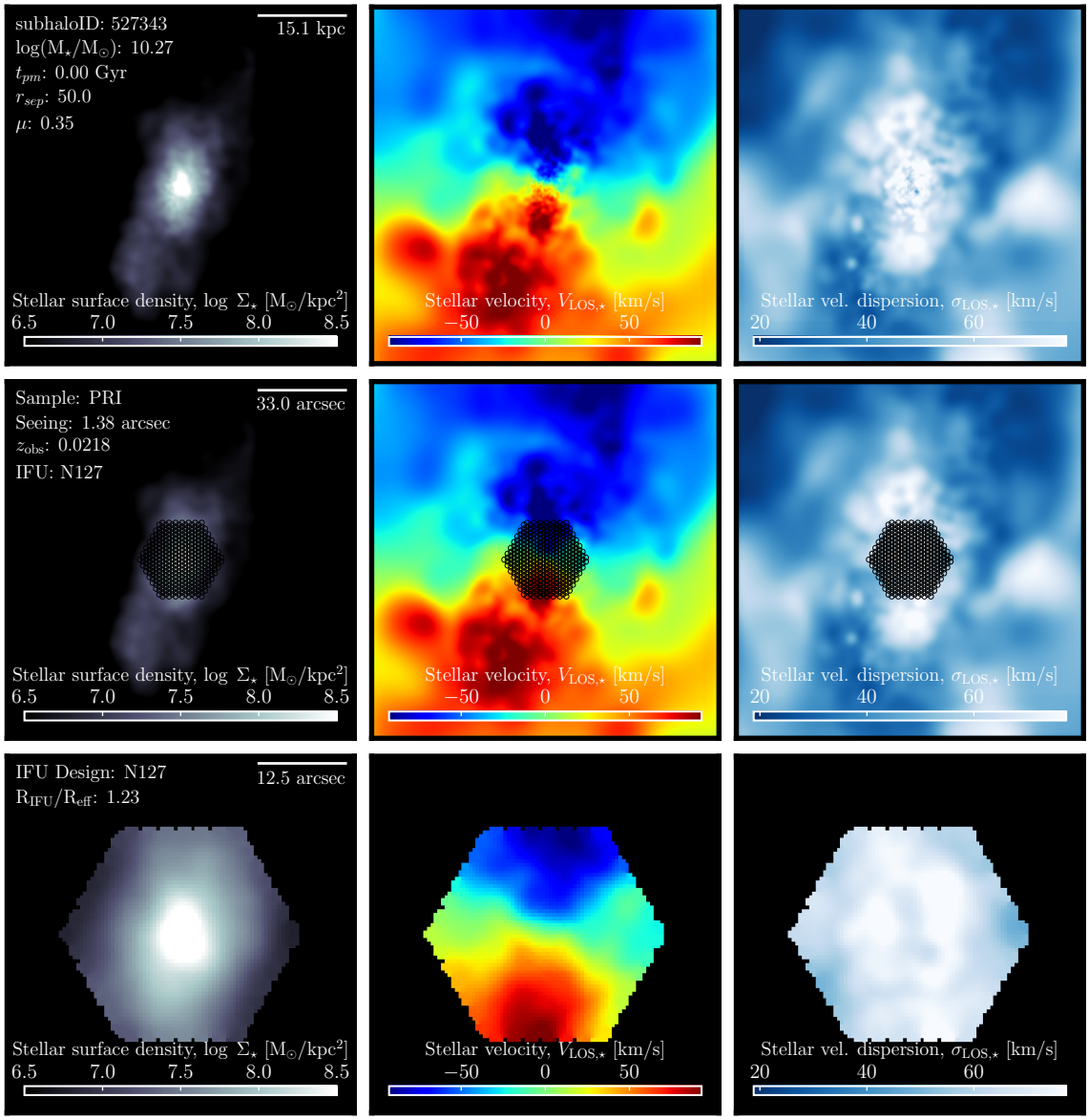

Figure G.70 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 527343, camera: 3 observed for the PRI sample.

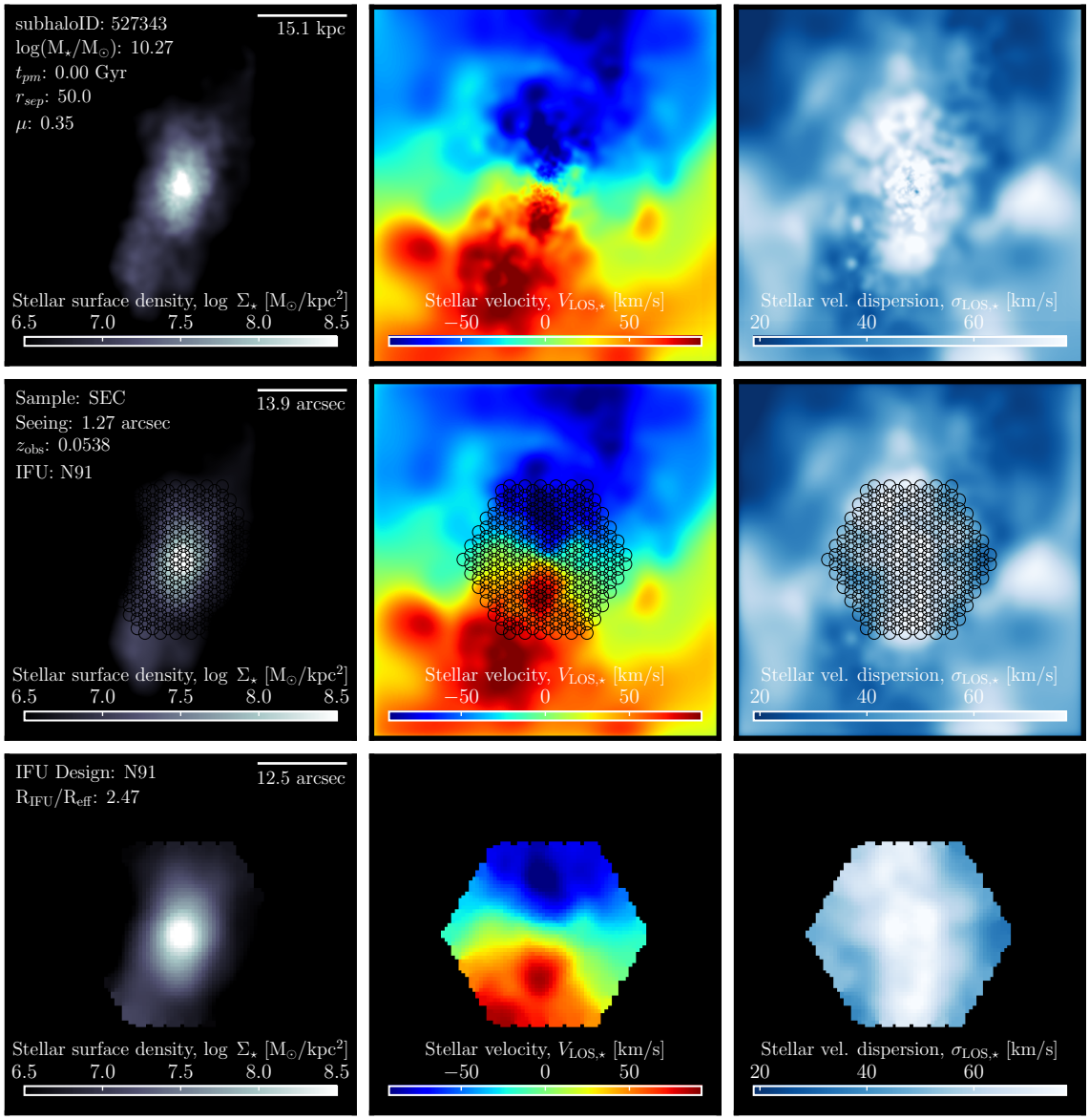

Figure G.71 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 527343, camera: 3 observed for the SEC sample.

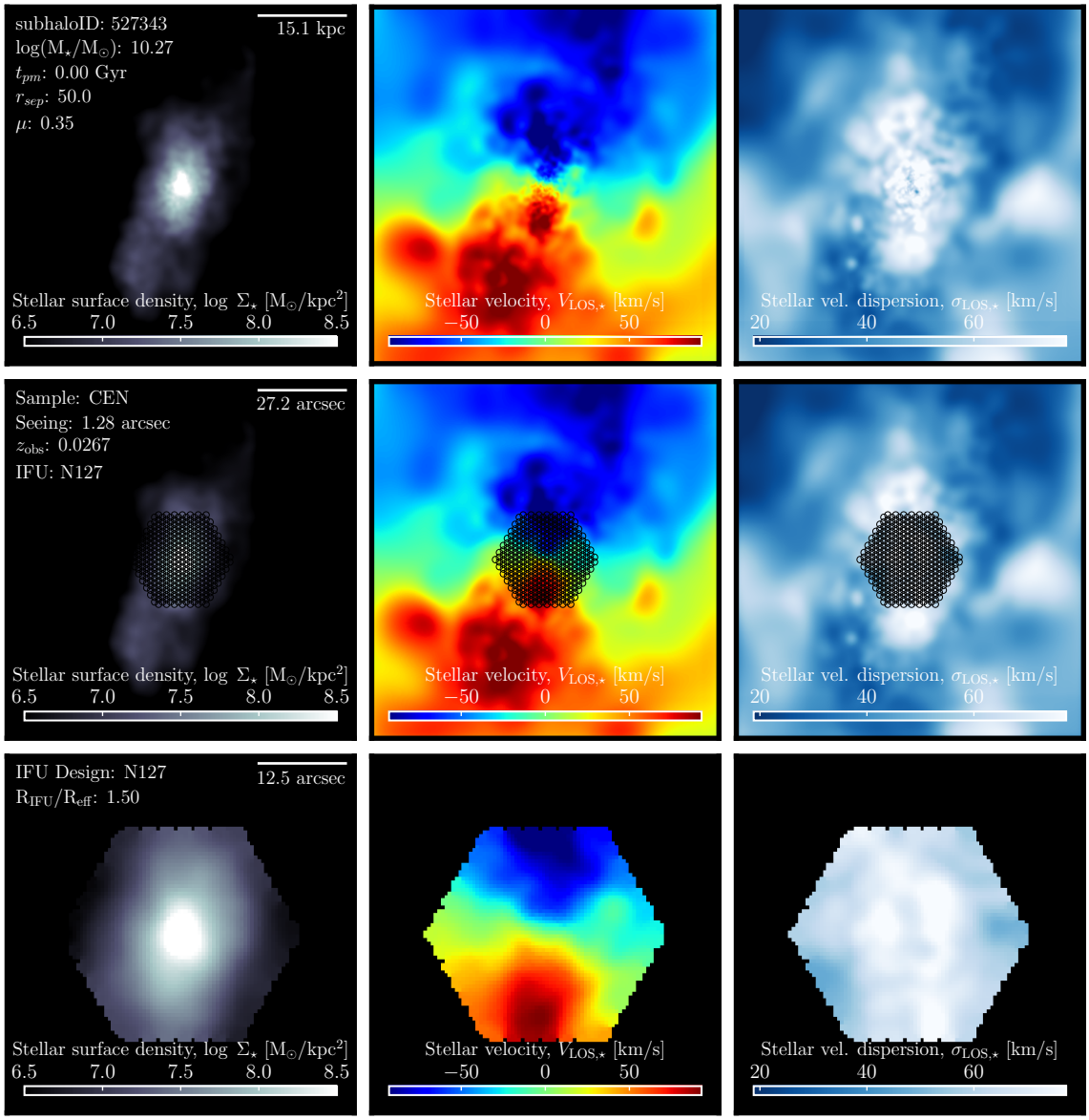

Figure G.72 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 527343, camera: 3 observed for the CEN sample.

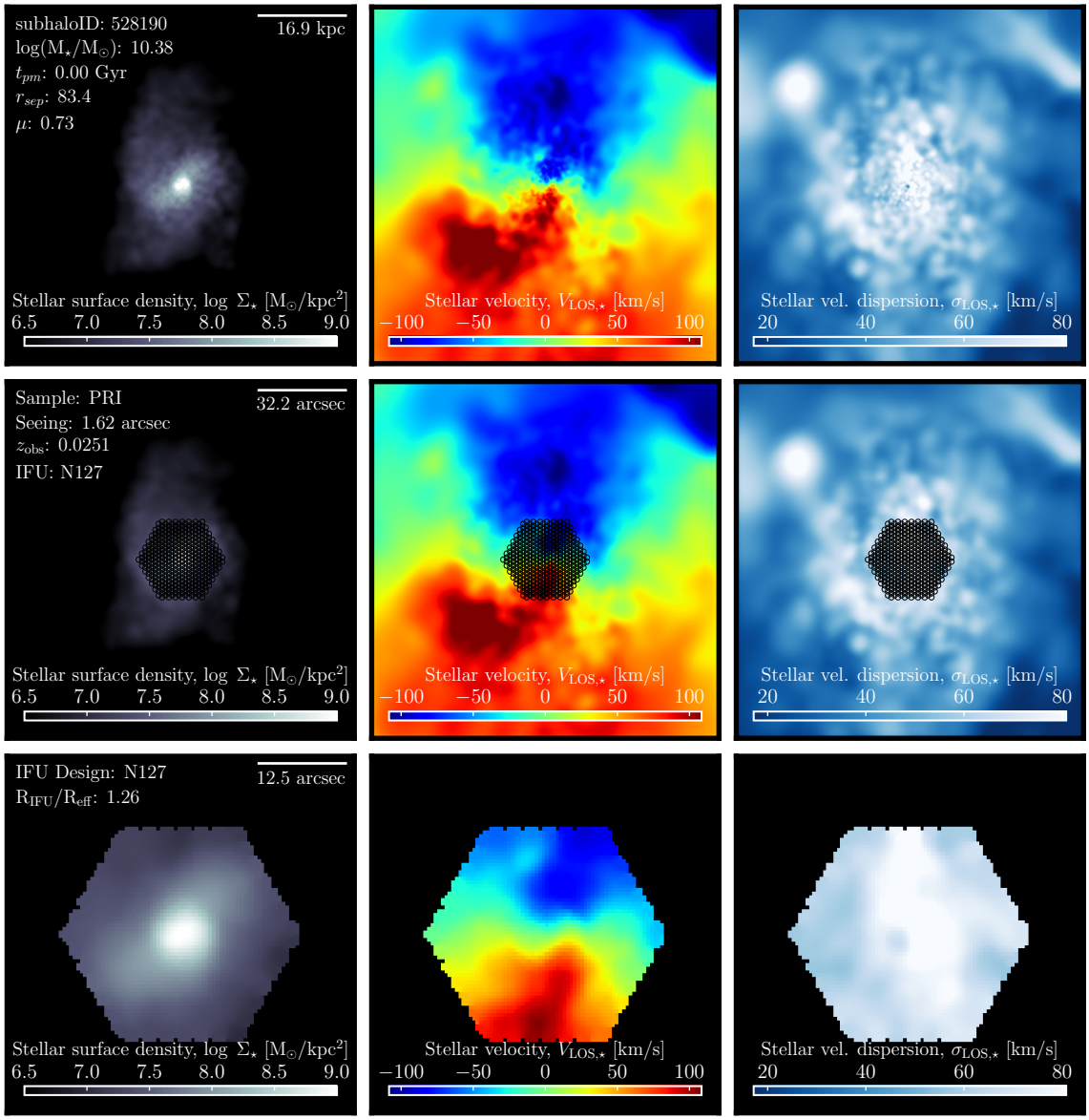

Figure G.73 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 528190, camera: 1 observed for the PRI sample.

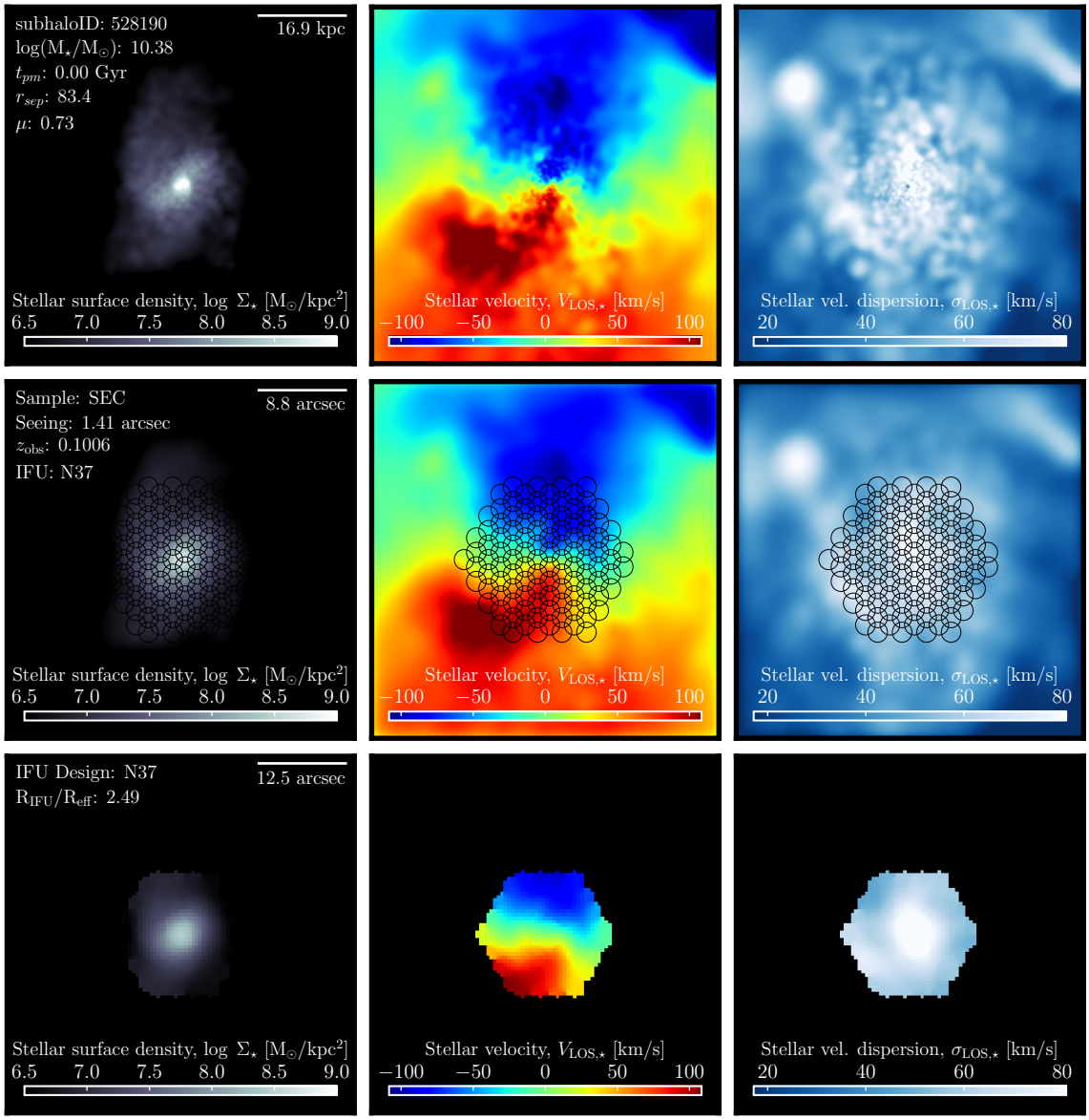

Figure G.74 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 528190, camera: 1 observed for the SEC sample.

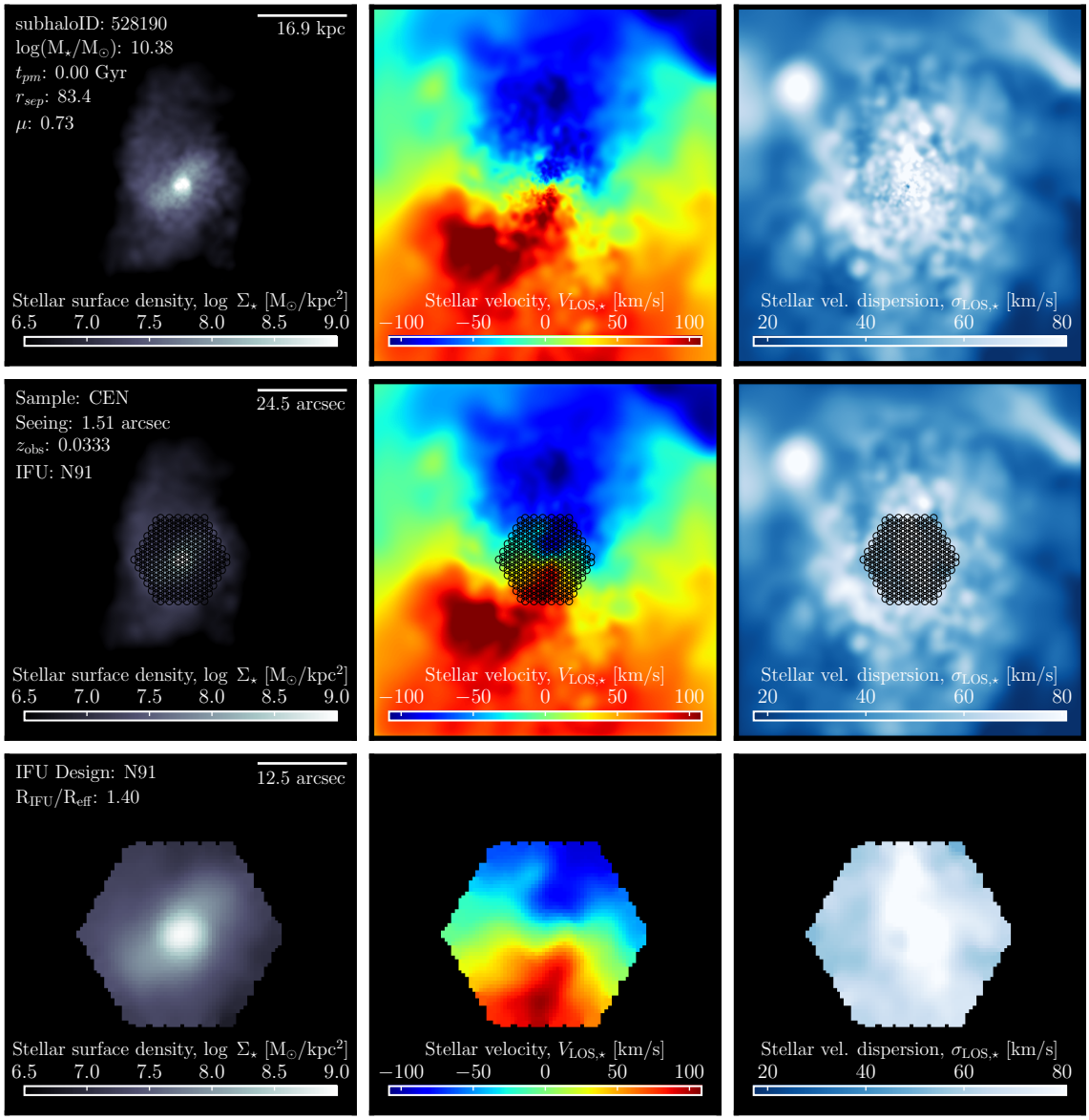

Figure G.75 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 528190, camera: 1 observed for the CEN sample.

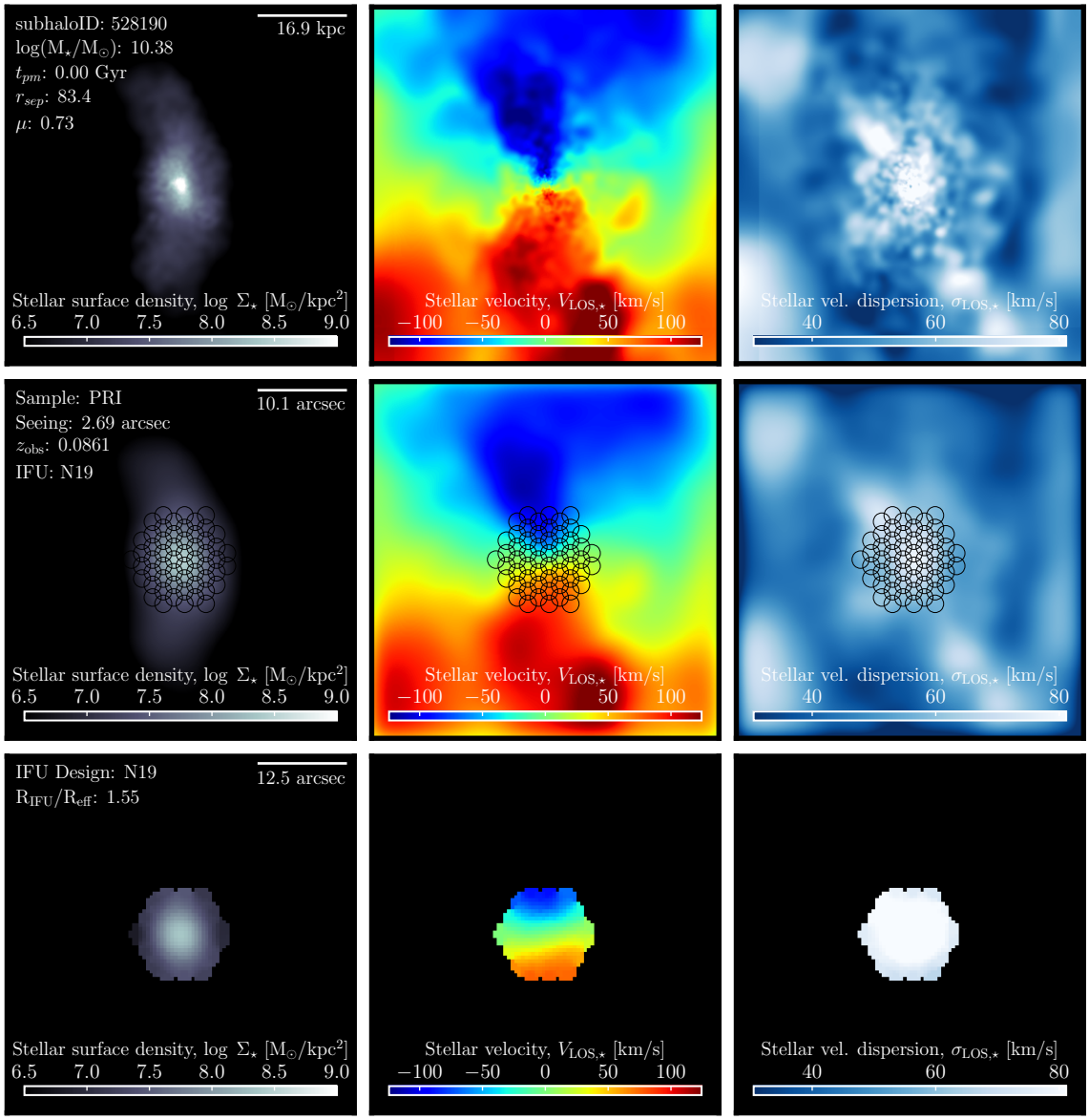

Figure G.76 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 528190, camera: 3 observed for the PRI sample.

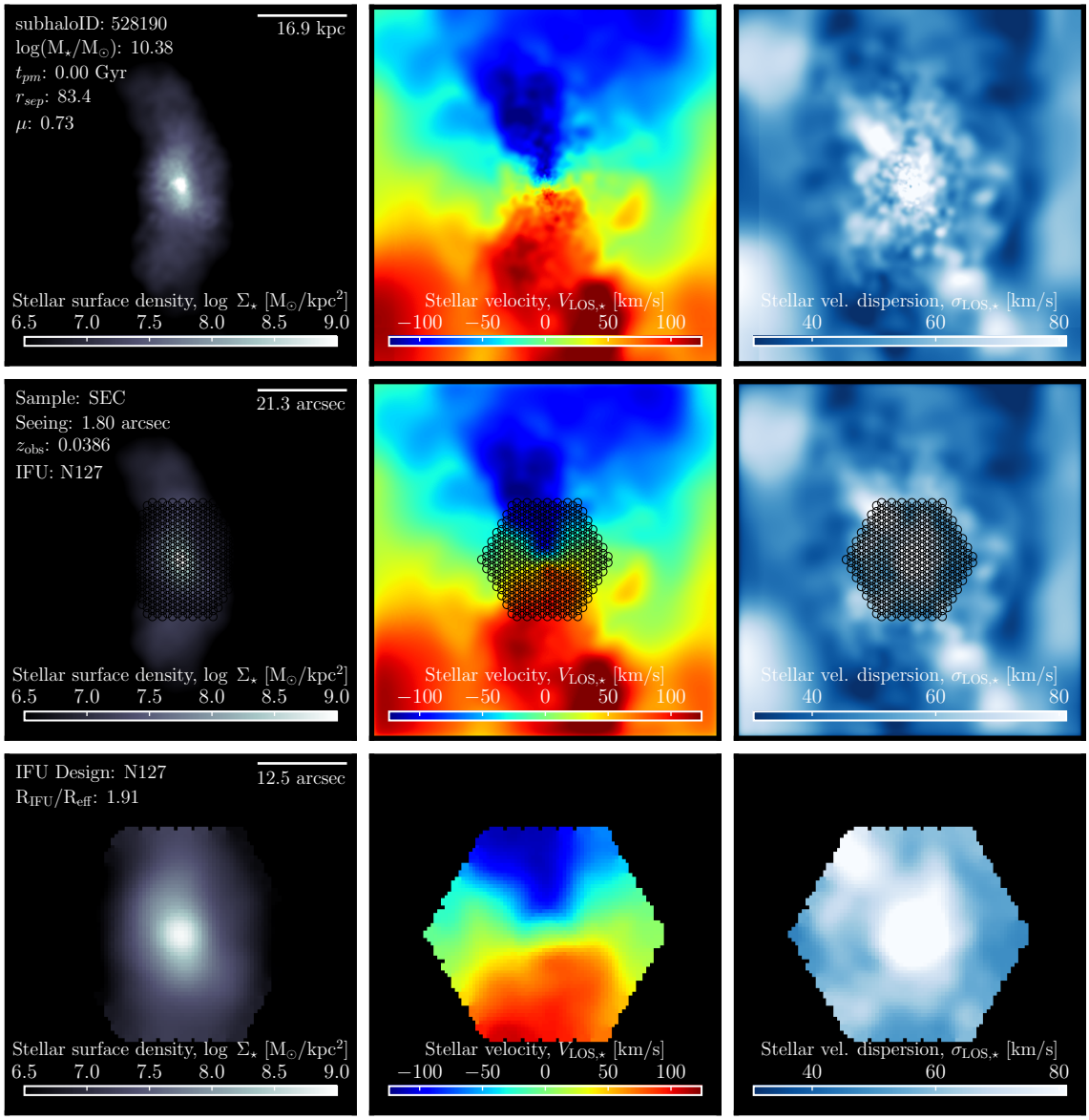

Figure G.77 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 528190, camera: 3 observed for the SEC sample.

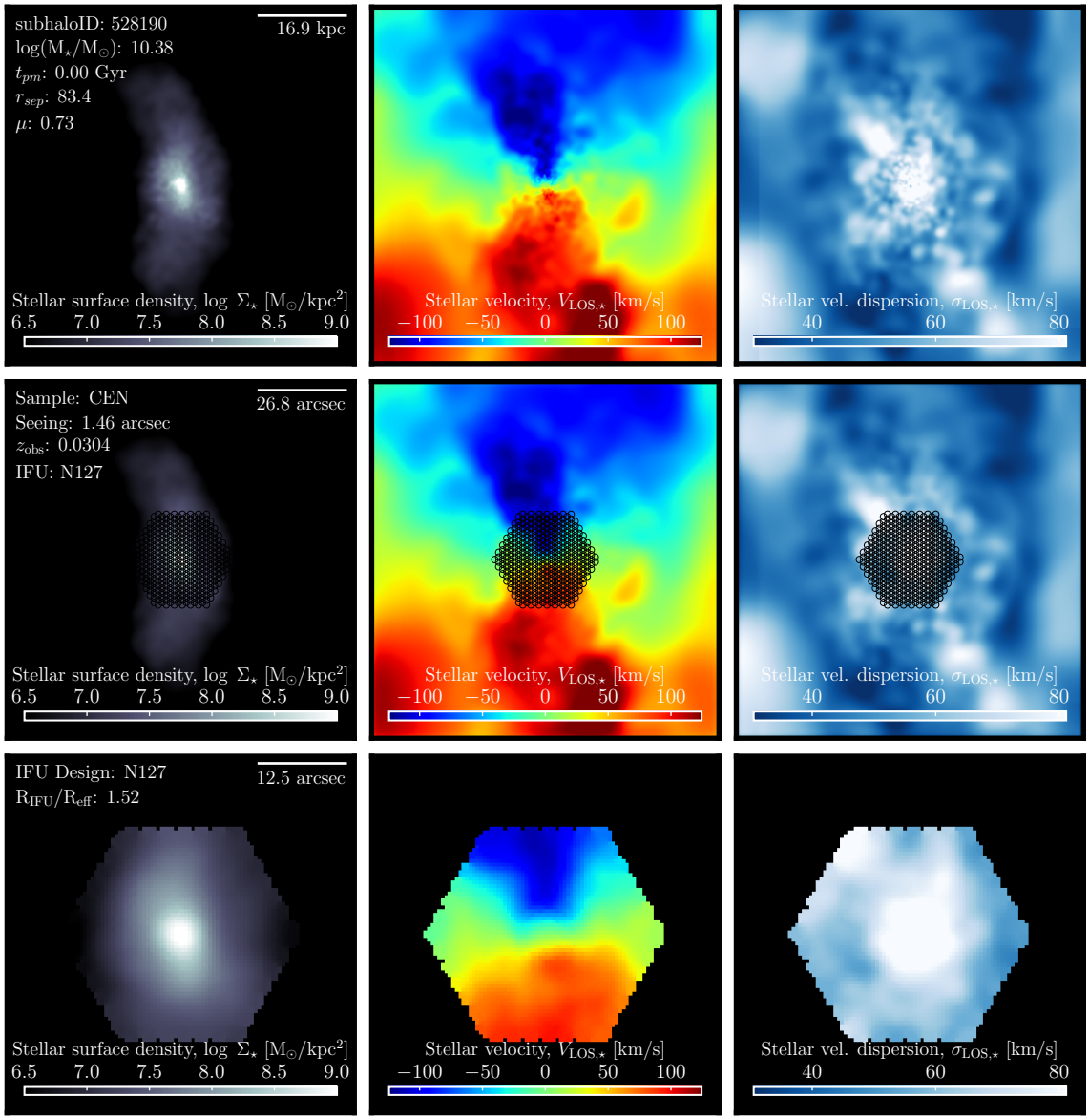

Figure G.78 Break-down of MaNGA synthetic stellar kinematic observations for TNG100-1 galaxy ID: 528190, camera: 3 observed for the CEN sample.