

**Using Machine Learning to Study the Relationship Between
Galaxy Morphology and Evolution**

by

Michael Andrew Peth

A dissertation submitted to The Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

May, 2016

© Michael Andrew Peth 2016

All rights reserved

Abstract

We can track the physical evolution of massive galaxies over time by characterizing the morphological signatures inherent to different mechanisms of galactic assembly. Structural studies rely on a small set of measurements to bin galaxies into disk, spheroid and irregular classifications. These classes are correlated with colors, SF history and stellar masses. Rare and subtle features that are lost in such a generic classification scheme are important for characterizing the evolution of galaxy morphology. We can connect the Hubble sequence observed for local galaxies to their high redshift progenitors to determine the full distribution of galaxy morphologies as a function of time over the entire lifetime of the Universe. To fully capture the complex morphological transformation of galaxies we need more useful classifications. To accomplish such a feat in a computationally tractable way we will need to convert galaxy images to low-dimensional representations of only a few parameters.

To overcome the limitations of the Hubble sequence, we use a principal component analysis of non-parametric morphological indicators (concentration, asymmetry, Gini coefficient, M_{20} , multi-mode, intensity and deviation) measured at rest-frame B -band

ABSTRACT

(corresponding to *HST/WFC3 F125W* at $1.4 < z < 2$) to trace the natural distribution of massive ($> 10^{10} M_{\odot}$) galaxy morphologies. Principal component analysis (PCA) quantifies the correlations between these morphological indicators and determines the relative importance of each. The first three principal components (PCs) capture $\sim 75\%$ of the variance inherent to our sample. We interpret the first principal component (PC) as bulge strength, the second PC as dominated by concentration and the third PC as dominated by asymmetry. PC1 is a better predictor of quenching than stellar mass, as good as other structural indicators (Sérsic-n or compactness). We divide the PCA results into groups using an agglomerative hierarchical clustering method. Distinguishing between these galaxy structural types in a quantitative manner is an important step towards understanding the connections between morphology, galaxy assembly and star-formation.

Using a random forest classification technique, we are able to distinguish mergers from non-merger galaxies in Pan-STARRS imaging using a variety of input features (PCs, non-parametric morphologies, sSFR, M_* , rest-frame color). Determining if a galaxy is a merger is important to understand how influential mergers are in building bulges and assembling galaxies. The galaxies were initially visually classified by users of Galaxy Zoo. Asymmetry is by far the most important indicator of whether a galaxy is experiencing a merger. The next most important features include: PC7, PC5, PC3, deviation and $d(G, M_{20})$. The importance of PC7 represents a very interesting result because PC7 is the least important PC but plays a huge role in determining whether

ABSTRACT

a galaxy is a merger.

Galaxy simulations can provide valuable insight into the mechanisms behind galaxy evolution. The VELA simulations and subsequent non-parametric morphological measurements provide a resource to study the connection between morphology (through the use of PC results) and physical properties (such as sSFR, gas fraction, etc.). We stack the results of a discrete cross correlation between PCs and physical parameters from 9 VELA galaxies. Each of the first three PCs correlates differently with these physical parameters: PC1 is correlated strongly with ex-situ stellar mass, the gas fraction and sSFR; PC2 is weakly anti-correlated with all physical properties; PC3 is strongly correlated with sSFR at all length scales and with gas fraction in the central kpc. The process of star-formation, gas accretion and bulge assembly is a messy picture that will require more simulate galaxies to further understand the process of galaxy evolution.

Primary Reader: Dr. Jennifer Lotz (Space Telescope Science Institute)

Secondary Reader:

Acknowledgments

I first need to thank my family: my mom, my dad, my sister and my grandma. All of you have been with me every step of the way. Without all the trips to the Air & Space Museum would I even be here writing this? There have been so many struggles along the way but each of you have kept me sane, kept me happy and made me feel loved. I have to thank my girlfriend, Ilana. No one has been able to help me believe in myself more than her. She has been a source of strength during the often difficult experience of writing this thesis and hunting for jobs.

I owe so much to my advisor Jennifer Lotz. Without her help, guidance and mentoring I would have been completely lost in graduate school. I feel like I won the lottery in terms of advisors. I couldn't have asked for someone more patient, thoughtful and understanding. Not to mention she is one of the most intelligent people I've ever met and I am constantly in awe of how she is able to understand and dissect any topic in discussion. I owe a debt of gratitude to Gregory Snyder and Alireza Mortazavi. Our weekly group discussion were very helpful and insightful. I can trace the origins of a good chunk of this theses to comments and suggestions

ACKNOWLEDGMENTS

made in those meetings. I also would like to thank our collaborators Peter Freeman and David Thilker who really provided me with the set of tools and impressive data (especially chapter 3). I need to thank my committee member and original co-advisor Harry Ferguson. I am very grateful to have been included in the CANDELS research team as I have been introduced to so many smart and interesting people, while also providing a chance to spread my scientific wings. None of this thesis would have been possible without contributions from those who crated the photometric, SED, visual morphology, VELA simulation and other assorted catalogs. I also need to thank my committee member Tamas Budavari who taught me so much about SQL, machine learning and data science in general in his course. I would like to thank the remainder of my thesis committee: Alex Szalay and Petar Maksimovic for reading my thesis, asking insightful questions and offering quality feedback. The thesis defense would not have been possible without the efforts of Jessica Rexroad and Kelley Key who really made it happen.

My fellow graduate students have been what has made this whole experience worth it. It all began with my friends who entered with me back in 2010: JT Mlack, Keith Redwine, Chris Martin, Kevin Grizzard and everybody else. We lived through the worst of the prelim exams, the POE and classes together, all while taking time to enjoy life and get together often. My 401 family (Raymond Simons, Rachael Alexandroff and Kirill Tchernyshyov) and the rest of the family tree (Duncan Watts, Roseanne Cheng, Erini Lambrides, David Jones and the rabble) hold a special place

ACKNOWLEDGMENTS

in my heart. My 401 family has been a source of inspiration in my research but also have been a source of so many great experiences. No one will ever be able to top the homemade fried chicken found at Bones 'n' Thrones. I truly would not be the person I am today without each of your friendships.

I thank the editors and anonymous reviewers of the *Monthly Notices of the Royal Astronomical Society* who were very helpful in getting chapter 2 published as Peth et al. (2016). This thesis used the following python libraries extensively: matplotlib, sci-kit learn, pandas, astropy, APLpy. I would like to thank the writers and maintainers of these particular libraries for creating such quality products.

Lastly, I need to thank *The Simpsons*, *Seinfeld*, *Curb Your Enthusiasm* and *The Sopranos*. When work became extremely stressful these shows provided me with some much needed comfort. And thank you Radiohead for releasing new music right as work on this thesis reached a fever pitch.

Dedication

for my family

DEDICATION

When I heard the learnd astronomer;
When the proofs, the figures, were ranged in columns before me;
When I was shown the charts and the diagrams, to add, divide, and
measure them;
When I, sitting, heard the astronomer, where he lectured with much ap-
plause in the lecture-room,
How soon, unaccountable, I became tired and sick;
Till rising and gliding out, I wanderd off by myself,
In the mystical moist night-air, and from time to time,
Lookd up in perfect silence at the stars.

– Walt Whitman

Contents

Abstract	ii
Acknowledgments	v
Contents	x
List of Tables	xv
List of Figures	xvii
1 Introduction	1
1.1 A Very Brief Overview of Galaxies	1
1.2 Physical Mechanisms Causing Galaxy Evolution	4
1.3 Galaxy Morphology as a Tool to Study Evolution	9
1.4 Using Machine Learning to Analyze Galaxy Morphology	12
1.5 Data Sets Used in This Analysis	16
2 Beyond Spheroids and Discs: Classifications of CANDELS Galaxy	

CONTENTS

Structure at $1.4 < z < 2$ via Principal Component Analysis	19
2.1 Introduction	19
2.2 Data	22
2.2.1 Sample Selection Criteria	23
2.2.2 Galaxies with FLAG=1	25
2.3 Morphological Measurements	25
2.3.1 Non-parametric Morphology	25
2.3.1.1 Petrosian Radius	27
2.3.1.2 Concentration	27
2.3.1.3 Asymmetry	28
2.3.1.4 Gini Coefficient	29
2.3.1.5 M_{20}	31
2.3.1.6 Multi-mode	32
2.3.1.7 Intensity	33
2.3.1.8 Deviation	34
2.3.2 Morphological Principal Components	34
2.3.3 Concentration - Sérsic Index Relationship	37
2.4 PCA-Morphology Group Properties	48
2.4.1 Defining PCA morphology groups	48
2.4.2 Morphological Error Estimation	50
2.5 PCA Morphology Groups at $z \sim 1.5$	55

CONTENTS

2.6 Discussion	74
2.6.1 Stellar Mass - Quenching Connection for groups	76
2.6.2 The relationship between PCA Classes and Visual/Sérsic Classifications	77
2.6.2.1 The Compact and Bulge-Dominated Galaxies: Groups 0, 6 and 9	78
2.6.2.2 The Disk-dominated Galaxies: Groups 1, 2 and 5 . .	80
2.6.2.3 The Intermediate Galaxies: Groups 4 and 8	81
2.6.2.4 Comparing the Irregular Galaxies of Groups 1 and 9	82
2.7 Summary	84
3 Merger Classifications of Pan-STARRS Galaxies Using Random Forest	96
3.1 Introduction	96
3.2 Data	101
3.2.1 Ground Based Surveys: Pan-STARRS and SDSS	101
3.2.2 Galaxy Zoo	102
3.3 Non-parametric Morphology of Pan-STARRS Galaxies	105
3.3.1 Principal Component Analysis of PanSTARRS Galaxies . . .	108
3.4 Random Forest Classifier	115
3.4.1 Random Forest Inputs	120
3.5 Results	121

CONTENTS

3.5.1	Random Forest Classifications	121
3.5.1.1	Random Forest Input Parameter Tests	123
3.5.1.2	Comparisons of RF on Different Subsamples	130
3.6	Discussion	140
3.6.1	Random Forest Classifications of MaNGA Galaxies	140
3.7	Summary and Conclusions	142
4	VELA Simulation Galaxy Morphologies	148
4.1	Introduction	149
4.2	Data	156
4.2.1	VELA Simulation	156
4.2.2	Image Processing and CANDELization	158
4.3	PCA-Morphology Groups	160
4.3.1	PC Group Demographics	161
4.3.2	PC Group Flow	165
4.4	Time Series Cross-Correlations	175
4.4.1	Discrete Correlation Function	179
4.4.2	Stacks of PCs and Physical Parameters Correlations	188
4.5	Discussion	212
4.5.1	Is PC1 an indicator of evolution?	213
4.5.2	How PC2 interacts with galaxy evolution	216
4.5.3	PC3 and Star-Formation	217

CONTENTS

4.6 Summary	219
5 Summary and Future Directions	222
5.1 Galaxy Morphological Classifications Using PCA	223
5.2 Random Forest Classifications of Pan-STARRS Galaxies	224
5.3 Studying Galaxy Morphology Using VELA Simulation Suite	226
5.4 Mergers Can Grow Bulges and Regulate Star-formation	227
5.5 Future Work	230
Bibliography	233
Vita	250

List of Tables

2.1	PC Weights with error estimates based on a bootstrap scattering method	38
2.2	Group percentages by mass range for both original group and “MC Group”	63
2.3	Demographics of Visual Classifications of Groups	64
2.4	Demographics of Sérsic Classifications of Groups	65
2.5	<i>UVJ</i> Quenched Fractions of Groups	66
3.1	Feature Importances of RF Classifications	128
3.2	Feature Importances of RF Classifications	129
3.3	Summary Statistics of RF Classifications	135
3.4	Summary Statistics of RF Classifications	136
3.5	False Positive Examples	145
4.1	VELA PCA Group Demographics at all viewing angles. The CAN-DELS sample is shown for comparison.	166

LIST OF TABLES

4.2 VELA PCA Group Demographics by Camera Angle. CAMERA0 represents the face-on view and CAMERA1 represents the edge-on view. The remaining cameras are from random angles. The CANDELS sample is shown for comparison.	167
---	-----

List of Figures

2.1	Histogram of F125W J-band Magnitude for galaxies with FLAG=1 and FLAG=0	26
2.2	Example of <i>MID</i> statistics	30
2.3	Concentration - Sérsic index for sample	41
2.4	Concentration - Sérsic Including PSF Effects	42
2.5	PC1 vs Sersic Index and Gini- M_{20} Bulge Strength	43
2.6	3D PC1-PC2-PC3 color-coded by group	44
2.7	Between cluster variance vs. N clusters	45
2.8	Magnitude vs. $\Delta(\text{GOODS} - \text{UDF})$ morphological statistics	46
2.9	Magnitude vs. $\Delta(\text{GOODS} - \text{UDF})$ morphological statistics	47
2.10	Group classification uncertainty	51
2.11	Rest-frame <i>UVJ</i>	52
2.12	Rest-frame $U - V$ vs. Stellar Mass	56
2.13	Gini – M_{20}	57
2.14	Concentration – Asymmetry	58

LIST OF FIGURES

2.15 Effective radii (kpc) - M_*	59
2.16 Cumulative quenched fraction rank	75
2.17 Group 6 F125W $1.36 < z < 1.97$ galaxies	87
2.18 Group 0 F125W $1.36 < z < 1.97$ galaxies	88
2.19 Group 9 F125W $1.36 < z < 1.97$ galaxies	89
2.20 Group 4 F125W $1.36 < z < 1.97$ galaxies	90
2.21 Group 8 F125W $1.36 < z < 1.97$ galaxies	91
2.22 Group 1 F125W $1.36 < z < 1.97$ galaxies	92
2.23 Group 2 F125W $1.36 < z < 1.97$ galaxies	93
2.24 Group 5 F125W $1.36 < z < 1.97$ galaxies	94
2.25 Group -1 F125W $1.36 < z < 1.97$ galaxies	95
3.1 Galaxy Zoo Decision Tree	106
3.2 Gini - M_{20} for PANSTARRS Mergers	109
3.3 Concentration - Asymmetry for PANSTARRS Mergers	110
3.4 Concentration - Asymmetry for PANSTARRS Mergers	111
3.5 PC1-PC2-PC3 Plot for PANSTARRS Galaxies	113
3.6 Histogram of PC groups for CANDELS and Pan-STARRS	114
3.7 Visualization of Random Forest Part 1	118
3.8 Visualization of Random Forest Part 2	119
3.9 Gini- M_{20} color-coded by RF Classification Probabilities	124
3.10 Gini- M_{20} coded by RF Classification Confusion Matrix	125

LIST OF FIGURES

3.11 Concentration-Asymmetry color-coded by RF Classification Probabilities	126
3.12 Concentration–Asymmetry coded by RF Classification Confusion Matrix	127
3.13 OOB Errors for Random Forests Using Different Numbers of Max Leaf Nodes	131
3.14 Summary Statistics for Random Forests Using Different Max Leaf Nodes	132
3.15 ROC Curve	133
3.16 Feature Importance Comparisons between Blue and Red Galaxies	137
3.17 Feature Importance Comparisons between Merged Galaxies and Merg- ing Pairs	138
3.18 Feature Importance Comparisons between the Full Sample and Blue/Red, Merged Galaxies/Merging Pairs	139
3.19 Images of False Positives	143
3.20 Segmentation Maps of False Positives	144
4.1 PC group histogram of VELA galaxies	168
4.2 Gini– M_{20} of VELA galaxies	169
4.3 Concentration–Asymmetry of VELA galaxies	170
4.4 PC group histogram of VELA galaxies with Asym Correction	171
4.5 PC group histogram of minor mergers	172
4.6 PCgroup(z) for all VELA galaxies	176
4.7 $f_{gas}(t)$ for all VELA galaxies	177
4.8 PC1 for all VELA galaxies	189

LIST OF FIGURES

4.9	PC2 for all VELA galaxies	190
4.10	PC3 for all VELA galaxies	191
4.11	sSFR for all VELA galaxies	192
4.12	f_{gas} for all VELA galaxies	193
4.13	\dot{f}_{gas} for all VELA galaxies	194
4.14	ex-situ M_* for all VELA galaxies	195
4.15	ex-situ \dot{M}_* for all VELA galaxies	196
4.16	\dot{M}_{dm} for all VELA galaxies	197
4.17	sSFR-PC1 Cross-correlation for VELA02	198
4.18	Gas Fraction - PC1 Cross-correlation for VELA02	199
4.19	Rate of gass mass into central kpc - PC1 of VELA02	200
4.20	Ex-situ stellar mass into central kpc - PC1 of VELA02	201
4.21	Rate Ex-situ stellar mass into central kpc - PC1 of VELA02	202
4.22	Rate of DM mass into central kpc - PC1 of VELA02	203
4.23	Stack of cross-correlations for PC1 - physical parameters for inner kpc	206
4.24	Stack of cross-correlations for PC1 - physical parameters for total galaxy	207
4.25	Stack of cross-correlations for PC2 - physical parameters for inner kpc	208
4.26	Stack of cross-correlations for PC2 - physical parameters for total galaxy	209
4.27	Stack of cross-correlations for PC3 - physical parameters for inner kpc	210
4.28	Stack of cross-correlations for PC3 - physical parameters for total galaxy	211

Chapter 1

Introduction

1.1 A Very Brief Overview of Galaxies

Initially after the Big Bang, the Universe was in a state of near but not perfect homogeneity with small quantum fluctuations present throughout. Following a period of rapid expansion in the Universe, known as inflation, these quantum fluctuations became amplified into regions of higher and lower density. At this point the Universe was radiation dominated and all primordial elements (such as hydrogen) were fully ionized. However, the ionized photons could not travel very far without Thomson scattering off a free electron. The continuing expansion cooled the Universe enough that it became energetically possible for protons and electrons to combine and form neutral hydrogen. This era, known as recombination (or decoupling), brought about the opportunity for baryonic matter assembly. Photons became decoupled from the

CHAPTER 1. INTRODUCTION

formerly charged particles and became free to propagate throughout the universe. These photons are visible as the Cosmic Microwave Background (CMB). The CMB is nearly uniform except for slight temperature fluctuations on the order of 10^{-5} K (Bennett et al., 2013; Mather et al., 1990; Planck Collaboration et al., 2014). These temperature fluctuations are the evidence of the density fluctuations of the post-inflation Universe.

The standard model of cosmology, known as Lambda-Cold Dark Matter (Λ CDM), posits the existence of “dark energy” which is responsible for counteracting the attractive effects of gravity and “cold dark matter” that only interacts with itself and other particles through gravity and does not radiate photons. Dark matter clumps grew from the perturbations in the density distribution of the Universe.

Dark matter is able to collapse in a dissipational manner (does not radiate away energy through photons) due to gravity and forms halos. The smallest dark matter halos are able to form first, later merging with one another to create progressively larger halos (White & Rees, 1978). This growth of dark matter halos is known as hierarchical assembly and is central to Λ CDM cosmology. Baryonic matter (in the form of gas) is accreted by these halos at which time the gas cools and fragments to form galaxy structures. Eventually, dark matter halos accrete enough gas to form what we know of as galaxies.

Galaxies can continually accrete material either smoothly or stochastically from the surrounding intergalactic medium. Smooth accretion in the form of cold gas dis-

CHAPTER 1. INTRODUCTION

tributed along dark matter filamentary structure is directly dumped onto the galaxy (Birnboim & Dekel, 2003; Dekel et al., 2009b). These so-called cold streams are among the main sources of gas for higher redshift galaxies (Dekel et al., 2009b). Stochastic accretion can occur in the form of merging galaxies (see §1.2).

Modern cosmological simulations (such as Illustris Vogelsberger et al., 2014 and EAGLE McAlpine et al., 2015) have successfully reproduced how observed galaxies form and grow in dark matter halos through the constant collapse of molecular clouds into stars and the gravitational attraction to form increasingly complex structures (Springel et al., 2005).

The most widely used visual classification scheme, the Hubble sequence, divides galaxies into ellipticals (also known as early-type galaxies), transitional phase (known as lenticular galaxies) and spiral galaxies (also known as late type galaxies) (Hubble, 1926). The elliptical galaxies vary in elongation from round to triaxial shapes and have smooth light profiles, star follow random orbits and appear spheroidal. The spiral galaxies consist of stars orbiting rotationally in spiral structures are subdivided by how tightly wound the spiral arms are and is a central bar exists. Typically the spectral color of a galaxy is related to the morphology: elliptical galaxies are composed of red and old stars, while spiral galaxies are composed of blue and young stars. Galaxies not fitting into this scheme are labeled as irregular. Irregular galaxies can be low mass galaxies or the result of the merger of two galaxies.

All of these galaxy characteristics beg the questions: why do galaxies look the

CHAPTER 1. INTRODUCTION

way they do? What physical mechanisms build galaxies into these specific structures and either create a large number of stars or prevent stars from forming?

1.2 Physical Mechanisms Causing Galaxy Evolution

There exists a strong correlation between the rate of star-formation and the amount of stellar mass, known as the “main sequence of star formation” as far back as $z \sim 2.5$ (Noeske et al., 2007; Wuyts et al., 2011). In this correlation, there exists a bi-modality in star-formation and stellar mass that is highly correlated with color and morphological type. Blue, star forming, primarily disk galaxies have star-formations and masses that follow a very tight relationship (e.g. Baldry & Glazebrook, 2003; Hogg et al., 2004; Bell et al., 2004). Meanwhile, red, low star-formation, primarily spheroidal galaxies fall below this relationship and have less star-formation than a bluer galaxy has for a specific mass and redshift. Galaxies with star-formation below the main sequence are known as “quenched”.

During the epoch known as “cosmic high noon” ($z=1.5 - 3$), the cosmic star formation rate is at a maximum and at which time nearly half of all stellar mass assembles (Madau & Dickinson, 2014). Galaxies were forming more stars per unit mass at higher redshift (Noeske et al., 2007). Even at this epoch, massive galaxies ($M_* > 10^{10} M_\odot$) begin to experience declining star formation, which is coupled with

CHAPTER 1. INTRODUCTION

an emergence of red central bulges (Kriek et al., 2006; van Dokkum et al., 2008; Kriek et al., 2009; Whitaker et al., 2012). Since “cosmic high noon” there has been a dramatic increase in the number of high mass quenched galaxies observed (e.g. Faber et al., 2007; Bell et al., 2012).

Any discussion of the overall galaxy morphology and star-formation characteristics would be incomplete without a discussion of bulges. Not all bulges are created equal, there are a few different structures which may collectively be called “bulges” but which are different from one another. There are “classical” bulges which resemble giant elliptical galaxies, but exist at the center of disk galaxies. The stars in these bulges are on random orbits and are redder than the stars in the disk. The light distribution is well described by the de Vaucouleurs law (surface brightness $\propto r^{1/4}$). A classical bulge is likely the final stage of the merger of two disk galaxies (Toomre, 1977; Kormendy & Kennicutt, 2004). Additionally, there are “pseudo-bulges” which are spheroidal and exist at the center of disk galaxies, however the stars orbit the center rotationally (similar to the outer disk). Pseudo-bulge light profiles are not well described by the de Vaucouleurs profile and are instead better fit by a Sersic profile ($\propto r$). Pseudo-bulges are likely the result of internal galaxy interactions such as bars and spiral structure (Kormendy & Kennicutt, 2004). Understanding the difference between these two types of bulges can have an impact on the likely formation mechanisms for a particular galaxy.

Bulges are also not the same at different redshifts. At higher redshift very small

CHAPTER 1. INTRODUCTION

($r \lesssim 1\text{--}3$ kpc) galaxies can resemble local elliptical galaxies but are actually a separate class, known as “compact” galaxies. Compact galaxies likely formed via gas inflows towards the central region of the galaxy. Quenched compact galaxies can have radii of 1 kpc or smaller (van der Wel et al., 2014a). Many $z \sim 3$ galaxies are compact elliptical galaxies with low amounts of star formation (van Dokkum et al., 2008, 2010; Whitaker et al., 2012). Compact, star forming galaxies have similar masses, kinematics, and abundances as quenched, red compact galaxies and are the likely progenitors (Barro et al., 2013, 2014b; Williams et al., 2014). Both types of compact galaxies are seen in hydrodynamical (Ceverino et al., 2014; Wellons et al., 2015) and semi-analytic (Brennan et al., 2015) simulations.

Why galaxies experience this reduction in star formation and bulge formation is hotly debated. Observations reveal a cosmic transition from blue and star forming disk galaxies to red and quenched spheroidal galaxies leading to an interesting “chicken or egg” problem: Do galaxies experience a morphological transformation that quenches star formation, or does star formation quenching lead to a fading disk? The mechanisms quenching star formation and affecting the morphology of galaxies are not fully understood but can be explained in a few different ways: major/minor mergers (e.g. Naab et al., 2006a; Hopkins et al., 2010); feedback from active galactic nuclei (AGN; e.g. Croton et al., 2006; Somerville et al., 2008a); secular processes (such as the spiral bar instabilities, star formation, gas recycling Kormendy & Kennicutt, 2004; Bournaud et al., 2007; Elmegreen et al., 2008; Genzel et al., 2008).

CHAPTER 1. INTRODUCTION

Mergers are defined by their mass ratios (major or minor) and their gas content (gas-rich or “wet” and gas-poor or “dry”). Each type of merger can influence star-formation and morphology in a different manner.

Major mergers (collisions between galaxies of roughly equivalent mass, mass ratio of $\lesssim 1:3$) can destroy disks by the gravitational interactions of the constituent galaxies and eventually reassemble into a relaxed spheroid. Galaxies with significant gas fractions interact which leads to peculiar features such as tidal tails, asymmetries, double nuclei, rings, shells (Toomre & Toomre, 1972). Major gas-rich galaxy mergers rapidly funnel gas into the cores of massive galaxies and feeds bulges (e.g. Sanders & Mirabel, 1996; Heckman et al., 2004). Gas-rich mergers provide a supply of star-forming fuel which can lead to starburst activity. Meanwhile, Gas-poor mergers are primarily responsible for the mass and size evolution of spheroids at $z < 2$ (Naab et al., 2006a, 2009).

Minor mergers (which are generally between galaxies with a mass ratio of $> 1:10$) may also disrupt morphologies, and gas-poor minor mergers must be more frequent than major mergers (Lotz et al., 2011; Papovich et al., 2012). Peculiar properties, such as low surface brightness tidal features, are often difficult to detect and require deep observations. The primary galaxy accretes stellar material from the satellite onto the outskirts (e.g. Naab et al., 2006b; Bell et al., 2006). Even the small amount of gas accreted in a minor merger is sufficient to trigger an AGN or starburst, and eventually quench star-formation (Kormendy & Richstone, 1995; Croton et al., 2006;

CHAPTER 1. INTRODUCTION

Somerville et al., 2008b).

Internal mechanisms, collectively referred to as secular processes, include the interactions of bars in a spiral galaxy rearranging disk gas (Kormendy & Kennicutt, 2004), and violent disk instabilities (VDIs, Kereš et al., 2005) leading to enhanced star-formation, irregular morphologies, angular momentum loss, rapid star-formation and supermassive black hole (SMBH) growth (Magorrian et al., 1998; Ferrarese & Merritt, 2000; Shankar et al., 2012; Elbaz & Cesarsky, 2003). In this scenario, the morphology of the galaxy is unaffected and the galaxy appears undisturbed and disk-like (Simard & Pritchett, 1998; Schawinski et al., 2011). Once the reservoir of gas is exhausted and star-formation is quenched, a disk structure can still exist.

Slow, long-term quenching mechanisms are required to keep galaxies quenched (Barro et al., 2013). This quenched state can be maintained by mechanisms such as mass quenching (Dekel & Birnboim, 2006; Bell et al., 2012) which is caused by the halo growing above a threshold mass of $10^{11} M_{\odot}$. At this mass, shocks are created which do not allow gas to cool sufficiently to form stars. Quenching can also be maintained by a sufficiently massive central bulge stabilizing the disk from further fragmentation and thus shutting down star formation (morphological quenching; Tacchella et al., 2015; Martig et al., 2009; Genzel et al., 2014). Additionally, AGN can provide strong jets that can heat the surrounding halo and thus prevent gas to cool and form stars (Cattaneo et al., 2009). On the other hand bulges, by themselves, have proven to be a “necessary but not sufficient” mechanism to shut down star-formation (Bell et al.,

CHAPTER 1. INTRODUCTION

2012; Fang et al., 2013).

Each mechanism leaves behind different clues (in the shape and structure of galaxies). Can we determine which mechanisms are important for a specific type of galaxy during a specific cosmic epoch? A possible answer is in the morphology of a galaxy. The shape and structure can tell us what processes have been important during a galaxy's history.

1.3 Galaxy Morphology as a Tool to Study Evolution

Morphology can offer clues that indicate how responsible mergers (and other mechanisms) are (or are not) in quenching galaxies and building bulges. Morphological classes (such as spheroids and disks) are correlated with colors, star-formation history and stellar masses. Significant correlations have been observed between star-formation rate, stellar mass and quantitative morphological measurements (Wuyts et al., 2011).

To study the processes driving evolution, we need a method to effectively and efficiently characterize the structures and shapes of galaxies. Visual classifications (such as the Hubble sequence) have been used since the discovery of galaxies, and have subsequently been adapted to fit modern surveys (e.g. Galaxy Zoo, Lintott et al., 2008a; Kartaltepe et al., 2015). These visual studies rely on the Hubble sequence to

CHAPTER 1. INTRODUCTION

classify galaxies and will have classifiers place galaxies into disk, spheroids, irregular and unknown categories. Visual classifications can find subtle structural elements possibly missed by an automated routine. However, human classifications of galaxies can be very time consuming and subjective.

However, galaxy structure at high redshift does not always correspond to the local Hubble sequence (Bruce et al., 2012; Bell et al., 2012; Kriek et al., 2009; Lee et al., 2013). Disk-dominated galaxies can appear clumpy (Förster Schreiber et al., 2009) and spheroid-dominated galaxies can be compact, very red and massive, but possess no extended envelope (e.g. van Dokkum et al., 2008). Therefore the standard Hubble sequence will miss rare and subtle features inherent to the morphology of high redshift galaxies and may need updating for high redshift.

Galaxies can appear vastly different between UV and optical wavelengths (e.g., Meurer et al., 1995). UV light traces bright stars and thus active star formation (since these stars are short-lived). Meanwhile, optical wavelengths longer than the Balmer (400 nm) break observe stars at a variety of ages. Progressively older stars dominate the galaxy spectral energy distribution (SED) at longer wavelengths. Additionally, dusty galaxies can have much of their optical light absorbed (Calzetti et al., 2000) and reradiated in the IR. To combat these wavelength-dependent morphological conditions it is important to observe galaxy morphology at a single rest-frame wavelength across redshift.

To combat the subjectivity of visual classifications, quantitative measurements

CHAPTER 1. INTRODUCTION

defining morphology have been created. The relationship between surface brightness and radius for elliptical galaxies ($I \propto r^{1/4}$) was first determined by de Vaucouleurs (1948). The de Vaucouleurs law was eventually generalized by Sersic (1968) to a Sérsic profile ($I \propto r^{1/n}$) with disk galaxies of $n=1$. Later studies decomposes the galaxy into bulge and disk profiles (Kormendy, 1977b) for even further discriminatory power between disks and bulge dominated galaxies. Many studies (e.g. Bell et al., 2012; van der Wel et al., 2012) fit a Sérsic profile to a galaxy for the purposes of classification.

GALFIT (Peng et al., 2002, 2010) is an automated technique often used to classify galaxies by fitting the galaxy light distribution to a Sérsic profile ($r^{-1/n}$) and is sensitive to small galaxies, can distinguish overlapping light profiles of nearby galaxies, incorporates the point spread function of a specific field/detector, and most importantly is easy to interpret. However, **GALFIT** assumes a symmetric and smooth light profile, which at times can be problematic. This assumption does not hold for irregular galaxies, merger remnants, and disk galaxies with bars or clumps.

Quantitative non-parametric morphological statistics characterize galaxy structure and do not assume an analytic light profile. This fact allows us to apply automated characterization to irregular galaxies as well. Examples of non-parametric morphological indicators include: concentration index (C , Bershady et al., 2000; Conselice et al., 2003), asymmetry (A , Conselice et al., 2000), Gini coefficient (G , Abraham et al., 2003; Lotz et al., 2004), M_{20} (Lotz et al., 2004), and three new statistics from Freeman et al. (2013): Multimode (M), Intensity (I), and Deviation (D). The

CHAPTER 1. INTRODUCTION

MID statistics have been found to be the most sensitive to mergers and clumpy star-formation, even at high redshift (Freeman et al., 2013). *CAS* is capable of identifying major mergers, while Gini– M_{20} can identify both major and minor mergers (just not to the same extent as the *MID* statistics, Conselice, 2014).

However, for many galaxies these statistics can be strongly correlated. Moreover, cosmological models of galaxy formation yield a picture in which these structures can evolve quickly along diverse paths, thereby motivating the need for a broad classification system (Snyder et al., 2015a). Therefore we require further analysis to understand the inherent relationships among these statistics and between galaxy assembly processes.

1.4 Using Machine Learning to Analyze Galaxy Morphology

In the upcoming years and decades, many new telescopes and surveys will become operational; such as the Large Synoptic Sky Telescope (LSST; Ivezić et al., 2008), the European Extremely Large Telescope (E-ELT), the Thirty Meter Telescope (TMT), and the Dark Energy Survey (DES), among others. Each of these telescopes will produce terabytes to petabytes of observational data nightly. Novel data analysis strategies will need to be created to account for the sheer deluge of information. These massive data sets will provide significant insights into every aspect of astrophysics to

CHAPTER 1. INTRODUCTION

a degree that only a decade ago may have seemed outlandish.

The sheer amount of images from future telescope surveys will make human visual classifications of galaxies an intractable problem. However, machine learning techniques are often successful at reproducing many of the results. In their review of data mining in astronomy, Ball & Brunner (2010), state the advantages as follows: simplicity, influence from prior information, pattern recognition, complimentary analysis and the simple ability to “get anything at all”. Complimentary analysis refers to the idea that different approaches to a problem will reduce the systematic errors inherent to any single approach.

To make sense of all this data, astronomers have begun to implement machine learning and data mining into their analysis. Data mining is simply a collection of techniques useful for analyzing and describing structured data (Ivezić et al., 2013). These techniques include: principal component analysis (PCA), clustering, unsupervised classification, amongst many others. Machine learning refers to a set of techniques that compare datasets to previously understood sets. These techniques include: random forest (RF), support vector machines (SVM), artificial neural networks (ANN) and maximum likelihood estimator.

There are two broad categories of machine learning techniques: supervised and unsupervised. Unsupervised techniques (such as principal component analysis, see Chapter 2) are helpful to reduce the dimensionality of a problem and to find relationships amongst the data. Supervised learning techniques, such as random forest

CHAPTER 1. INTRODUCTION

(Breiman, 2001), support vector machines (Vapnik & Vapnik, 1998), and artificial neural networks (ANN; Ripley, 1981, 1988), use a training set of labeled data to build a framework for which to classify unlabeled data.

Principal component analysis (PCA) is a simple way to reduce the dimensionality, break internal degeneracies and find the natural distributions of data in parameter space. To eliminate degeneracies inherent in these morphological statistics we performed a PCA using 7 non-parametric morphology measurements on 1244 galaxies from $1.36 < z < 1.97$. PCA has been shown to efficiently classify galaxies (e.g. Taghizadeh-Popp et al., 2012; the Zurich Estimator of Structural Types (ZEST), Scarlata et al., 2007a). A few studies immediately capitalized on the ZEST classifications to study the number density evolution of disk galaxies (Sargent et al., 2007), the luminosity function evolution for elliptical galaxy progenitors (Scarlata et al., 2007b), and the evolution of the galaxy merger rate to $z \sim 1$ (Kampczyk et al., 2007).

The Zurich Estimator of Structural Types (ZEST; Scarlata et al., 2007a) uses a PCA of 5 non-parametric morphological diagnostics: Gini coefficient, M_{20} , concentration, asymmetry, and ellipticity. They classify $\sim 56,000$ bright ($I_{AB} < 24$) COSMOS into spheroidal, disk and irregular galaxy types while additionally calculating a bulginess, elongation, irregularity and clumpiness parameter for each galaxy. The classifications are used to demonstrate redshift evolution (since $z \sim 1$) of the galactic luminosity function (LF) for galaxies of different classes. Their analysis concluded that the average volume density of disk galaxies remains constant. However, the stel-

CHAPTER 1. INTRODUCTION

lar populations of these systems are brightened at earlier epochs. Only the bright, ($M_B < -21.5$) end of the irregular and the early-type galaxies remains roughly consistent with the LF of local galaxies. At fainter magnitudes, irregular and early-type galaxies show evolution from $z = 0$ to 0.7.

Similarly, Taghizadeh-Popp et al. (2012) uses PCA to describe the entire zoo of galaxy morphologies with a single parameter. Which they derived from a set of observational derived quantities: mass-to-light ratio, surface brightness, concentration, star-formation rate, specific star-formation rate, $g-r$ and $u-r$. Their analysis labels, ranks and classifies galaxies by a single arc-length value.

Supervised methods such as random forest have been used to classify galaxies (e.g. Lahav et al., 1995; Freeman et al., 2013). The random forest technique was developed by Breiman (2001) as a supervised method for classification. The random forest classifier is learned from a labeled training set representing a random sample of the total sample. The split best differentiating mergers from non-mergers among the random subset of the features in each node defines the optimal classifiers. Random forest inherently provides probabilities which we can use to investigate the effect thresholds have on the completeness and quality of classifications.

Supervised techniques require a basis set of data in which all subsequent classifications are founded upon. Freeman et al. (2013) uses the CANDELS visual classifications (Kartaltepe et al., 2015) to build a classification schema out of non-parametric morphologies for separating mergers from non-mergers. The M , I , and D statistics

CHAPTER 1. INTRODUCTION

are more useful than Gini and M_{20} at identifying disturbed morphologies. Lahav et al. (1995) compared visual classifications of galaxies by world experts (such as de Vaucouleurs) to classifications by an Artificial Neural Network. The Sloan Digital Sky Survey was used as a training set (over 143 million objects) to separate galaxies from stars (Ball et al., 2004). Huertas-Company et al. (2015) uses convolutional neural networks to classify galaxies based on non-parametric morphological measurements from CANDELS.

Supervised techniques are not just used to classify galaxies but can be used to infer values such as photometric redshifts and galaxy stellar masses. Kamdar et al. (2016a,b) use random forest regression of semi-analytic models of galaxies to make predictions of observable galaxy properties from pure dark matter simulations. Carliles et al. (2010) uses random forest trained upon SDSS galaxies to calculate photometric redshifts.

1.5 Data Sets Used in This Analysis

The Cosmic Assembly Near Dawn Extragalactic Legacy Survey (CANDELS, PIs: S. Faber and H. Ferguson; Grogin et al. 2011 and Koekemoer et al. 2011) provides a wealth of data from 5 heavily studied fields (UDS, EGS, COSMOS and GOODS-North+South) with observations by the *Hubble Space Telescope* (HST). Space based observations from HST provide the highest resolution ever for a sample of high red-

CHAPTER 1. INTRODUCTION

shift galaxies . High resolution is critical for observations of low surface brightness structural features. Without which, morphological evolution would be incredibly difficult. Observations by the Wide Field Camera 3 (WFC3) in Near-Infrared bands, F125W (*J*) and F160W (*H*), combined with observations from the Advanced Camera for Surveys (ACS) in UV-Visible bands, F814W (*i*) and F606W (*V*) constitute the new measurements in the CANDELS program.

We focus on high mass ($M_* > 10^{10} M_\odot$) galaxies, brighter than $H < 24.5$. We restrict ourselves to only redshift ranges that correspond our observed morphologies to a single rest-frame waveband. Constant rest-frame morphologies are crucial for understanding possible evolution in the stellar structures of galaxies while preventing strong wavelength, and therefore redshift biases.

High redshift observations can be extended to low redshifts through the use of large all sky surveys. In the next few years, the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) will provide a dataset of up to 50,000 galaxies (with an addition 3,000 from a Medium Deep Survey) that will need to be analyzed. Pan-STARRS will take frequent and repetitive wide-field images over nearly the entire visible sky. Additionally, the Sloan Digital Sky Survey (SDSS; York et al., 2000a), is a very well established program that has observed over a million low redshift galaxies. The observations of low redshift galaxies from SDSS and PAN-STARRS will offer a critical baseline for comparison to the high redshift galaxies from CANDELS.

The VELA simulations are a suite of zoom-in hydro-cosmological simulations of

CHAPTER 1. INTRODUCTION

moderately massive galaxies calculated using Eulerian gas dynamics and an N-body Adaptive Refinement tree (ART, Kravtsov et al., 1997; Kravtsov, 2003). The VELA simulations are described in depth by Ceverino et al. (2010a); Ceverino & Klypin (2009); Ceverino et al. (2012); Dekel et al. (2013); Ceverino et al. (2014). The simulation outputs have been processed (using **SUNRISE** Jonsson, 2006; Jonsson et al., 2010 and **CANDELization** Mozena, 2013) to resemble observed galaxies at high redshift by CANDELS (Snyder et al., 2015b). The VELA simulations offer a new avenue to study individual galaxy evolution from $1 \lesssim z \lesssim 3$ and how physical processes are directly related to morphology.

Chapter 2

Beyond Spheroids and Discs:

Classifications of CANDELS

Galaxy Structure at $1.4 < z < 2$ via

Principal Component Analysis

2.1 Introduction

Massive galaxies today form stars at a lower rate than in the past due to many factors. However, we do not have a complete accounting of the processes quenching the star-formation in galaxies. An increase in the mass/number densities (Tomczak et al., 2014; van der Wel et al., 2014b) of massive, red galaxies implies stars are not

CHAPTER 2. PCA MORPHOLOGY

forming to the same extent they once were. Each of these observations attempt to connect of observed color (or star-formation rate) and stellar masses to morphology. The star-formation rate - stellar mass ($\text{SFR}-M_*$) relationship shows star-forming galaxies at $z \sim 0$ follow a “main sequence” (Brinchmann et al., 2004; Wuyts et al., 2011). Galaxies on the main sequence are bluer and have lower Sérsic-indices than galaxies below the relation. Massive galaxies with low SFRs are red and have high Sérsic indices and bulge strengths. The $\text{SFR}-M_*$ morphology relation has been shown to hold out to $z \sim 2.5$ (Wuyts et al., 2011). However, bulge strength has been described as a “necessary but not sufficient” condition for quenching star-formation in $z \lesssim 2.2$ galaxies (Bell et al., 2012).

If the presence of a bulge is not sufficient to fully quench a galaxy, other factors such as size may be important for shutting down star-formation. At redshifts $z \sim 1.5$, galaxies of sufficiently high mass and small size are quenched (Barro et al., 2013). This suggests a relationship between so-called “compactness” ($\Sigma_{1.5} = M/r_e^{1.5}$) and the specific star-formation rate (sSFR) . However, the number density of these compact galaxies has been decreasing with the age of the Universe.

The mechanisms for quenching star-formation and transforming the morphology of galaxies are not fully understood. Proposed mechanisms include: major mergers (e.g. Naab et al., 2006a; Hopkins et al., 2010); minor mergers (e.g. Taniguchi, 1999; Hopkins & Hernquist, 2009; Villforth et al., 2013); secular processes (for review see Kormendy & Kennicutt, 2004; Cisternas et al., 2011); AGN feedback (e.g. Silk &

CHAPTER 2. PCA MORPHOLOGY

Rees, 1998; Schawinski et al., 2006); and mass quenching (Dekel & Birnboim, 2006; Bell et al., 2012). Comprehensive models of galaxy formation can yield a reasonable link between galaxy morphology and star formation (e.g. Snyder et al., 2015b) but we do not yet have a perfect accounting of how all these processes might contribute.

As a result, two evolutionary tracks have been developed to explain the disappearance of compact, quenched galaxies: (1) major mergers at $z \sim 2\text{-}3$ quickly cause a galaxy to quench, which later grow through minor mergers and gas accretion; (2) violent disk instabilities/secular processes/minor mergers at $z \sim 1.5$ cause a slower decline in star-formation and simultaneous size growth before the quiescent phase.

To study the processes driving evolution, we need a method to effectively and efficiently characterize the structures and shapes of galaxies.

Quantitative non-parametric morphological statistics characterize galaxy structure and do not assume an analytic light profile. This fact allows us to apply automated characterization to irregular galaxies as well. Examples of non-parametric morphological indicators include: concentration index (C , Bershady et al., 2000; Conselice et al., 2003), asymmetry (A , Conselice et al., 2000), Gini coefficient (G , Abraham et al., 2003; Lotz et al., 2004), M_{20} (Lotz et al., 2004), and three new statistics from Freeman et al. (2013): Multimode (M), Intensity (I), and Deviation (D). The MID statistics have been found to be sensitive to mergers and clumpy star-formation, even at high redshift (Freeman et al., 2013).

In this chapter, we use PCA and hierarchical clustering to classify galaxies based

CHAPTER 2. PCA MORPHOLOGY

on their structure. These classifications allow us to characterize galaxies by more subtle means than the traditional Hubble sequence scheme. We can test the mechanisms which cause galaxies to reassemble and/or influence star-formation by tracking how morphologies change across time. This places vital constraints on the physical mechanisms assembling galaxies and quenching star-formation.

All magnitudes are quoted in the AB system. A standard Λ CDM cosmology of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.3$ is used throughout this work.

2.2 Data

The Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS, PIs: S. Faber and H. Ferguson; Grogin et al. 2011 and Koekemoer et al. 2011) observed 5 heavily studied fields (of which we use UDS, GOODS-S and COSMOS) with the *Hubble Space Telescope* (HST). High resolution imaging by Wide Field Camera 3 (WFC3) in near-infrared bands F125W (J) and F160W (H), combined with observations from the Advanced Camera for Surveys (ACS) in visible bands F606W (V) and F814W (I_w) constitute the new measurements in the CANDELS program. For the purposes of our study, we initially focus only on the F125W WFC3 images. Future work will study the evolution of galaxy morphology at a consistent rest-frame wavelengths.

We use the CANDELS H -band (F160W) selected multi-wavelength catalogs (UDS,

CHAPTER 2. PCA MORPHOLOGY

Galmetz et al., 2013; GOODS-S, Guo et al., 2013; COSMOS, Nayyeri et al., in prep), photometric redshifts (Dahlen et al., 2013), non-parametric morphologies (this work), Sérsic parameters (van der Wel et al., 2012), visual classifications (Kartaltepe et al., 2015), rest-frame photometry, and stellar masses (this work). The limiting magnitude for HST/WFC3 F125W and F160W are 27.35 and 27.45 respectively with FWHM of $\sim 0.135''$ and $\sim 0.15''$ respectively. Galmetz et al. (2013) outlined the techniques used to create the photometric catalogs.

The photometric redshift catalogs of Dahlen et al. (2013) are the combination of multiple different photometric redshift calculating codes and techniques which reduce the scatter of photometric redshifts (to $\sigma \sim 0.03$, with an outlier fraction of 3 percent). Throughout the rest of this paper, we use z to denote the average photometric redshift in these CANDELS catalogs (Mobasher et al., 2015).

Rest-frame $U - V - J$ colors were calculated by the sed-fitting code EAZY (Brammer et al., 2008), using the empirical local galaxy templates of Brown et al. (2014). Stellar masses were computed with FAST (Kriek et al., 2009), assuming Bruzual & Charlot (2003) delayed exponential star-formation histories, a Chabrier (2003) initial mass function, Calzetti et al. (2000) dust attenuation, and solar metallicities.

2.2.1 Sample Selection Criteria

We select bright ($H < 24.5$), massive ($M_* > 10^{10} M_\odot$) galaxies with $1.36 < z < 1.97$ galaxies measured in F125W (J). This band approximately corresponds to rest-

CHAPTER 2. PCA MORPHOLOGY

frame optical B -band at these redshifts. This redshift range provides a large sample of galaxies measured in a constant rest-frame waveband, and offers a high enough redshift to have a different morphological distribution from a local sample. At this redshift and magnitude, the CANDELS surveys are mass-complete down to $10^{10} M_{\odot}$ (Wuyts et al., 2011). In our sample of UDS, COSMOS and GOODS-S there are a total of 6269 galaxies with $H < 24.5$ and $M_* > 10^{10} M_{\odot}$. Of those galaxies 1539 are within our redshift range ($1.36 < z < 1.97$).

The following affect our sample completeness: high signal-to-noise (per pixel) measurements ($S/N > 4$), an internal morphology quality flag = 0, and a well measured concentration (i.e. $C \neq -99$) requirement. The quality flag requirement removes objects from the sample with discontiguous segmentation maps resulting from low surface brightness, and/or poor masking of bright neighbors. In §2.2.2 we include a brief discussion of galaxies with a quality flag = 1. The concentration requirement removes the contamination from poorly measured galaxies on the overall PCA. For some galaxies, r_{20} (and thus C) can not be accurately measured because either the object is too small, or there is a bright point source disrupting the light profile (see §2.3.1.2). The concentration requirement reduces the total of galaxies in the sample to 1482. The FLAG requirement reduces the sample to 1250. The signal-to-noise, FLAG and well measured concentration requirements together reduce our final sample to 1244 galaxies.

2.2.2 Galaxies with FLAG=1

Galaxies with non-contiguous segmentation maps receive a FLAG=1 designation. The disconnected segmentation maps could be the result of a few factors: the light of a nearby bright galaxy encroaching on a galaxy, low surface brightness or low signal-to-noise. For this reason their non-parametric morphology measurements are likely to be unreliable. Fig. 2.1 is the normalized histogram of magnitudes for galaxies with either FLAG=0 or FLAG=1. We also show the fraction of galaxies per magnitude bin. The number of galaxies with FLAG=1 galaxies as a fraction of all galaxies increases up to magnitude 24.5, which is the brightness limit of the survey. For these reasons we leave these galaxies out of our sample in this work, but we will investigate these galaxies in a future work.

2.3 Morphological Measurements

2.3.1 Non-parametric Morphology

We focus on non-parametric morphology statistics: concentration, asymmetry, Gini coefficient, M_{20} , along with three new statistics from Freeman et al. 2013: multi-mode, intensity and deviation. The code for calculating the morphological statistics (originally developed by Lotz et al. 2008) has been modified to include the new statistics and accommodate much larger input images. The code is applied to the

CHAPTER 2. PCA MORPHOLOGY

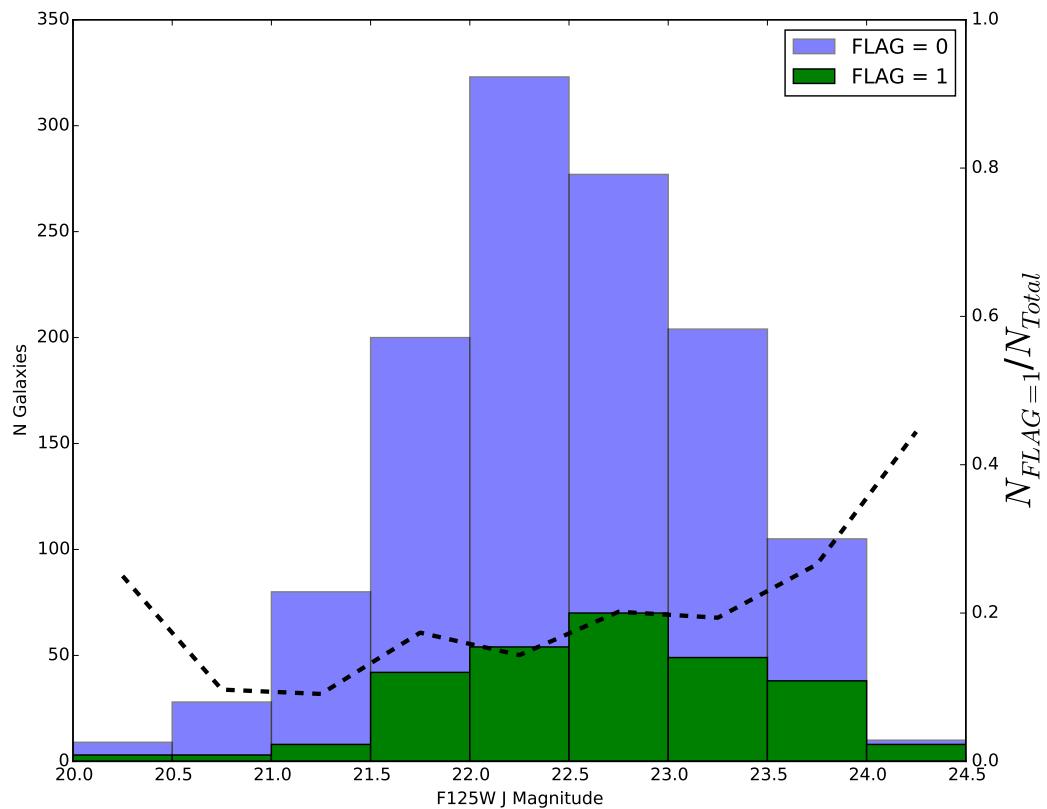


Figure 2.1 Histogram of F125W J-band Magnitude for galaxies with FLAG=1 and FLAG=0 and a plot of the fraction of all galaxies with FLAG = 1 designation per magnitude bin (*black dashed line*).

CHAPTER 2. PCA MORPHOLOGY

CANDELS F125W mosaics using the F160W detected catalogs and segmentation maps as the input.

2.3.1.1 Petrosian Radius

The Petrosian radius r_p is the radius we set to where the surface brightness μ is 20% of the mean interior surface brightness (Eq. 2.1; Petrosian, 1976). The Petrosian radius is more robust to surface brightness dimming than isophotal sizes are. We can measure the same physical portions for galaxies at a variety of redshifts (e.g. Lotz et al., 2004).

$$0.2 = \frac{\mu(r_p)}{\bar{\mu}(r < r_p)} \quad (2.1)$$

2.3.1.2 Concentration

The concentration index (C ; Bershady et al., 2000; Conselice et al., 2003) is the ratio of the circular radius containing 80% (r_{80}) of a galaxy's light (as measured within 1.5 Petrosian radii) to the radius containing 20% (r_{20}) of the light (Eq. 2.2). A large concentration value indicates a majority of light is concentrated at the center of the galaxy. Elliptical galaxies and bulge-dominated spirals have high concentration values. However, a spiral or irregular galaxy with diffuse light profile and weak/no bulge will have low concentration values.

$$C = 5 \log \left(\frac{r_{80}}{r_{20}} \right) \quad (2.2)$$

For some galaxies r_{20} (and thus C) can not be accurately measured because either the object is too small, or there is a bright point source disrupting the light profile. These galaxies instead have unphysical concentration values ($C < 0$) and are not included in the definition of our principal components (see §2.2.1).

2.3.1.3 Asymmetry

Asymmetry (A ; Conselice et al., 2000) measures the difference between the image of a galaxy ($I_{x,y}$) and the galaxy rotated by 180 degrees ($I_{180(x,y)}$; Eq. 2.3). This determines a ratio of the amount of light distributed symmetrically to all light from the galaxy. A is calculated from a sum of all pixels within 1.5 Petrosian radii from the center of the galaxy. We then correct by B_{180} , which is the average asymmetry of the background. An initial guess for the center of rotation is defined by the physical center, but is updated through an iterative process. This process continues until a global minimum value for A is found (Conselice et al., 2000).

$$A = \frac{\sum_{x,y} |I_{(x,y)} - I_{180(x,y)}|}{2 \sum |I_{(x,y)}|} - B_{180} \quad (2.3)$$

Due to their uniform morphologies and lack of structure elliptical galaxies typically have small asymmetry values ($A \sim 0.02$). Meanwhile spiral galaxies usually have values between $A \sim 0.07$ to 0.2 (Conselice, 2014). This statistic is most useful

CHAPTER 2. PCA MORPHOLOGY

for identifying irregular galaxies because they appear lopsided or ragged. Visually inspected merger remnants can have $A \gtrsim 0.3$ (Conselice et al., 2003). The asymmetry statistic is more sensitive to gas-rich mergers than to gas-poor or minor mergers (Lotz et al., 2010a,b).

If the local background is high and the galaxy is sufficiently low surface brightness then negative A values are measured. This is consistent with measurement errors (see §2.4.2).

2.3.1.4 Gini Coefficient

The Gini coefficient (G ; Lorenz, 1905; Abraham et al., 2003; Lotz et al., 2004) is a statistic adapted from economics that measures the equality of light distribution in a galaxy. The Gini coefficient is defined by the Lorenz curve of the galaxy's light distribution, and is not affected by spatial position. This implies that only the amount of light distribution matters, which differentiates the Gini coefficient from the concentration statistic (Conselice, 2014).

The pixels are ranked by increasing flux value, then G is determined by Eq. 2.4, where n is the number of pixels in the galaxy's segmentation map, X_i is the pixel flux at the rank i pixel and \bar{X} is the mean pixel value.

$$G = \frac{1}{\bar{X}n(n-1)} \sum_i^n (2i - n - 1)X_i \quad (2.4)$$

A galaxy with equally distributed light will have a Gini coefficient approaching 0.

CHAPTER 2. PCA MORPHOLOGY

Conversely, a galaxy with a large fraction of light concentrated on a few pixels will have a Gini coefficient closer to 1. Elliptical galaxies and galaxies with bright nuclei have high Gini coefficients, while disks and galaxies with a uniform surface brightness will have low Gini coefficients.

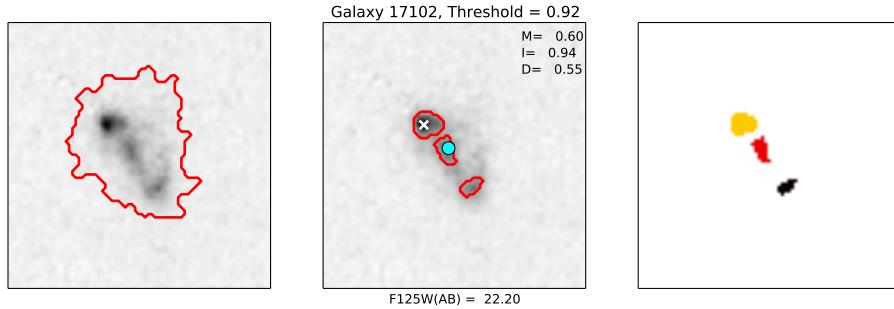


Figure 2.2 $F125W$ (AB) = 22.2 CANDELS galaxy image is shown to demonstrate the M , I and D statistics. The left panel shows the image of the galaxy outlined by the segmentation map created using our morphology code. The middle panel shows red outlines describing the clumps found when calculating the M statistic. The white X displays the location of the brightness distribution peak, and the cyan circle represents the location of the intensity centroid used to calculate the D statistic (§2.3.1.8). The right panel color codes the clumps for easy identification. This galaxy is highly disturbed and is broken into 3 bright regions, with the brightness peak well separated from the intensity centroid. The threshold value (q_l) in this case is 0.92, which represents the threshold where the M statistic was maximized.

CHAPTER 2. PCA MORPHOLOGY

2.3.1.5 M_{20}

The second order moment of the brightest regions of a galaxy (M_{20} ; Lotz et al., 2004) traces the spatial distribution of any bright clumps. When used in tandem with the Gini coefficient, M_{20} can be an effective tool for differentiating galaxies with bright off-center clumps (such as irregular galaxies) from those with one bright central region (such as the bulge of a spiral galaxy). We define the regions representing the brightest 20% of the galaxy (Eq. 2.5), and then calculate the spatial distribution of those pixels as an offset from the central pixel. The center is defined as the position minimizing M_{tot} .

$$\sum_i f_i < 0.2f_{tot} \quad (2.5)$$

$$M_{tot} = \sum_i^n M_i = \sum_i^n f_i [(x_i - x_c)^2 + (y_i - y_c)^2] \quad (2.6)$$

Finally we calculate the second order moment (Eq. 2.7).

$$M_{20} = \log \left(\frac{\sum_i M_i}{M_{tot}} \right) \quad (2.7)$$

Values for the M_{20} statistic are generally between -0.5 and -2.5. Elliptical galaxies have M_{20} closer to -2.5 signifying a lack of bright-off center clumps. Meanwhile disk galaxies can have $M_{20} > -1.6$ when, for example, bright star-forming knots are present. Similar to concentration, M_{20} is biased low for galaxies where the brightest 20% light is unresolved.

2.3.1.6 Multi-mode

The multi-mode (M) statistic is the ratio, in pixels, of the two brightest regions of a galaxy (adapted from Freeman et al., 2013). Bright regions are determined via a threshold method where q_l represents the normalized flux value, and $l\%$ of pixel fluxes are less than q_l . This creates a new binary image $g_{i,j}$ where 1 represents fluxes larger than q_l and 0 represents fluxes less than q_l (Eq. 2.8).

$$g_{i,j} = \begin{cases} 1 & f_{i,j} \geq q_l \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

The number of pixels in contiguous groups of pixels with value 1 are then sorted in descending order by area. The 2 largest groups ($A_{l,(2)}$ and $A_{l,(1)}$) define an area ratio R_l :

$$R_l = \frac{A_{l,(2)}}{A_{l,(1)}} \quad (2.9)$$

The previous two steps are recomputed for various normalized flux levels l , and the M statistic is the maximum R_l value (Eq. 2.10). Values approaching 1 represent multiple nuclei, while values near 0 are single nuclei systems.

$$M = \max R_l \quad (2.10)$$

This formulation is slightly revised from Freeman et al. (2013) to limit the M statistic to values between 0 and 1. Freeman et al. (2013) multiplies Eq. 2.9 by

CHAPTER 2. PCA MORPHOLOGY

an additional factor of $A_{l,(2)}$ to limit the effect of hot pixels. However, this adds a size dependent factor to the calculation. Because we wish to measure M values for galaxies at a variety of angular distance scales, it is important to have a size independent measure. For illustrative purposes, Fig. 2.2 shows an example of how the MID statistics are calculated. In small galaxies that are poorly resolved $A_{l,(1)}$ is very small (approaching zero) and we set $M=-99$. We have tested the result of setting $M=-99$ values to $M=0$ but find the PC weights and group assignments are very similar to the original values.

2.3.1.7 Intensity

Intensity (I) is the ratio, in flux, of the two brightest regions (Freeman et al., 2013). The galaxy image is first smoothed by a symmetric bivariate Gaussian kernel. Regions are defined using maximum gradient paths, where the surrounding eight pixels of every pixel are inspected and the path of maximal intensity increase is followed until a local maximum is reached. Regions consist of pixels linked to a unifying local maximum. The fluxes within these groups are summed and sorted into descending order (by total flux) leading to our intensity ratio:

$$I = \frac{I_{(2)}}{I_{(1)}} \quad (2.11)$$

Similar to the M statistic, elliptical galaxies with a bright bulge have $I \sim 0$, while disk galaxies with bright clusters of star-formation are more likely to have I values

approaching 1.

2.3.1.8 Deviation

Deviation (D) measures the distance between the intensity centroid of a galaxy and the center of the brightest region (Freeman et al., 2013, Eq. 2.12 and Eq.2.13). Disk and spheroidal galaxies have deviation values near 0 because their central bulges typically possess the brightest pixels. On the other hand, a high deviation value indicates a galaxy has bright star forming knots significantly separated from the intensity centroid (e.g. Fig. 2.2).

$$(x_{cen}, y_{cen}) = \left(\frac{1}{n_{seg}} \sum_i \sum_j i f_{i,j}, \frac{1}{n_{seg}} \sum_i \sum_j j f_{i,j} \right) \quad (2.12)$$

The deviation statistic D is the Euclidean distance (in pixels) between the intensity centroid and brightest pixel scaled by a crude estimate of a galaxy's radius based upon the number of pixels comprising the galaxy.

$$D = \sqrt{\frac{\pi}{n_{seg}}} \sqrt{(x_{cen} - x_{l_{(1)}})^2 + (y_{cen} - y_{l_{(1)}})^2} \quad (2.13)$$

2.3.2 Morphological Principal Components

Principal component analysis (PCA) is a linear transformation of multivariate data. This defines a set of uncorrelated axes, called principal components (PCs), which are ranked by the variance they capture (Pearson, 1901; Ivezić et al., 2013). A

CHAPTER 2. PCA MORPHOLOGY

linear combination of the original data and eigenvector solutions (also called weights) project the original data on to the PCs. Principal component analysis is a simple way to reduce the dimensionality and find the natural distributions of data in parameter space. PCA is able to determine the correlations between the input data and can find relationships missed by other means.

We begin by “whitening” the data, i.e. we subtract the mean of each morphological measurement and divide by the standard deviation of each feature. By dividing our data by feature variance we remove the effects of mixed units. We calculate the singular value decomposition ($x_{ij} = V\Sigma V^T$, SVD) of the “whitened” data matrix (\mathbf{x}_{ij}). An SVD decomposes the original data into a diagonal matrix (Σ) containing eigenvalues (e) and a non-diagonal matrix \mathbf{V} containing the expansion coefficients (aka weights). The eigenvalues determine how important each principal component is to explaining the original data set. The eigenvectors are rank ordered by their associated eigenvalue. We then project our “whitened” data onto our new eigenbasis to calculate the principal component scores, which inform us how similar are data points to each other (PC_i , Eq. 2.14).

$$PC_i = \sum_{j=1}^N V_{ji}x_j (i = 1, \dots, N) \quad (2.14)$$

Table 2.1 shows the correlations and importance of different statistics across the eigenvector solutions of the principal component analysis. The scree value ($e^2 / \sum e^2$) represents the amount of variance captured by a single principal component. The

CHAPTER 2. PCA MORPHOLOGY

scree values demonstrate that the first 3 PCs account for >75% of the variance in the data. The fact that PC1 only accounts for 40% of the variance shows that more than a single parameter is needed to define a galaxy. The error estimates are the result of the scattering method described in §2.4.2.

PC1 is highly dependent upon M , I , D , M_{20} and the Gini coefficient. We interpret PC1 as a “bulge strength” indicator given the correlation with $G - M_{20}$ and the importance of the MID statistics. Fig. 2.5 shows the relationship between PC1, Sérsic index and the Gini- M_{20} “bulge strength” (Eq. 2.15 and 2.16) the vector of correlations between Gini and M_{20} ; Snyder et al., 2015b). Galaxies with low PC1 values have high Sérsic indices and high F indicative of strong bulges, while galaxies with higher PC1 values have progressively smaller bulges and more prevalent disc properties (see §2.5 for more on the physical and visual properties of specific groups). We observe two correlations between F and PC1 which corresponds to different groups of galaxies. Additionally, the two parallel stripes of data seen in Fig. 2.5 are the result of $M=-99$ outlier values shifting PC1. We have tested the result of setting $M=-99$ values to $M=0$ and find that the PC eigenweights and the group classifications are very similar to our original values.

$$F = -0.693M_{20} + 4.95G - 3.85 \quad (2.15)$$

$$F(G, M_{20}) = \begin{cases} |F| & G \geq 0.14M_{20} + 0.778 \\ -|F| & G < 0.14M_{20} + 0.778 \end{cases} \quad (2.16)$$

PC2 is highly dependent upon concentration, and is larger for galaxies with bright centers and extended envelopes. PC3 is dominated by asymmetry and is larger for disturbed galaxies. The other principal components are harder to interpret, but are also less important as evidenced by their lower scree values. It is interesting to note PC1 defines a bulge strength but is not dependent on concentration (Eq. 2.2). Concentration for very small ($r_e < 2$ kpc), high Sérsic ($n > 2.5$) galaxies is strongly biased down (see § 2.3.3). This bias is potentially important for $\sim 14\%$ of our sample.

We performed tests on how PCA results are affected by whitening the data set using the interquartile range (IQR) statistic instead of a standard deviation. The eigenvectors calculated using either whitening method are mainly consistent. However, we chose to use the standard deviation to whiten our data because the PC weights are more volatile when calculated with an IQR whitened data set. In particular, the weight in PC3 describing concentration has a variance nearly nine times larger when calculated for an IQR-scaled data set compared to a standard deviation-scaled data set.

2.3.3 Concentration - Sérsic Index Relationship

Andrae et al. (2011) demonstrated the correlation between concentration and

CHAPTER 2. PCA MORPHOLOGY

Table 2.1 PC Weights with error estimates based on a bootstrap scattering method

Parameter	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Scree value	0.41	0.19	0.15	0.08	0.06	0.06	0.05
Concentration	-0.06 ± 0.02	0.74 ± 0.01	-0.35 ± 0.03	0.19 ± 0.04	-0.31 ± 0.11	0.03 ± 0.12	-0.43 ± 0.07
M ₂₀	0.48 ± <0.01	-0.03 ± 0.02	-0.12 ± 0.02	0.16 ± 0.07	-0.67 ± 0.19	0.07 ± 0.19	0.52 ± 0.09
Gini	-0.45 ± 0.01	0.27 ± 0.02	0.12 ± 0.02	0.45 ± 0.05	0.11 ± 0.16	-0.46 ± 0.13	0.53 ± 0.07
Asymmetry	0.00 ± <0.01	0.41 ± 0.03	0.82 ± 0.02	-0.31 ± 0.03	-0.18 ± 0.05	0.18 ± 0.05	0.06 ± 0.03
Multi-mode	0.38 ± <0.01	0.45 ± 0.02	-0.27 ± 0.02	-0.30 ± 0.07	0.56 ± 0.11	0.14 ± 0.15	0.40 ± 0.07
Intensity	0.49 ± <0.01	0.04 ± 0.01	0.13 ± 0.01	-0.13 ± 0.03	0.02 ± 0.10	-0.82 ± 0.15	-0.24 ± 0.07
Deviation	0.43 ± <0.01	0.00 ± 0.01	0.30 ± 0.01	0.73 ± 0.04	0.31 ± 0.15	0.25 ± 0.10	-0.18 ± 0.06

CHAPTER 2. PCA MORPHOLOGY

Sérsic- n . However, this relationship does not appear to hold for our high redshift sample. Fig. 2.3 shows a less established relationship for concentration and Sérsic- n in our galaxy sample. We show that concentration is biased low for very small ($r_e < 2$ kpc), high Sérsic n galaxies ($n > 2.5$) which represents $\sim 14\%$ of our sample. We also find many $z \sim 1.5$ galaxies with high concentration and low Sérsic- n that deviate from the Andrae et al. (2011) relation and are not easily explained by measurement bias.

The PSF for F125W has a full width half-maximum (FWHM) of $\sim 0.135''$. For many galaxies, r_{20} is smaller than the PSF (and in some cases even r_e is smaller than the PSF). Fig. 2.3 shows that high Sérsic galaxies make up some of the smallest objects in our sample. These small galaxies can have $r_{80} \sim 0.48''$, which is only a few times larger than the PSF.

We wish to test the effect of the size of the PSF can have on measuring the concentration index, particularly for small galaxies. To accomplish this we take a pure Sérsic surface brightness light profile $I \sim \exp[(r/r_e)^{1/n}]$ with $r_e = 10$ kpc and calculate the Petrosian Radius (Eq. 2.1), r_{80} , r_{20} and thus concentration. We convolve the pure Sérsic profile with a gaussian with the same FWHM as the PSF. This convolution has little effect on the concentration for large galaxies. However, we noticed in Fig. 2.3 that many galaxies have very small r_e values which could lead to why concentration values are lower than anticipated. To test this hypothesis we convolve the Sérsic surface brightness profile of a small galaxy ($r_e = 1$ kpc and 2 kpc) with a gaussian

CHAPTER 2. PCA MORPHOLOGY

with the FWHM of the PSF. This will allow us to observe the effect of convolving the surface brightness profile of a small galaxy with a PSF of comparable size.

Fig. 2.4 shows the concentration - Sérsic relation present in our galaxy sample and is color coded by the ratio of the size of the PSF to the effective radius of a galaxy. The solid red line in Fig. 2.4 shows the relation between concentration and Sérsic calculated for a pure Sérsic surface brightness profile with $r_e = 10$ kpc (first demonstrated in Andrae et al. 2011). The thin-thick and thick dashed lines in Fig. 2.4 show the concentration - Sérsic relation for a surface brightness profile (of a $r_e = 1$ kpc or 2 kpc galaxy) convolved with a gaussian with the FWHM of the PSF. Galaxies with high FWHM/R_e ratios (i.e. the galaxy has a comparable physical size to the PSF) fall noticeably below the concentration-Sérsic relation for a pure Sérsic surface brightness profile. The flatter concentration-Sérsic relation of the small galaxy surface brightness profiles convolved with the PSF closely follows the concentration and Sérsic values we measure for our sample. As the physical size of a galaxy decreases the concentration values are increasingly depressed. We take this as evidence that small galaxies (those with physical sizes similar to the PSF, $r_e \sim 1\text{-}2$ kpc) are most affected by the PSF. Thus the reason the concentration values for our galaxies are smaller than the relation of Andrae et al. (2011) is likely due to the small physical sizes of many galaxies in our sample.

Up to 14% of our total sample maybe be quite small (roughly the size of the PSF, $r_e < 2$ kpc) and have a high Sérsic index ($n > 2.5$) leading to an artificially depressed

CHAPTER 2. PCA MORPHOLOGY

concentration value. Many of these galaxies ($\sim 80\%$) are in group 6. After correcting the concentration values these galaxies would instead be classified into group 0. This suggests that a portion of the group 6 galaxies would instead be group 0 if we had higher resolution images. However, this implies only $\sim 26\%$ of all group 6 galaxies would be reclassified as group 0 so there is still a notable distinction between these two groups.

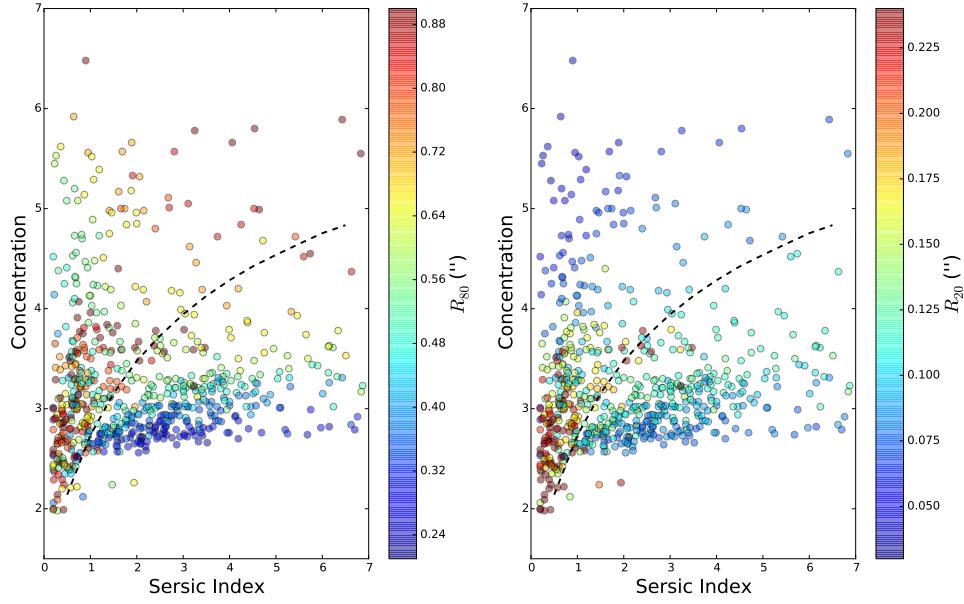


Figure 2.3 WFC3 125W measured concentration versus F125W Sérsic index (van der Wel et al., 2012) color coded by (*left panel*) R_{80} and (*right panel*) R_{20} for the entire sample. Our $z \sim 1.5$ galaxies generally follow a shallower relation than previously shown in Andrae et al. 2011 (*black dashed line*).

CHAPTER 2. PCA MORPHOLOGY

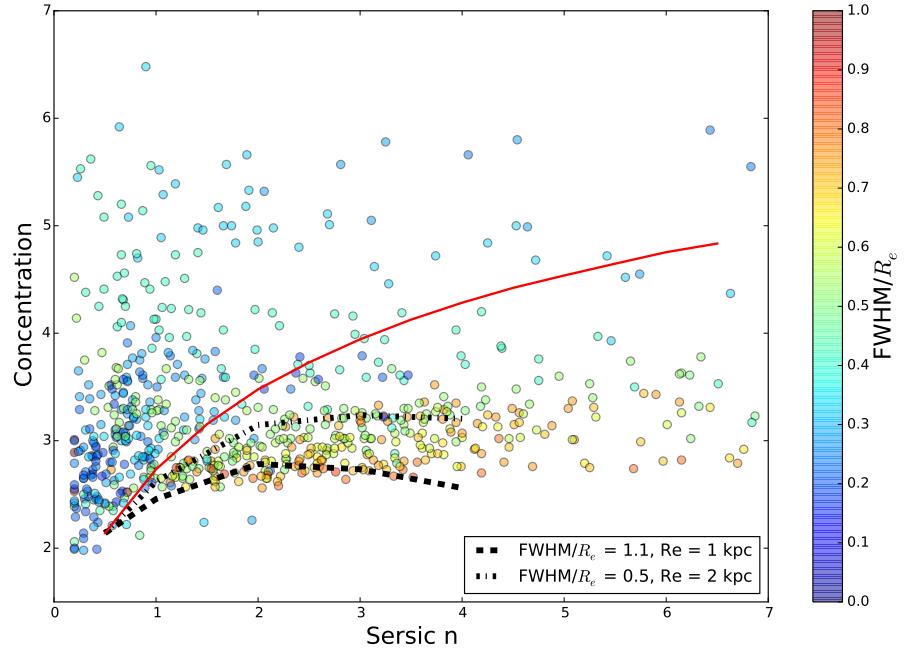


Figure 2.4 Concentration versus Sérsic index relation color-coded by the ratio of the PSF FWHM ($\sim 0.135''$) to the effective radius of a galaxy. We plot the numerically defined relationship for a pure Sersic profile (Andrae et al., 2011., *red line*) and the corrections to the pure Sersic profile when the PSF FWHM is 50% the size of a 2 kpc galaxy (*thick-thin dashed line*) and when the PSF is as large as a 1 kpc galaxy (*thick dashed line*). The relative size of the PSF to a galaxy has a large impact on the concentration values for galaxies with higher Sérsic indices.

CHAPTER 2. PCA MORPHOLOGY

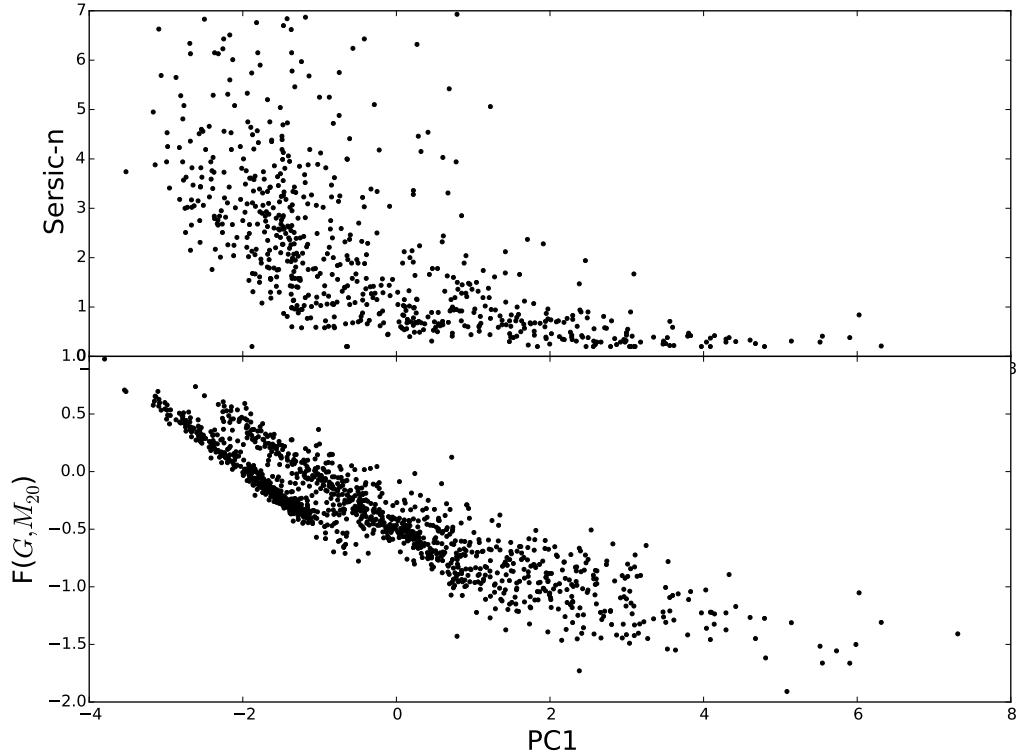


Figure 2.5 PC1 v. Sérsic Index and PC1 v. the Gini- M_{20} bulge strength metric (F , Snyder et al., 2015b). PC1 is anti-correlated with Sérsic index and the Gini- M_{20} Bulge factor, F , and thus low PC1 values are indicative of a strong central bulge. Small galaxies can have $M = -99$ which shifts PC1 and leads to the two parallel stripes. See §2.5 for more on how group 6 galaxies are different from the remainder of the sample.

CHAPTER 2. PCA MORPHOLOGY

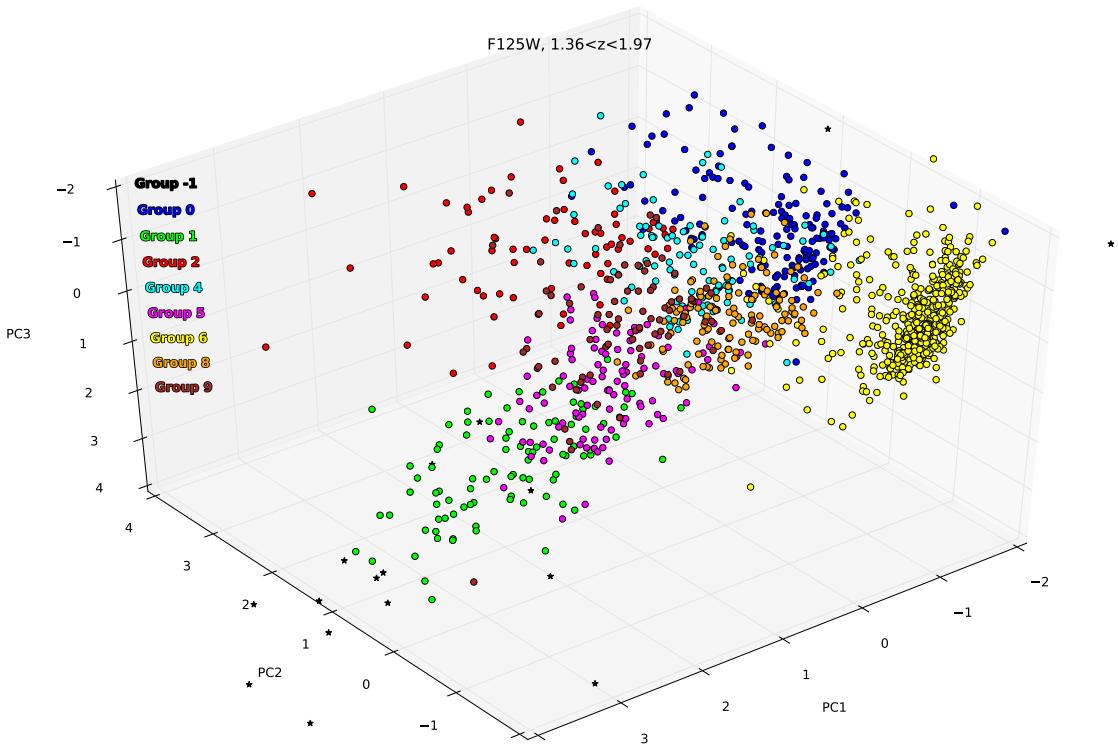


Figure 2.6 PC1 v. PC2 v. PC3 for our sample of $M_* > 10^{10} M_\odot$, $1.36 < z < 1.97$ galaxies, color-coded by their hierarchical cluster definitions. PC1 anti-correlates with bulge strength, PC2 is dominated by concentration, and PC3 is dominated by asymmetry (see Table 2.1). Group -1 galaxies (*black stars*) are outliers from remaining groups, initially they comprised groups 3 and 7.

CHAPTER 2. PCA MORPHOLOGY

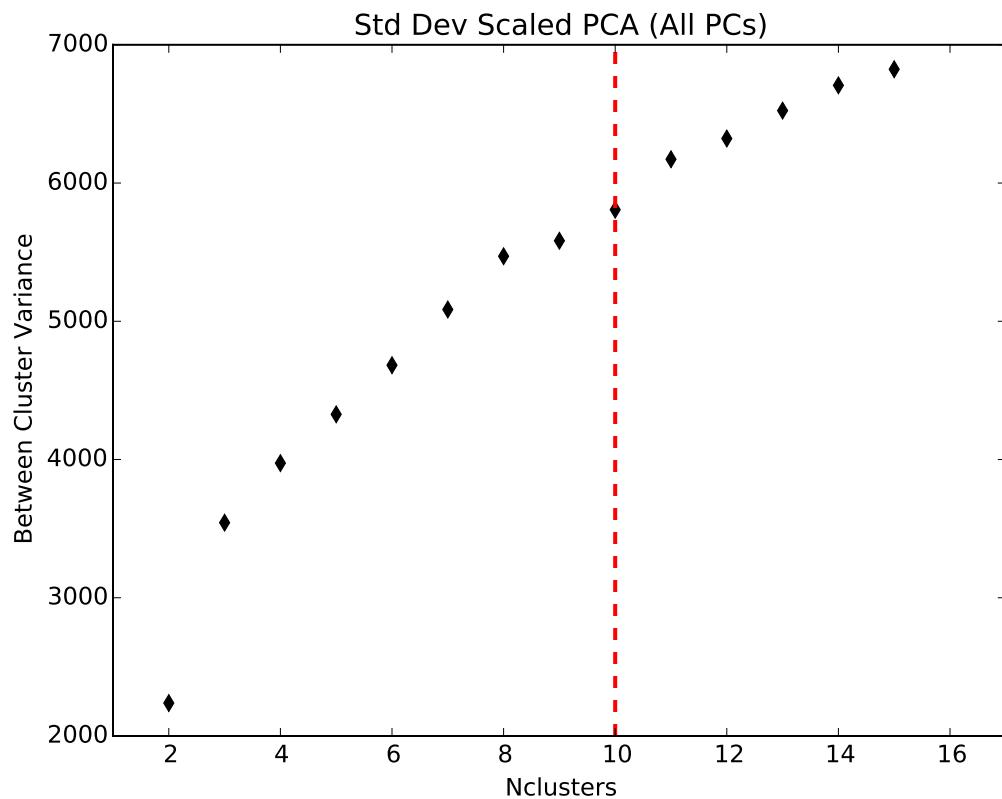


Figure 2.7 The amount of between-cluster variance as a function of the number of clusters grouped by the Ward hierarchical agglomerative clustering routine. The between-cluster variance is the sum of the distances from the centroid of each cluster to the centroid of all the data. Eventually this value grows to the total variance in the data when the number of clusters equals the number of data points.

CHAPTER 2. PCA MORPHOLOGY

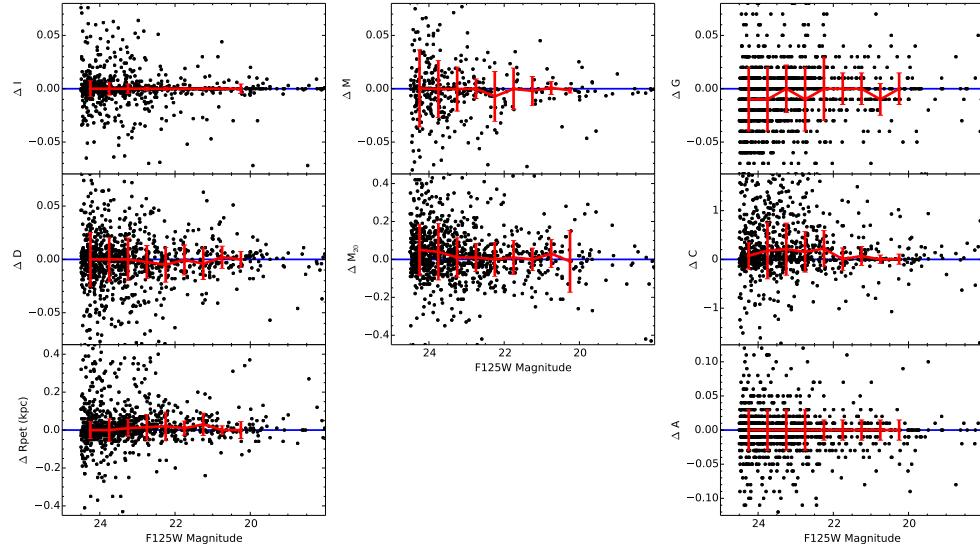


Figure 2.8 Magnitude and Surface Brightness vs. $\Delta[\text{GOODS} - \text{UDF}]$ morphological statistics as measured in wide-field imaging of GOODS-S compared to the deep imaging of UDF. Red error bars represent the median Δ morphology value binned in magnitude (or surface brightness) bins of 0.5. Error bars represent the median absolute deviation of each bin corresponding to a 1σ deviation.

CHAPTER 2. PCA MORPHOLOGY

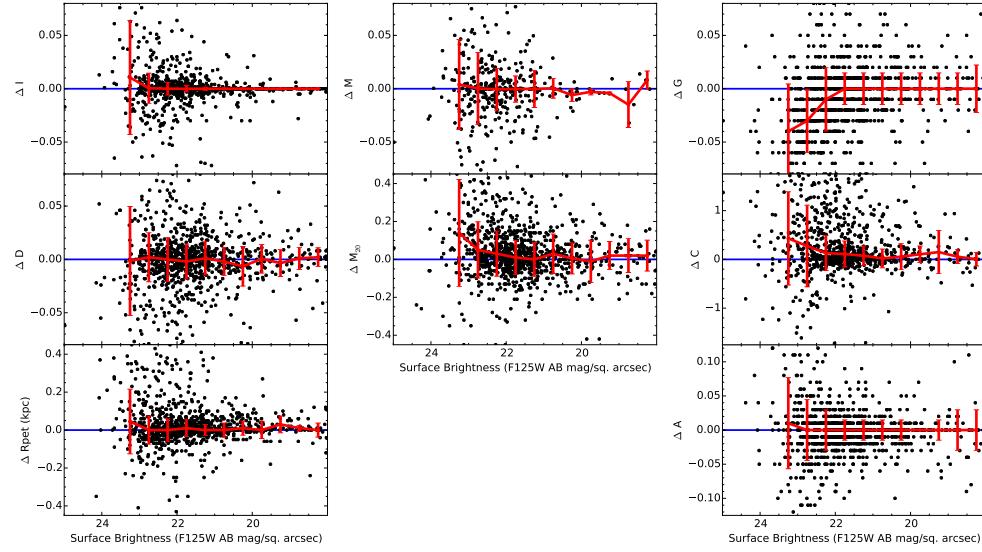


Figure 2.9 Magnitude and Surface Brightness vs. Δ [GOODS - UDF] morphological statistics as measured in wide-field imaging of GOODS-S compared to the deep imaging of UDF. Red error bars represent the median Δ morphology value binned in magnitude (or surface brightness) bins of 0.5. Error bars represent the median absolute deviation of each bin corresponding to a 1σ deviation.

2.4 PCA-Morphology Group Properties

2.4.1 Defining PCA morphology groups

Studies using PCA usually only select the top eigenvectors to represent the data. However, this is not a requirement of the analysis. In our case, the number of variables is not very large and thus retaining the entire parameter space is not a computationally expensive procedure. We aim to reconstruct the full set of galaxy morphological correlations at other redshift ranges by using all 7 PC dimensions to represent the data set. The correlations from the higher PC eigenvectors may be important at different redshifts. When the goal of PCA is to cluster data then reducing the number of features based on the amount of variance captured is not the only option (Jolliffe, 1986; Ben-Hur & Guyon, 2003). In these cases more principal components can better recreate the original data set.

The morphologies of galaxies are not inherently discrete, but rather lie on a continuum. However, it is often useful to bin galaxies into discrete morphological groups. Fig. 2.6 shows the distribution of galaxies when projected onto the first three principal axes. Except for a large distinct cluster of data points, most of our sample are not well separated, requiring the need for an objective data dependent grouping method.

To classify galaxies in distinct groups, we employ the Ward hierarchical agglomerative clustering routine of `scikit-learn` (Pedregosa et al., 2011). Hierarchical clustering (specifically agglomerative clustering) treats each galaxy as its own cluster,

CHAPTER 2. PCA MORPHOLOGY

which are then merged with nearby clusters while minimally increasing the in-cluster variance. Mergers of adjacent clusters continue until the desired number of groupings are attained. We define 10 groups, 2 of which are very sparsely populated, with only a combined 12 galaxies. The sparsely populated clusters consist of extreme outliers from the other 8 clusters. For this reason, we group all outliers into a single cluster.

Fig. 2.7 shows the amount of between-cluster variance calculated for various numbers of clusters. Typically, the optimal number of clusters chosen corresponds to the turnover in this distribution (the point where the increase in between-cluster variance begins to slow; Everitt & Hothorn, 2006) which occurs at ~ 10 clusters. Increasing the number of clusters any further does not provide any more discriminatory power and only complicates interpretations of the final results. We must note that there is no definitive criterion to help define how many clusters are to be defined in the data set.

The hierarchical clustering algorithm defines the groups based on the distribution of the data. In order to reproduce the same group definitions for new objects with potentially different distributions (e.g. different redshifts), we use a convex hull method to define the original group boundaries in principal component space. A convex hull defines the smallest area containing a set of points. We define convex hulls using the 10 clusters determined by Ward's method for our $z \sim 1.5$ galaxy sample. In practice we disregard the 2 sparsely populated clusters and instead group all of those galaxies into the outlier class.

CHAPTER 2. PCA MORPHOLOGY

We use all 7 PCs to define the convex hull. Calculating convex hulls in 7-dimensional space is computationally intensive and currently impossible for large data sets, thus we outline a simple workaround. We define a convex hull based on 2 PC dimensions at a time and test whether a galaxy falls within the boundaries of a group using all combinations of 2 PC dimensions. The group a galaxy falls in the most times is determined to be its group. If more than one group is equally likely, the smallest distance from the galaxy's position in PC space to the center of the possible groups is used to determine group membership. Galaxies that are misclassified following the convex hull method generally exist on the boundaries of a convex hull. We present the PYTHON code determining the group membership based on convex hull groupings¹.

2.4.2 Morphological Error Estimation

The *Hubble Ultra Deep Field* (UDF) consists of deep imaging on a portion of the shallower GOODS-S field (Koekemoer et al., 2011). We measure the same galaxies using different depth images to the determine reliability of morphological measurements as a function of signal-to-noise and magnitude. The non-parametric morphologies of galaxies are measured both in the deep UDF region and the GOODS-S observations. We calculate the differences of GOODS-S morphologies from UDF morphologies. We then bin galaxies in magnitude (or surface brightness) to find the average difference

¹<https://github.com/mikepeth/PyML>

CHAPTER 2. PCA MORPHOLOGY

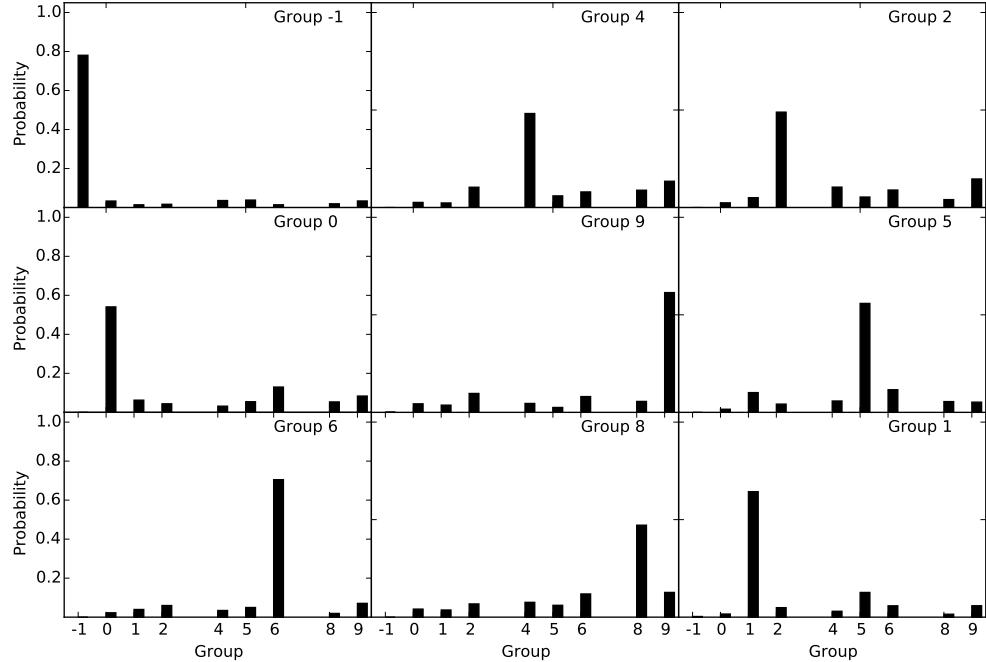


Figure 2.10 Group classification uncertainty, based on bootstrapped morphology measurement errors. Each galaxy's non-parametric morphologies are randomly scattered based on gaussians with widths based on errors found in Figs. 2.8 and 2.9 . The principal components and group membership are redetermined 250 times. The resulting MC group distributions for each originally defined group are shown. Groups 1, 6, and -1 are the most robust to measurement errors, whereas half of Groups 2, 4, 5 and 8 galaxies are scattered into other groups. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

CHAPTER 2. PCA MORPHOLOGY

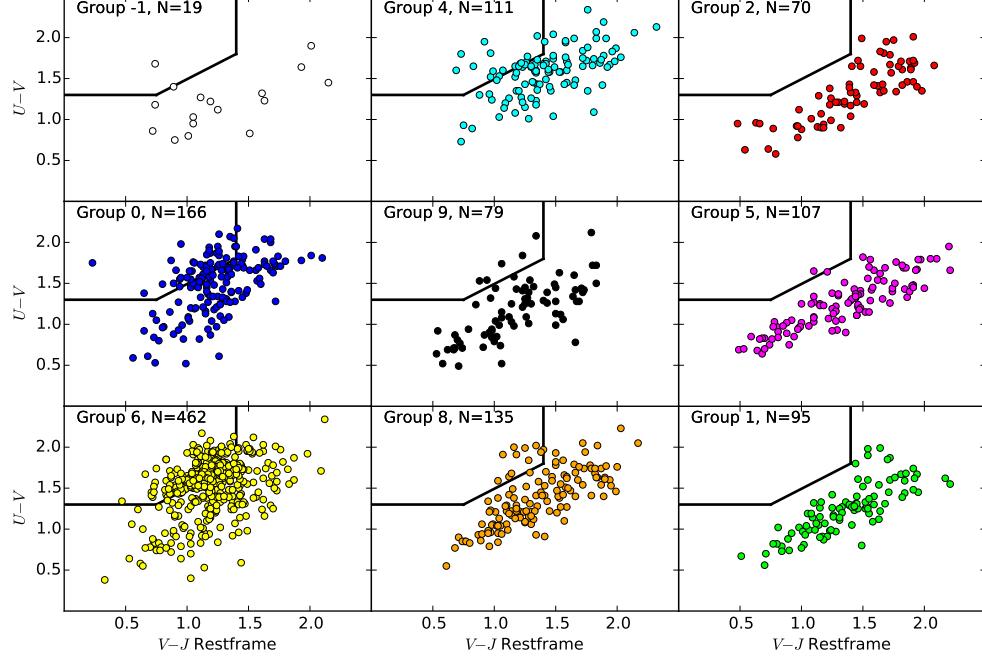


Figure 2.11 Rest-frame UVJ diagram for $M_* > 10^{10} M_\odot$, $1.36 < z < 1.97$ galaxies for each group. A UVJ diagram is used to separate quenched galaxies from star-forming galaxies (Williams et al., 2009). Quenched galaxies reside in the upper left trapezoid. Star-forming galaxies follow a sequence of increasing dust from the bottom left to the upper right. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1). The majority of quenched galaxies are in group 6, with some quenched galaxies found in groups 0 and 8. As PC1 increases we observe a decrease in the fraction of quenched galaxies.

CHAPTER 2. PCA MORPHOLOGY

and median absolute deviation, which we define as the error for that morphological measurement.

Figs. 2.8 and 2.9 show that larger and brighter galaxies are (unsurprisingly) well measured morphologically. The median absolute deviations (red error bars) show a majority of galaxies have statistics that do not vary widely between shallow and deep images. In general, the morphological offsets seen in Figs. 2.8 and 2.9 are very small. (For similar study see Fig. 6 in Grogan et al., 2011.)

Now that we have calculated the principal morphological components and resulting morphology groups, we can test their robustness to measurement errors. We use Monte Carlo resampling test to randomly scatter our initial morphological measurements by Gaussians with sigma equal to the median absolute deviation for each morphological measurement (Figs. 2.8 and 2.9). We then perform a principal component analysis for this new data set and repeat this process 250 times. We project the scattered data on the original PC weights and then classify the galaxies based on the originally defined convex hulls (§2.4.1) each time. The group with a plurality of the reclassifications is defined to be the “Monte Carlo” (MC) group. Fig. 2.10 can thus be seen as the probability distribution function for a galaxy of a certain group to be classified into a group via the convex hull method. Group 6 is the most robustly classified group. Only group 4 galaxies are reclassified as such following the Monte Carlo scattering to less than a majority of times (however still a large plurality of times). The plots are separated by group and ordered roughly by PC1 horizontally

CHAPTER 2. PCA MORPHOLOGY

and PC2 vertically. The largest PC1 values and smallest PC2 values are in group 6 galaxies.

Table 2.1 shows that the most important principal components (PC1-3) have typical resampled deviations $\leq 10 - 15\%$ of their weights. Higher principal component dimension display greater variability, but are also less important to our group classifications.

Every galaxy has an MC reclassification with a probability associated with it and the group with a plurality of the reclassifications is defined as the MC resampled group. Regardless of the probability, the reclassification is either the same or different from the original group designation. This similarity or difference in classification determines the completeness and purity of the classification scheme. The MC resampled classifications are 90.8% complete and 90.4% pure relative to original group classifications. The completeness and purity are highest when all 7 PCs are used to define the groups instead of only 3 PCs. Representing the data set with 3 PCs slightly drops the completeness and purity scores to 88.3 and 89.4%. In contrast, the completeness and purity values significantly drop to 25.9 and 20.3 percent when PCs are calculated from IQR-scaled data. The volatility of the eigenvectors calculated from an IQR-scaled data set is the cause of these poor reclassification results. This evidence leads to our conclusion that using all 7 standard-deviation scaled PC eigenvectors will result in more definitive groups.

Note that this does not include the systematic biases e.g. those due to the PSF.

CHAPTER 2. PCA MORPHOLOGY

This bias likely important for groups 6 and 0.

2.5 PCA Morphology Groups at $z \sim 1.5$

The connection between morphology and star-formation has been well studied (Wuyts et al., 2011; Kriek et al., 2009; Brinchmann et al., 2004). Late-type galaxies are typically still actively forming stars, whereas early-type galaxies have had their star-formation quenched. However, there are examples of red, quenched disks and blue, star-forming ellipticals which are important rare “transitional” classes.

We use a UVJ color-color diagram (Fig. 2.11) to classify galaxies as “star-forming” and “quenched” using the bimodality of these two types of galaxies seen in $U - V$ and $V - J$ rest-frame colors (Labbé et al., 2005; Wuyts et al., 2007; Williams et al., 2009). Star-forming galaxies follow a sequence determined by dust extinction. The panels are arranged in Fig. 2.11 so that PC1 increases along the x-axis and PC2 increases along the y-axis. Most groups are primarily comprised of star-forming galaxies. Groups with lower PC1 values are more compact and quenched. Similarly, a UV -Mass diagram separate star-forming from quenched galaxies (Fig. 2.12). Again galaxies with lower PC1 values are more massive and more quenched.

Previous studies (e.g. Lotz et al., 2004; Conselice et al., 2000; Lee et al., 2013) utilize $G - M_{20}$ (Fig. 2.13) or Concentration-Asymmetry (Fig. 2.14) diagrams to classify galaxies into early and late-type categories. In our study we use these tools to

CHAPTER 2. PCA MORPHOLOGY

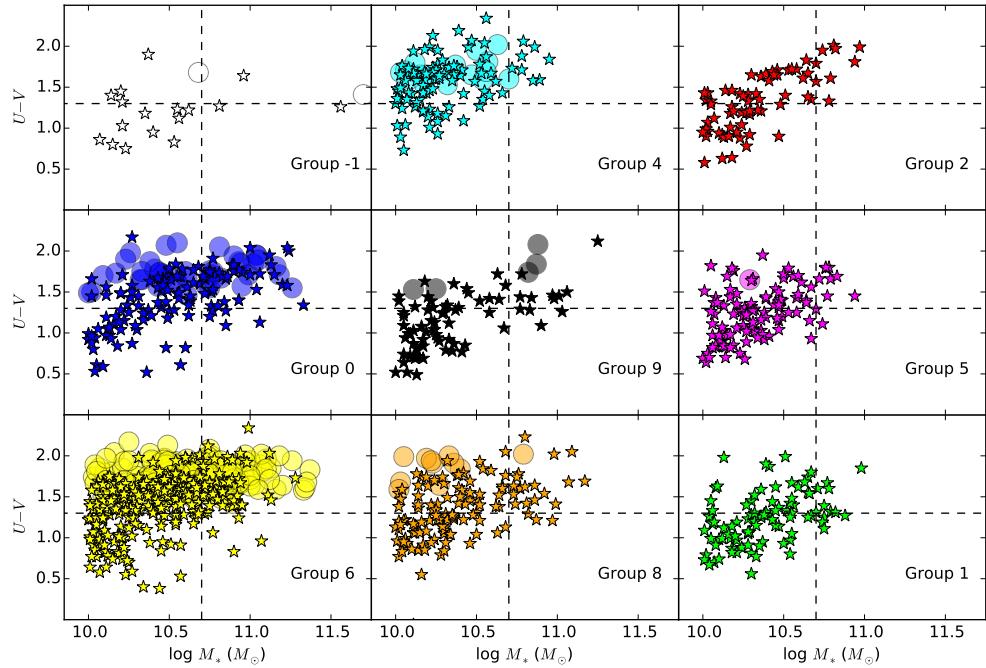


Figure 2.12 Rest-frame $U - V$ vs. Stellar Mass diagram for $M_* > 10^{10} M_\odot$, $1.36 < z < 1.97$ galaxies for each cluster group. Galaxies classified by UVJ as star-forming (*stars*) and quenched (*circles*) are shown for each group. The dashed line in $U - V$ represents the approximate dividing line between quenched and star-forming galaxies. Groups 6, 0, and 9 have the greatest fractions of galaxies with large masses (dashed line, $M_* > 5 \times 10^{10} M_\odot$). The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

CHAPTER 2. PCA MORPHOLOGY

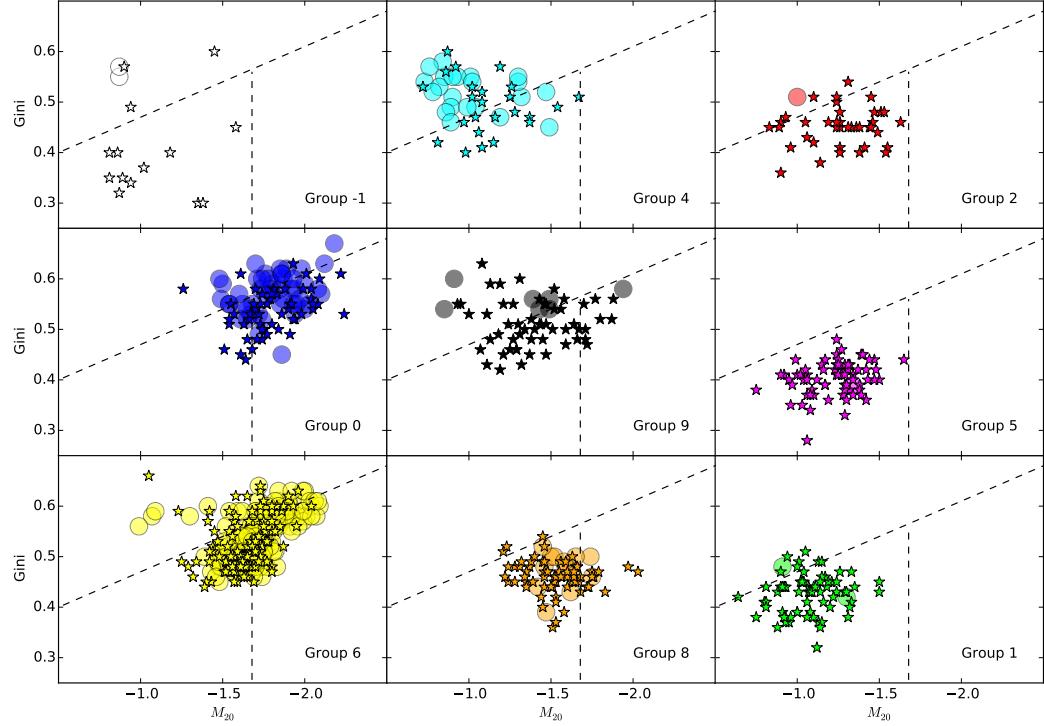


Figure 2.13 $G - M_{20}$ for each group. Overplotted are the dividing lines between: mergers (*top left corner*), bulge-dominated (*right-most region*), and disk-dominated (*bottom left region*) modified from Lotz et al. (2004). Group 0 fully occupies the bulge-dominated region of the plot. Symbols same as Fig. 2.12. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

CHAPTER 2. PCA MORPHOLOGY

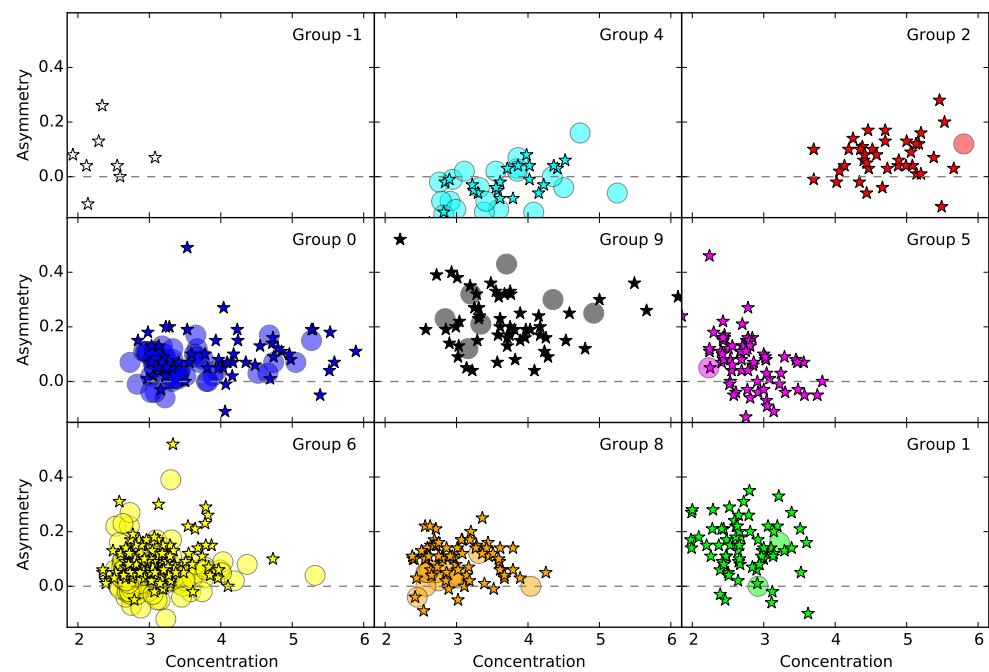


Figure 2.14 Concentration - Asymmetry for each group. Plotting symbols same as Fig. 2.12. Groups 9 and 1 have the highest asymmetry, while group 0 has the highest concentration. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

CHAPTER 2. PCA MORPHOLOGY

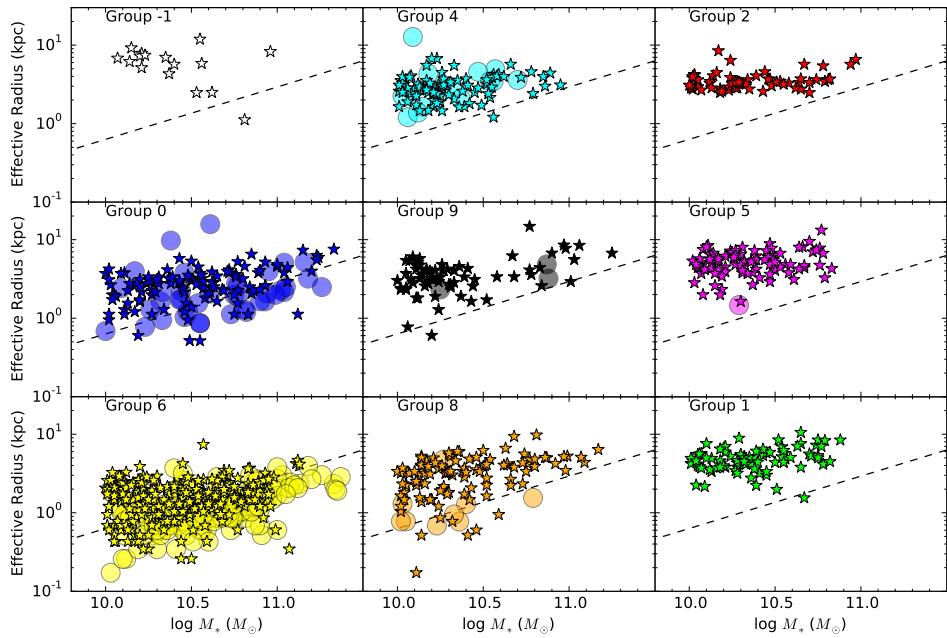


Figure 2.15 Effective radii (kpc, as measured in WFC3 F160W by van der Wel et al., 2012) vs. stellar mass for each group. Dotted lines represent the “compact” criteria ($M/r_e^{1.5} < 10.3 M_\odot \text{ kpc}^{-1.5}$) of Barro et al. (2013). Almost all group 6 galaxies are very compact, with most galaxies smaller than 2 kpc. Groups 0 and 8 have a number of borderline compact galaxies. The remaining groups have only a few compact galaxies at most. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

CHAPTER 2. PCA MORPHOLOGY

reinforce how effective our PCA groups are at separating different classes of galaxies.

In Fig. 2.13 the dotted lines signify classification regions adapted from Lotz et al. (2004) for $z \sim 1-2$ galaxies observed by *HST*. Mergers are in the upper left region, late-type galaxies are in the lower region and early-type galaxies are in the wedge-shaped region on the rightmost portion of the $G - M_{20}$ diagram. $C - A$ diagrams (for review see Conselice, 2014) have been used to differentiate giant ellipticals (which live in regions of large C and small A) from spirals (with progressively smaller C and larger A) and from mergers (which are the most asymmetric but the least centrally concentrated).

For our group descriptions in the following sections we will refer heavily to Fig. 2.11 - 2.15, the example galaxies of Fig. 2.17 - 2.25 and Tables 2.2 - 2.5. For these figures the locations of each subplot represents the approximate position of that group in PCA space. From left to right, PC1 increases which is indicative of an increase in bulge strength. From bottom to top, PC2 increases thus concentration increases.

Tables 2.2-2.5 describe the group demographics in terms of stellar mass, visual classification (Kartaltepe et al., 2015), Sérsic indices (van der Wel et al., 2012) and quenched fraction. These demographics are both listed in terms of the original group (as determined by the hierarchical clustering method, left columns) and in terms of the MC group (determined using the scattering method, right columns). The agreement between the galaxy demographics in the original groups and scattered MC groups shows the group characteristics are quite robust to noise. Table 2.2 shows that high

CHAPTER 2. PCA MORPHOLOGY

PC1 (disk-dominated) groups have very few high mass galaxies. Meanwhile, low PC1 (compact/bulge-dominated) groups have a larger fraction of high mass galaxies.

We use the CANDELS visual classifications (Kartaltepe et al., 2015) to determine the demography of the PCA groups in disk, spheroidal and irregular galaxy classes. For a galaxy to be counted as a “disk”, “spheroid” or “irregular” it must have been classified by at least two-thirds of the classifiers as such, and less than one-third as the other classes. A “disk+spheroid” is classified as *both* a disk and a spheroid by at least two-thirds of the classifiers. The “other” class represents everything that does not belong to the other 4 categories. The fractions of galaxies in each morphological type are shown in Table 2.3.

Sérsic fits have been used extensively to classify galaxies into early- and late-type categories (van der Wel et al., 2012; van Dokkum et al., 2010; Patel et al., 2011; Peng et al., 2002). Typically, $n=2.5$ is used to divide late-type ($n < 2.5$) and early-type ($n > 2.5$) galaxies. Table 2.4 shows the percentage of galaxies representing a certain classification for each group as a percentage of the group population (van der Wel et al., 2012). Similar to visual classification, the percentage of galaxies with disk-dominated morphologies decreases with decreasing PC1 values.

Table 2.5 and Fig. 2.11 show that in this redshift range ($1.36 < z < 1.97$) and mass range ($\gtrsim 10^{10} M_{\odot}$) only 23% of galaxies are quenched. Table 2.5 shows that the quenched fraction for a group is anti-correlated to PC1 and PC2.

Fig. 2.15 shows the effective radii (kpc) - stellar mass relation for each group. In

CHAPTER 2. PCA MORPHOLOGY

this figure, PC1 and PC2 are strongly correlated a galaxy's compactness. Group 6 galaxies are by far the most compact, with the largest fraction of quenched galaxies. As PC1 and PC2 increase the number of quenched galaxies in each group decreases.

Group 6

Constituting 37% of the entire sample, group 6 is by far the most populated group at $z \sim 1.5$ (example postage stamps in Fig. 2.17) . Group 6 galaxies are characterized by their compact sizes ($r_e \sim 1.57 \pm 0.81$ kpc) and smooth features. Many of these galaxies are barely resolved by *HST WFC3* which leads to their structureless appearance. Therefore, the structural properties of this group should be interpreted with caution, since it is possible that unresolved features in these galaxies would cause them to be classified as a different group if we had access to higher resolution observations. 43% of the group is quenched, which represents 72% of all quenched galaxies at this redshift. Groups 0 and 4 are the only other group with a >10% fraction of quenched galaxies.

Group 6 galaxies also dominate the high mass galaxies at this epoch, constituting 48% of galaxies with $5 \times 10^{10} M_\odot < M_* < 10^{11} M_\odot$ and 49% of galaxies with $M_* > 10^{11} M_\odot$.

Group 6 galaxies have low concentrations ($C \sim 3.03 \pm 0.40$), moderate Gini coefficients ($G \sim 0.53 \pm 0.05$), low M_{20} ($\sim -1.67 \pm 0.17$), extremely low MID values ($I \sim 0.00 \pm 0.02$, $D \sim 0.06 \pm 0.04$), and low asymmetry values ($A \sim 0.05 \pm 0.06$). The

CHAPTER 2. PCA MORPHOLOGY

Table 2.2 Group percentages by mass range for both original group and “MC Group”

		$\log M_* < 10.5$			$10.5 < \log M_* < 11.0$			$\log M_* > 11.0$			Group Percentage	
Group		$\log M_* < 10.5$		$10.5 < \log M_* < 11.0$		$10.5 < \log M_* < 11.0$		$\log M_* > 11.0$		$\log M_* > 11.0$		
Low PC1	6	54.9 ^{+4.3} _{-4.0}	58.5 ^{+4.6} _{-4.3}	41.4 ^{+3.7} _{-3.4}	38.2 ^{+3.8} _{-3.5}	3.8 ^{+1.4} _{-1.0}	3.3 ^{+1.4} _{-1.0}	33.2 ⁽³¹⁹⁾	30.2 ^(290.4)	30.2 ^(290.4)	30.2 ^(290.4)	
	0	44.4 ^{+7.0} _{-6.1}	45.9 ^{+8.7} _{-7.4}	41.7 ^{+6.9} _{-5.9}	40.1 ^{+8.2} _{-6.9}	13.9 ^{+4.4} _{-3.5}	14.0 ^{+5.4} _{-4.1}	11.2 ⁽¹⁰⁸⁾	8.0 ^(77.1)	8.0 ^(77.1)	8.0 ^(77.1)	
	9	69.7 ^{+11.3} _{-9.8}	65.1 ^{+7.5} _{-6.8}	21.2 ^{+7.0} _{-5.5}	29.2 ^{+5.3} _{-4.5}	9.1 ^{+5.2} _{-3.6}	5.7 ^{+2.8} _{-2.0}	6.9 ⁽⁶⁶⁾	13.5 ^(129.9)	13.5 ^(129.9)	13.5 ^(129.9)	
Mid PC1	4	68.8 ^{+10.3} _{-9.0}	68.6 ^{+11.3} _{-9.8}	31.2 ^{+7.4} _{-6.1}	29.3 ^{+8.0} _{-6.4}	0.0 ^{+2.3} _{-0.6}	2.1 ^{+3.6} _{-1.9}	8.0 ⁽⁷⁷⁾	6.8 ^(65.6)	6.8 ^(65.6)	6.8 ^(65.6)	
	8	69.8 ^{+7.9} _{-7.1}	71.2 ^{+9.3} _{-8.3}	27.8 ^{+5.3} _{-4.5}	26.8 ^{+6.1} _{-5.1}	2.4 ^{+2.2} _{-1.4}	2.0 ^{+2.6} _{-1.5}	13.1 ⁽¹²⁶⁾	9.9 ^(95.3)	9.9 ^(95.3)	9.9 ^(95.3)	
		69.5 ^{+1.0} _{-0.5}	70.1 ^{+5.6} _{-5.0}	29.1 ^{+1.0} _{-0.4}	27.8 ^{+3.8} _{-3.1}	1.5 ^{+0.9} _{-0.3}	2.0 ^{+1.6} _{-0.9}					
High PC1	1	62.9 ^{+9.1} _{-8.0}	61.8 ^{+8.1} _{-7.2}	36.0 ^{+7.2} _{-6.1}	35.2 ^{+6.4} _{-5.4}	1.1 ^{+2.5} _{-1.2}	3.0 ^{+2.7} _{-1.7}	9.3 ⁽⁸⁹⁾	11.3 ^(108.7)	11.3 ^(108.7)	11.3 ^(108.7)	
	2	81.0 ^{+12.4} _{-10.8}	67.8 ^{+9.8} _{-8.6}	17.5 ^{+6.7} _{-5.1}	29.8 ^{+6.9} _{-5.7}	1.6 ^{+3.5} _{-1.7}	2.4 ^{+3.0} _{-1.7}	6.5 ⁽⁶³⁾	8.7 ^(84.0)	8.7 ^(84.0)	8.7 ^(84.0)	
	5	75.5 ^{+9.2} _{-8.2}	67.3 ^{+8.7} _{-7.8}	23.5 ^{+5.6} _{-4.6}	30.0 ^{+6.2} _{-5.2}	1.0 ^{+2.2} _{-1.1}	2.7 ^{+2.7} _{-1.6}	10.6 ⁽¹⁰²⁾	10.6 ^(101.8)	10.6 ^(101.8)	10.6 ^(101.8)	
Outliers	72.4 ^{+0.8} _{-0.4}	65.4 ^{+3.0} _{-2.6}	26.4 ^{+0.8} _{-0.3}	31.9 ^{+2.2} _{-1.8}	1.2 ^{+0.7} _{-0.2}	2.7 ^{+0.9} _{-0.6}						
	-1	50.0 ^{+28.7} _{-20.0}	48.5 ^{+35.0} _{-23.1}	50.0 ^{+28.7} _{-20.0}	51.4 ^{+35.6} _{-23.7}	0.0 ^{+15.0} _{-4.2}	0.1 ^{+20.3} _{-5.7}	1.2 ⁽¹²⁾	0.9 ^(8.9)	0.9 ^(8.9)	0.9 ^(8.9)	
N Galaxies	600		323		39			962				

Note: The left hand columns for each mass range represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications.

CHAPTER 2. PCA MORPHOLOGY

Table 2.3 Demographics of Visual Classifications of Groups

	Group	Disks	Spheroids	Irregulars	D+Sph	Other
Low PC1	6	13.6 ^{+2.9} _{-2.4}	25.8 ^{+1.1} _{-0.4}	52.3 ^{+5.2} _{-4.7}	43.0 ^{+0.1} _{-0.5}	1.4 ^{+1.3} _{-0.8}
	0	44.6 ^{+8.8} _{-7.4}	45.5 ^{+3.6} _{-1.3}	21.6 ^{+6.6} _{-5.2}	20.0 ^{+3.5} _{-1.1}	1.4 ^{+3.0} _{-1.5}
	9	45.2 ^{+12.4} _{-10.0}	46.6 ^{+2.4} _{-0.9}	9.5 ^{+7.3} _{-4.7}	20.9 ^{+2.3} _{-0.8}	26.2 ^{+10.1} _{-7.6}
Mid PC1	4	24.5 ^{+0.6} _{-0.2}	34.2 ^{+0.6} _{-0.2}	40.0 ^{+0.6} _{-0.2}	33.7 ^{+0.6} _{-0.2}	4.5 ^{+0.6} _{-0.2}
	8	62.5 ^{+13.0} _{-10.9}	64.1 ^{+4.6} _{-1.7}	8.3 ^{+6.3} _{-4.1}	9.6 ^{+4.3} _{-1.3}	0.0 ^{+3.7} _{-1.0}
	7	75.3 ^{+10.4} _{-9.2}	71.5 ^{+3.3} _{-1.3}	8.6 ^{+4.5} _{-3.2}	11.1 ^{+3.0} _{-0.9}	4.9 ^{+3.8} _{-2.4}
High PC1	1	70.5 ^{+1.6} _{-0.7}	68.5 ^{+2.1} _{-0.9}	8.5 ^{+1.4} _{-0.4}	10.5 ^{+1.8} _{-0.6}	3.1 ^{+1.4} _{-0.4}
	2	71.2 ^{+11.4} _{-9.9}	61.6 ^{+2.6} _{-1.1}	1.5 ^{+3.4} _{-1.6}	11.1 ^{+2.4} _{-0.7}	16.7 ^{+6.4} _{-4.9}
	5	76.3 ^{+16.2} _{-13.6}	60.8 ^{+3.6} _{-1.4}	7.9 ^{+7.4} _{-4.6}	18.7 ^{+3.4} _{-1.1}	5.3 ^{+6.7} _{-3.8}
Outliers	-1	81.5 ^{+1.3} _{-0.6}	64.6 ^{+1.1} _{-0.5}	2.4 ^{+1.1} _{-0.3}	12.9 ^{+0.9} _{-0.3}	8.9 ^{+1.1} _{-0.3}
	0	50.0 ^{+28.1} _{-24.8}	40.0 ^{+31.1} _{-9.0}	25.0 ^{+21.1} _{-18.1}	34.0 ^{+31.1} _{-8.9}	12.5 ^{+27.9} _{-13.6}
	Total Fraction	49% (31.3)	24% (149)	5% (35)	15% (56)	7% (42)

Note: Visual classification from Kartaltepe et al. (2015) for UDS and GOODS-S (no classifications for COSMOS galaxies). For a galaxy to be visually classified 2/3 observers need to agree. ‘Other’ classification refers to galaxies failing the 2/3 agreement requirement. The left hand columns for each visual classification represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications.

CHAPTER 2. PCA MORPHOLOGY

Table 2.4 Demographics of Sérsic Classifications of Groups

Group	0 < n < 1			1 < n < 2.5			2.5 < n < 4			n > 4		
	6	37.9 ^{+4.5} _{-4.0}	21.2 ^{+1.1} _{-0.4}	22.9 ^{+3.6} _{-3.1}	35.7 ^{+1.2} _{-0.5}	21.6 ^{+3.5} _{-3.0}	26.1 ^{+1.2} _{-0.5}	10.0 ^{+2.5} _{-2.1}	14.0 ^{+1.1} _{-0.4}	10.0 ^{+2.5} _{-2.1}	14.0 ^{+1.1} _{-0.4}	14.0 ^{+1.1} _{-0.4}
Low PC1	0	31.5 ^{+7.6} _{-6.3}	10.2 ^{+3.7} _{-1.1}	23.1 ^{+6.7} _{-5.4}	36.0 ^{+3.8} _{-1.3}	18.5 ^{+6.2} _{-4.8}	26.1 ^{+3.8} _{-1.2}	19.4 ^{+6.3} _{-4.9}	27.7 ^{+3.8} _{-1.2}	19.4 ^{+6.3} _{-4.9}	27.7 ^{+3.8} _{-1.2}	27.7 ^{+3.8} _{-1.2}
	9	59.1 ^{+13.8} _{-11.4}	33.9 ^{+2.6} _{-1.0}	25.8 ^{+10.0} _{-7.6}	41.6 ^{+2.7} _{-1.0}	3.0 ^{+5.5} _{-2.8}	13.9 ^{+2.5} _{-0.8}	4.5 ^{+6.0} _{-3.4}	10.6 ^{+2.5} _{-0.8}	4.5 ^{+6.0} _{-3.4}	10.6 ^{+2.5} _{-0.8}	10.6 ^{+2.5} _{-0.8}
		39.2 ^{+0.6} _{-0.2}	22.5 ^{+0.6} _{-0.2}	23.3 ^{+0.6} _{-0.2}	37.2 ^{+0.7} _{-0.3}	18.6 ^{+0.6} _{-0.2}	24.9 ^{+0.6} _{-0.2}	11.4 ^{+0.6} _{-0.2}	15.4 ^{+0.6} _{-0.2}	11.4 ^{+0.6} _{-0.2}	15.4 ^{+0.6} _{-0.2}	15.4 ^{+0.6} _{-0.2}
Mid PC1	4	51.9 ^{+12.1} _{-10.0}	39.2 ^{+5.1} _{-1.7}	27.3 ^{+9.4} _{-7.3}	43.0 ^{+5.1} _{-1.7}	6.5 ^{+5.9} _{-3.7}	10.7 ^{+5.0} _{-1.5}	6.5 ^{+5.9} _{-3.7}	7.1 ^{+5.0} _{-3.7}	7.1 ^{+5.0} _{-3.7}	7.1 ^{+5.0} _{-3.7}	7.1 ^{+5.0} _{-3.7}
	8	61.9 ^{+9.6} _{-8.4}	49.7 ^{+3.5} _{-1.3}	20.6 ^{+6.1} _{-4.9}	35.2 ^{+3.4} _{-1.2}	4.8 ^{+3.7} _{-2.4}	10.9 ^{+3.3} _{-1.0}	1.6 ^{+2.9} _{-1.5}	4.2 ^{+3.3} _{-0.9}	4.2 ^{+3.3} _{-0.9}	4.2 ^{+3.3} _{-0.9}	4.2 ^{+3.3} _{-0.9}
		58.2 ^{+1.6} _{-0.7}	45.6 ^{+2.2} _{-0.9}	23.1 ^{+1.5} _{-0.5}	38.3 ^{+2.2} _{-0.9}	5.4 ^{+1.4} _{-0.4}	10.8 ^{+2.0} _{-0.6}	3.4 ^{+1.4} _{-0.4}	5.3 ^{+2.0} _{-0.6}	5.3 ^{+2.0} _{-0.6}	5.3 ^{+2.0} _{-0.6}	5.3 ^{+2.0} _{-0.6}
High PC1	1	82.0 ^{+10.7} _{-10.7}	68.5 ^{+3.0} _{-1.2}	3.4 ^{+4.0} _{-2.3}	12.9 ^{+2.8} _{-0.9}	0.0 ^{+2.7} _{-0.8}	10.9 ^{+2.8} _{-0.8}	1.1 ^{+3.2} _{-1.5}	7.6 ^{+2.7} _{-0.8}	7.6 ^{+2.7} _{-0.8}	7.6 ^{+2.7} _{-0.8}	7.6 ^{+2.7} _{-0.8}
	2	69.8 ^{+15.6} _{-13.0}	43.0 ^{+3.9} _{-1.4}	14.3 ^{+8.8} _{-6.0}	33.2 ^{+3.9} _{-1.3}	1.6 ^{+5.5} _{-2.4}	12.2 ^{+3.8} _{-1.1}	3.2 ^{+6.1} _{-3.1}	11.6 ^{+3.8} _{-1.1}	11.6 ^{+3.8} _{-1.1}	11.6 ^{+3.8} _{-1.1}	11.6 ^{+3.8} _{-1.1}
	5	86.3 ^{+12.6} _{-11.1}	68.4 ^{+3.4} _{-1.4}	4.9 ^{+4.5} _{-2.8}	14.5 ^{+3.2} _{-1.0}	0.0 ^{+2.8} _{-0.8}	9.5 ^{+3.1} _{-0.9}	0.0 ^{+2.8} _{-0.8}	7.6 ^{+3.1} _{-0.9}	7.6 ^{+3.1} _{-0.9}	7.6 ^{+3.1} _{-0.9}	7.6 ^{+3.1} _{-0.9}
Outliers		80.9 ^{+1.3} _{-0.6}	61.3 ^{+1.2} _{-0.5}	6.4 ^{+1.1} _{-0.3}	19.1 ^{+1.1} _{-0.4}	0.4 ^{+1.1} _{-0.3}	10.8 ^{+1.1} _{-0.3}	1.2 ^{+1.1} _{-0.3}	8.7 ^{+1.1} _{-0.3}	8.7 ^{+1.1} _{-0.3}	8.7 ^{+1.1} _{-0.3}	8.7 ^{+1.1} _{-0.3}
	-1	83.3 ^{+44.5} _{-31.5}	66.0 ^{+31.3} _{-9.2}	0.0 ^{+22.5} _{-6.2}	0.0 ^{+30.9} _{-8.6}	0.0 ^{+22.5} _{-6.2}	0.0 ^{+30.9} _{-8.6}	16.7 ^{+29.3} _{-15.2}	34.0 ^{+31.1} _{-8.9}	34.0 ^{+31.1} _{-8.9}	34.0 ^{+31.1} _{-8.9}	34.0 ^{+31.1} _{-8.9}
Total Fraction		60% (527)		20% (179)		12% (103)		8% (68)				

Note: The left hand columns for each Sérsic-index range represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications. Due to the small sizes of certain galaxies, not every galaxy has a measured Sérsic fit.

CHAPTER 2. PCA MORPHOLOGY

Table 2.5 UVJ Quenched Fractions of Groups

	Group	Quenched	Star-Forming	
Low PC1	6	$43.5^{+3.1}_{-2.9}$	$39.3^{+3.3}_{-3.0}$	$56.5^{+3.5}_{-3.3}$
	0	$25.9^{+4.4}_{-3.8}$	$27.0^{+6.5}_{-5.3}$	$74.1^{+7.0}_{-6.4}$
	9	$7.6^{+4.4}_{-3.0}$	$15.1^{+3.5}_{-2.9}$	$92.4^{+11.6}_{-10.3}$
Mid PC1		$35.4^{+0.3}_{-0.1}$	$31.3^{+1.0}_{-0.8}$	$64.6^{+0.3}_{-0.1}$
	4	$15.3^{+4.5}_{-3.6}$	$17.2^{+4.9}_{-3.9}$	$84.7^{+9.2}_{-8.3}$
	8	$8.1^{+3.1}_{-2.4}$	$13.9^{+4.2}_{-3.4}$	$91.9^{+8.6}_{-7.9}$
High PC1		$11.4^{+0.8}_{-0.2}$	$15.5^{+2.2}_{-1.7}$	$88.6^{+0.9}_{-0.4}$
	1	$0.0^{+1.9}_{-0.5}$	$8.3^{+3.1}_{-2.3}$	$100.0^{+10.8}_{-9.8}$
	2	$0.0^{+2.6}_{-0.7}$	$15.0^{+4.3}_{-3.5}$	$100.0^{+12.8}_{-11.4}$
	5	$0.9^{+2.1}_{-1.0}$	$13.9^{+4.2}_{-3.4}$	$99.1^{+10.1}_{-9.2}$
		$0.4^{+0.7}_{-0.2}$	$12.1^{+1.2}_{-0.9}$	$99.6^{+0.8}_{-0.4}$
	-1	$10.5^{+13.4}_{-7.6}$	$14.2^{+15.7}_{-9.3}$	$89.5^{+26.2}_{-20.9}$
Total Fraction		23% (281)	77% (962)	

Note: Quenched/star-forming classifications based on Fig. 2.11. The left hand columns for quenched/star-forming classifications represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications.

CHAPTER 2. PCA MORPHOLOGY

$G - M_{20}$ diagram classifies the majority of these galaxies as borderline disk/spheroidal (with occasional irregular classification). However, M_{20} values are potentially biased because the 20% light is not resolved. These galaxies have large average Sérsic indices ($\bar{n} \sim 3.11$).

Group 6 is comprised of the highest percentage of visually identified spheroids (52 percent) and disk+spheroids (26 percent), and also has the lowest percentage of disks (13 percent) of any group.

Upwards of 26% of group 6 galaxies are small ($r_e < 2$ kpc) with high Sérsic ($n > 2.5$) which could result in an underestimation of concentration and PC2 values. These galaxies would instead be classified into group 0.

Group 0

Group 0 galaxies are characterized by a strong bulge component which is surrounded by a faint smooth extended component (example postage stamps in Fig. 2.18). A significant fraction of group 0 galaxies are quenched galaxies (26 percent; Table 2.5). Although group 0 galaxies make up only 13% of the galaxies in the sample, they constitute 38% of the galaxies more massive than $10^{11} M_\odot$ (Table 2.2).

These galaxies have high concentration values ($C \sim 3.80 \pm 0.78$), low M_{20} ($\sim -1.80 \pm 0.17$), high Gini coefficients ($G \sim 0.55 \pm 0.04$), low deviations ($D \sim 0.06 \pm 0.04$), low multi-modes ($M \sim 0.03 \pm 0.04$), low intensities ($I \sim 0.03 \pm 0.04$) and low asymmetries ($A \sim 0.06 \pm 0.07$). This group of galaxies is the only class to fall almost

CHAPTER 2. PCA MORPHOLOGY

entirely into the spheroidal region of the $G - M_{20}$ diagram.

Visually, these galaxies have a large disk+spheroid fraction (33 percent), a large spheroid fraction (35 percent), a moderate disk fraction (31 percent) and a very low irregular fraction (1 percent). Parametric fits find that group 0 galaxies have moderately sized effective radii ($r_e \sim 3.13 \pm 1.97$ kpc) and large average Sérsic indices ($\bar{n} \sim 3.87$). The visual classifications and distribution of Sérsic indices agree with $G - M_{20}$ measurements and thus describe the prototypical group 0 galaxy as bulge-dominated with a faint disk component or extended envelope.

Group 9

Group 9 is characterized by their asymmetric, irregular morphologies and strong bulge component (example postage stamps in Fig. 2.19). These galaxies make up a significant portion of the $M_* > 10^{11} M_\odot$ galaxies (15 percent). However, most of these galaxies are lower mass ($M_* < 3 \times 10^{10} M_\odot$). Only 8% of group 9 galaxies are quenched.

These galaxies have moderate concentrations ($C \sim 3.70 \pm 0.70$), moderate Gini coefficient ($G \sim 0.52 \pm 0.05$), moderate M_{20} ($\sim -1.40 \pm 0.27$), moderate MID values ($M \sim 0.14 \pm 0.14$, $I \sim 0.21 \pm 0.18$, $D \sim 0.19 \pm 0.09$) and high asymmetry ($A \sim 0.21 \pm 0.10$). These galaxies lie along the $G - M_{20}$ merger/disk galaxy dividing line and also overlap with the spheroidal region.

Group 9 galaxies have large radii ($r_e \sim 3.67 \pm 1.64$ kpc) and moderately low

CHAPTER 2. PCA MORPHOLOGY

average Sérsic indices ($\bar{n} \sim 2.11$).

This group is the most visually irregular group (24 percent), and has a relatively low disk fraction (41 percent), spheroid fraction (13 percent) and disk+spheroid fraction (11 percent). These statistics and visual classifications imply many galaxies have bright off-center clusters, in addition to bright central bulges.

Group 4

Group 4 galaxies are low-mass, smooth, extended galaxies with moderate central concentrations (example postage stamps in Fig. 2.20). Although mostly star-forming, group 4 contains some quenched galaxies (~ 11 percent). Some galaxies are extended and also quenched; meaning they are rare “red disk” population. None of the group 4 galaxies are more massive than $M_* > 10^{11} M_\odot$. Primarily these galaxies are lower mass ($M_* < 3 \times 10^{10} M_\odot$).

Group 4 has moderate concentrations ($C \sim 3.53 \pm 0.66$), moderate Gini coefficients ($G \sim 0.49 \pm 0.04$), high M_{20} ($\sim -1.11 \pm 0.24$), low intensities ($I \sim 0.05 \pm 0.05$), small multi-mode values ($M \sim 0.07 \pm 0.07$), low deviations ($D \sim 0.10 \pm 0.07$), and low asymmetry ($A \lesssim 0$).

Group 4 galaxies have moderate effective radii ($r_e \sim 3.13 \pm 1.63$ kpc) and medium average Sérsic indices ($\bar{n} \sim 2.68$).

Group 4 members are primarily visually classified as disks (51 percent) or disk+spheroids (24 percent) and are less classified as spheroids (17 percent) or irregulars (0 percent).

Group 8

Group 8 galaxies are an interesting class of bulge+disk systems with dominant and smooth disks (example postage stamps in Fig. 2.21). This class is dominated by low-mass star-forming galaxies, but also includes low-mass ($< 3 \times 10^{10} M_\odot$) quenched galaxies (~ 8 percent). Very few galaxies have stellar masses $> 5 \times 10^{10} M_\odot$.

Group 8 galaxies have small concentrations ($C \sim 3.05 \pm 0.43$), moderate Gini coefficients ($G \sim 0.46 \pm 0.03$), moderate M_{20} ($\sim -1.56 \pm 0.17$), low but non-zero MID values ($M \sim 0.06 \pm 0.06$, $I \sim 0.10 \pm 0.11$, $D \sim 0.09 \pm 0.05$), and low asymmetry values ($A \sim 0.08 \pm 0.07$). On the $G - M_{20}$ diagram these galaxies fall within the disk-dominated region but are close to the spheroidal/disk dividing line.

Sérsic fits to this class find moderate sizes ($r_e \sim 3.48 \pm 1.89$ kpc) and low average Sérsic indices ($\bar{n} \sim 1.46$).

Group 8 is dominated by visually-classified disks (74 percent) with only a modest fraction of spheroids (10 percent). A small number of galaxies are quenched and compact which overlaps with groups 0 and 6.

Group 1

Group 1 galaxies are primarily large disks and irregulars with bright off-center star-forming knots (example postage stamps in Fig. 2.22). None of these galaxies are quenched based on their UVJ colors. The distribution of masses is heavily weighted

CHAPTER 2. PCA MORPHOLOGY

towards lower mass galaxies with very few objects more massive than $3 \times 10^{10} M_{\odot}$.

Group 1 galaxies have low concentration values ($C \sim 2.76 \pm 0.40$), low Gini coefficients ($G \sim 0.43 \pm 0.04$), high M_{20} ($\sim -1.07 \pm 0.17$), moderately high asymmetry values ($A \sim 0.13 \pm 0.11$), large multi-mode values ($M \sim 0.40 \pm 0.27$), high deviations ($D \sim 0.37 \pm 0.13$) and large intensities ($I \sim 0.61 \pm 0.24$). The high A and MID statistics indicate many of these galaxies have bright off-center clusters and are potentially irregular.

The visual classifications and Sérsic indices primarily classify this group as disk galaxies and/or irregulars. Group 1 is dominated by visually-classified disks (72 percent) and has a relatively large fraction of irregulars (16 percent). This group has very few spheroids or bulge-dominated disk galaxies. Their effective radii are large for this redshift and mass ($r_e \sim 5.35 \pm 1.43$ kpc). This group has low average Sérsic indices ($\bar{n} \sim 0.63$) imply a large disk and irregular population.

Group 2

Group 2 galaxies are primarily low-mass, star-forming, smooth disk galaxies with high central concentrations and few visually detected star-forming knots (example postage stamps in Fig. 2.23). None of these galaxies are quenched. The mass distribution for this group is a steeply declining function where there are only a few galaxies with masses $> 3 \times 10^{10} M_{\odot}$.

Group 2 galaxies have large concentrations ($C \sim 4.81 \pm 0.62$), low Gini coefficients

CHAPTER 2. PCA MORPHOLOGY

($G \sim 0.45 \pm 0.04$), moderate M_{20} ($\sim -1.20 \pm 0.24$), low asymmetry ($A \sim 0.06 \pm 0.08$), low deviations ($D \sim 0.16 \pm 0.09$), moderate multi-modes ($M \sim 0.16 \pm 0.21$), and a wide spread of intensity values ($I \sim 0.29 \pm 0.29$). On the $G - M_{20}$ diagram these galaxies fall within the disk-dominated and irregular portion of the diagram. However, their high C values suggest a bright nuclear component.

The visual classifications show this group is dominated by disks (76 percent) and only small fractions of irregular galaxies (~5 percent) and disk+spheroid galaxies (~8 percent). They have mid-sized effective radii ($r_e \sim 3.52 \pm 0.83$ kpc) and mid-to-low average Sérsic indices ($\bar{n} \sim 1.10$).

Group 5

Group 5 galaxies are primarily low-mass ($M_* < 3 \times 10^{10} M_\odot$), star forming, extended disk galaxies with a weak bulge component (example postage stamps in Fig. 2.24). This group has a negligible fraction of quenched galaxies (~1 percent).

Group 5 is mostly comprised of low concentration values ($C \sim 2.87 \pm 0.42$), low Gini coefficients ($G \sim 0.40 \pm 0.03$), low/moderate M_{20} ($\sim -1.20 \pm 0.19$), a wide spread in multi-mode ($M \sim 0.26 \pm 0.24$), large intensity values ($I \sim 0.52 \pm 0.28$), low deviation values ($D \sim 0.12 \pm 0.06$), and low asymmetry values ($A \sim 0.03 \pm 0.12$). On the $G - M_{20}$ diagram these galaxies fall solidly within the disk-dominated region.

The defining feature of this group is its large typical size ($r_e \sim 5.47 \pm 1.81$ kpc).

CHAPTER 2. PCA MORPHOLOGY

Group 5 galaxies have very low average Sérsic indices ($\bar{n} \sim 0.65$); implying a disk-dominated/irregular population.

Visual classification indicate group 5 is comprised almost entirely of disks (95 percent), and a few irregulars (3 percent). This group has no visually identified bulge-dominated or spheroidal galaxies and are not clumpy.

Group -1

The original groups 3 and 7 were comprised of only a few galaxies each (19 in total, example postage stamps in Fig. 2.25). They were outliers from the remaining groups and are combined into a single outlier group. These galaxies are most likely outliers because they have at least one poorly measured (or missing) morphological parameter (especially the multi-mode statistic). These galaxies have a low surface brightness, very large radii ($r_e \sim 6.73 \pm 2.30$ kpc), low concentration ($C \sim 2.21 \pm 0.74$), high intensity ($I \sim 0.44 \pm 0.39$), high M_{20} ($\sim -0.99 \pm 0.26$), low Gini coefficient ($G \sim 0.41 \pm 0.10$), extremely high deviations ($D \sim 0.69 \pm 0.49$) and high multi-modes ($M \sim 0.53 \pm 0.39$). The deviation values can separate group -1 galaxies from all the other groups.

2.6 Discussion

The spatial distribution of light for galaxies is a snapshot of the orbital paths of the constituent stars, gas, and dust. The morphology of a galaxy informs us of the merger and gas-accretion history in ways integrated colors, spectral-energy distributions and stellar mass cannot directly probe.

Using a Sérsic index, bulge-dominated galaxies are traditionally defined to have $n > 2.5$ (e.g. Bruce et al., 2014a). For the purposes of our PC classifications we define galaxies with low PC1 values as bulge-dominated (the constituents of groups 0, 6 and 9). These two definitions lead to differences in the characteristics of what are defined as ‘bulge-dominated’ and we will explore these differences in the following sections.

The connection between morphology and star-formation has been well studied (Wuyts et al., 2011; Kriek et al., 2009; Brinchmann et al., 2004). Late-type galaxies are typically still actively forming stars, whereas early-type galaxies have had their star-formation quenched. However, there are examples of red, quenched disks and blue, star-forming ellipticals which are important rare “transitional” classes. In our study we delve deeper into the correlations between morphological type and star-formation and how the connection between them is not always clear-cut.

CHAPTER 2. PCA MORPHOLOGY

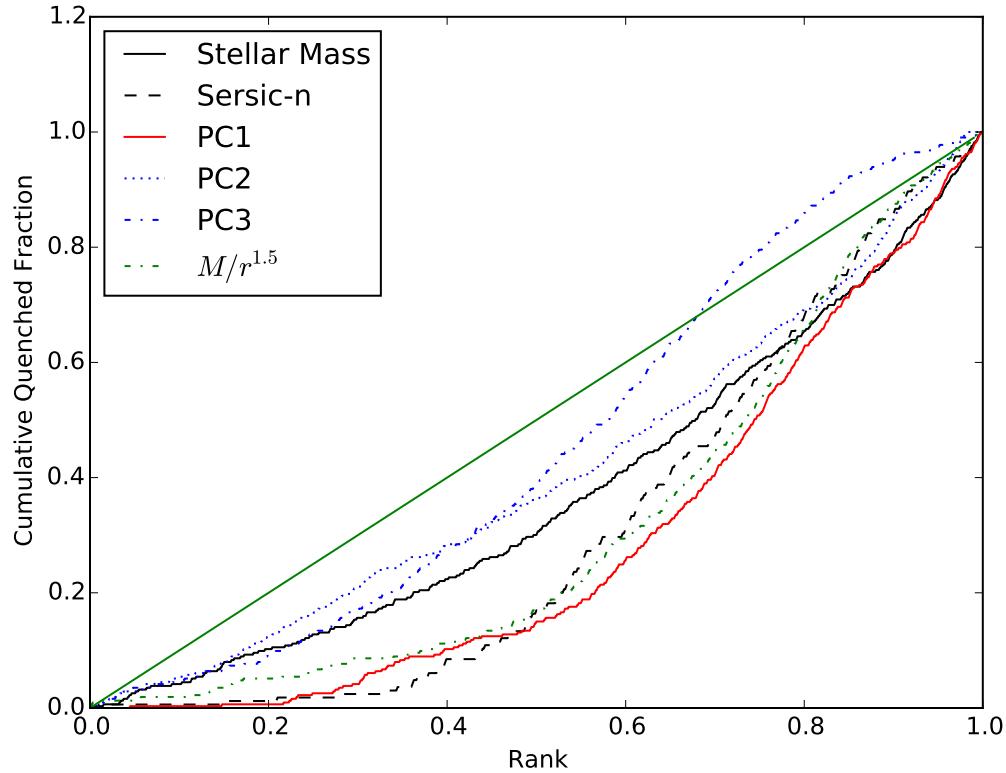


Figure 2.16 Cumulative quenched fraction rank ordered by various metrics: PC1, PC2, PC3, stellar mass, Sérsic-n and “compactness”. The green solid line represents no correlation between quenched fraction and rank. Sérsic-n and PC1 have a similar CQF shape, where n is less contaminated by quenched galaxies at low values but PC1 is less contaminated at high values.

2.6.1 Stellar Mass - Quenching Connection for groups

Fig. 2.16 shows the cumulative distribution of the quenched fraction rank-ordered by “compactness” ($M/r_e^{1.5} < 10.3 M_\odot \text{ kpc}^{-1.5}$; Barro et al., 2013), stellar mass, Sérsic- n , PC1, PC2 and PC3. For every galaxy we assign a 0 to star-forming galaxies and $1/n_{\text{quenched}}$ for quenched galaxies (as determined by the UVJ diagram, Fig. 2.11) and then these values are cumulatively summed. We observe a flat trend in stellar mass and PC2, and a much steeper trend in PC1 and Sérsic- n . However, Sérsic- n has previously been shown to correlate well with quenching (e.g. Wuyts et al., 2011; Bell et al., 2012). The similarities in steepness between the PC1 and Sérsic- n curves show PC1 is an equivalently useful predictor of quenching.

We also investigate the relationship between quenching and PC1 through the color-mass relation. In Fig. 2.12 we observe a correlation between the increase in the fraction of massive galaxies ($> 5 \times 10^{10} M_\odot$) for a specific group and the magnitude of PC1 (bulge strength). The amount of quenched galaxies correlates more strongly with PC1 (bulge strength) than PC2 (concentration). Similar results have been found for $z \sim 1$ -2 galaxies (Bell et al., 2012; Barro et al., 2013; Lang et al., 2014). Unsurprisingly, the most massive galaxies are also the most likely to be quenched. For instance, group 6 has the largest amount of red galaxies and many massive galaxies ($> 5 \times 10^{10} M_\odot$). The only other groups with a substantial number of quiescent galaxies are groups 0 (26 percent) and 4 (15 percent). Group 0 galaxies are primarily bulge-dominated with a faint disk. However, group 6 galaxies are more massive

CHAPTER 2. PCA MORPHOLOGY

($> 5 \times 10^{10} M_{\odot}$) than group 0 galaxies. Furthermore, a much larger percentage of these massive galaxies in group 6 are quenched. Group 9 galaxies are slightly less massive ($\log M_* \sim 10.4$), but still generally have a strong bulge component (as determined by PC1).

Group 4 and 8 galaxies fall between the extremes of the bulge-dominated groups (0, 6 and 9) and the disk-dominated groups (1, 2 and 5) in stellar mass, bulge strength and quenched fraction (see Tables 2.2 - 2.5). The galaxies of groups 4 and 8 are more bulge-dominated than the disk-dominated galaxies which would explain the larger quenched fraction. Groups 4 and 8 galaxies are not as massive as those in the bulge-dominated groups 0, 6 and 9 (Table 2.2) and are not quenched to the same extent either (Table 2.5).

2.6.2 The relationship between PCA Classes and Visual/Sérsic Classifications

PCA, in conjunction with our group finding algorithm, provides a distinct picture of galaxy structure from Sérsic index and visually based classifications. This classification scheme also separates quenched compact galaxies (group 6) from larger, smooth proto-elliptical systems (group 0), and star-forming disk-dominated clumpy galaxies (group 1) from star-forming bulge-dominated asymmetric galaxies (group 9). Separating clumpy star-formers and bulge dominated star forming galaxies has great

CHAPTER 2. PCA MORPHOLOGY

importance for understanding the mechanisms that formed these galaxies and the potential avenues for evolution available to them.

Based upon the visual and Sérsic classifications, our groups belong to 3 distinct types: the “disk-dominated” galaxies of groups 1, 2, and 5; the “compact/bulge-dominated” galaxies of groups 0, 6, and 9; and the “intermediate” galaxies of groups 4 and 8. For the purposes of our discussion we refer the reader to Figs. 2.11 - 2.15, the example galaxies of Figs. 2.17 - 2.25 and Tables 2.2 - 2.5.

2.6.2.1 The Compact and Bulge-Dominated Galaxies: Groups 0, 6 and 9

Galaxies in groups 0, 6, and 9 display a variety of visual classifications, but have a single unifying characteristic: many of these galaxies are bulge-dominated. Group 6 galaxies are very small and compact ($r_e \sim 1.57 \pm 0.81$ kpc) with no discernible stellar envelope. Group 0 galaxies are slightly larger ($r_e \sim 3.13 \pm 1.97$ kpc) than group 6, and display evidence for an extended stellar envelope. Groups 0 and 6 display some distinguishing characteristics as well. Group 6 galaxies lower measured concentrations ($C \sim 3.04 \pm 0.40$) than those in group 0 ($C \sim 3.80 \pm 0.78$). The small sizes and lower concentrations for group 6 galaxies are due to the fact that r_{20} measurements are near or below the resolution limit of the survey. Additionally, r_{80} measurements are very small for group 6 compared to galaxies in all other groups (see §2.3.3).

The size-mass (Fig. 2.15) relation for these two groups is different as well. Group

CHAPTER 2. PCA MORPHOLOGY

6 galaxies are smaller but overlap in masses with group 0 galaxies. Thus many more group 6 galaxies are compact using the Barro et al. (2013) definition. Compact galaxies in group 6 are quenched, whereas the quenched galaxies of group 0 are more extended.

Group 6 galaxies are visually classified as bulge-dominated (either pure spheroid or disk+spheroid morphology) >78 % of the time. However, a Sérsic cut of $n > 2.5$ yields only 35%. Similarly for group 0 galaxies, 66% of galaxies are visually classified as bulge-dominated, but a Sérsic classification only indicates 43% are bulge-dominated galaxies. Meanwhile, group 9 galaxies are the most visually disturbed group (26% irregular) and have bright central bulges determined by PC1.

Classifications based on PCs provide a slightly different picture from those based on Sérsic- n or visual inspection. A PC classification determines $\sim 57\%$ of galaxies are compact/bulge-dominated (groups 0, 6 and 9) while visual classifications determine $\sim 47\%$ of galaxies are bulge-dominated (either pure spheroids or disk+spheroids) and Sérsic indices classify $\sim 25\%$ of galaxies as bulge-dominated ($n > 2.5$). The differences between the classification schemes are subtle but important because they mean each is probing a slightly different subset of galaxies.

The compact/bulge-dominated nature and high masses of these 3 groups could imply an evolutionary connection. In this scenario, galaxies begin as group 6 galaxies, a naked core with no extended envelope or structure. Following a gas-rich merger disturbed tidal features become visible and the galaxy becomes classified as group 9.

CHAPTER 2. PCA MORPHOLOGY

After a sufficient time for the gas to settle in a disk or spheroidal envelope ($\gtrsim 1.5$ Gyr) the galaxy would appear as a group 0 galaxy. In this scenario, the quenched galaxies of group 6 have star-formation reignited following the merger, only to once again fade during the disk settlement period. Mergers would thus be a major mechanism for triggering disk growth.

2.6.2.2 The Disk-dominated Galaxies: Groups 1, 2 and 5

Groups 1, 2 and 5 all have an overwhelmingly large percentage of visually classified disk galaxies (72%, 76%, and 96% respectively). Sérsic classifications largely agree with the visual classifications for these groups. The only difference is that Sérsic classifications yield more disk-dominated galaxies ($1 < n < 2.5$) than visual classifications would indicate. Non-parametric morphologies determine these disk galaxies have varying degrees of clumpiness and disturbances.

Group 1 galaxies are the most disturbed of the “disk-dominated” groups. They have the largest asymmetries ($A \sim 0.13 \pm 0.11$), multi-modes ($M \sim 0.40 \pm 0.27$), intensities ($I \sim 0.61 \pm 0.24$) and deviations ($D \sim 0.37 \pm 0.13$). They are more often visually classified as irregular (16 percent), but have a weaker bulge component (indicated by their larger M_{20} values, $\sim -1.07 \pm 0.17$) than groups 2 and 5.

Of the remaining disk-dominated groups, group 5 galaxies have much higher M and I statistics ($M \sim 0.26 \pm 0.24$ and $I \sim 0.52 \pm 0.28$) than those in group 2 ($M \sim 0.16 \pm 0.21$ and $I \sim 0.29 \pm 0.29$). However, these two groups have similar asymmetry

CHAPTER 2. PCA MORPHOLOGY

values ($A \sim 0.05$), M_{20} values (~ -1.2), and deviations ($D \sim 0.1$).

The disk-dominated galaxies of groups 1, 2 and 5 are on average less massive, bluer in $U - V - J$ and larger than the compact/bulge-dominated galaxies of groups 0, 6 and 9.

2.6.2.3 The Intermediate Galaxies: Groups 4 and 8

Groups 4 and 8 represent an intermediate PC class between the compact/bulge-dominated morphologies of groups 0, 6 & 9 and the disk-dominated groups 1,2 & 5. Group 4 and 8 both have a population of quenched galaxies. However, the quenched galaxies of group 8 are smaller than those of group 4.

Both groups 4 and 8 have a large fraction of galaxies with $n < 2.5$ (72% and 80%, respectively). However, group 8 galaxies are more likely to be visually classified as disks than group 4 galaxies (74% compared to 51 percent). Meanwhile, group 4 galaxies are more likely be visually classified as bulge-dominated (41% compared to 15 percent). However, the differences between groups should taken with caution as the small numbers of galaxies in these groups reduces the significance of the percentages.

For groups 4 and 8 the classifications based upon non-parametric morphologies do not always agree with classifications based on Sérsic indices or visual inspection. Group 8 has a much smaller average M_{20} value ($M_{20} \sim -1.56 \pm 0.17$) than group 4 ($M_{20} \sim -1.11 \pm 0.24$). This indicates the bulges of group 8 galaxies are large and possibly dominate the morphology. However, Sérsic indices and visual classifications

CHAPTER 2. PCA MORPHOLOGY

would suggest there is no sizable bulge component for most of these galaxies. Group 4 galaxies have high concentrations, low Sérsic indices and are the least well defined group by bootstrap measures (see Fig. 2.10). Meanwhile, the $G - M_{20}$ diagram suggests a population of irregular galaxies while visual classifications find no irregular galaxies. The bright nuclear components may be the result of an AGN or starburst activity.

2.6.2.4 Comparing the Irregular Galaxies of Groups 1 and 9

The galaxies of groups 1 and 9 are the most likely to be classified visually as irregular. While group 1 is defined by star-forming disk-dominated clumpy galaxies, group 9 is defined by star-forming bulge-dominated asymmetric galaxies with tidal features. These subtle morphological differences are missed by Sérsic index, $C - A$ and $Gini - M_{20}$ based classifications and potentially offer clues as to the formation and evolutionary tracks of these galaxies.

Group 9 galaxies display tidal features and irregular disks but their strong central bulge is missed by Sérsic fits. Group 9, itself, shows the power of our PCA classifications to find interesting subtypes of galaxy morphology. Group 9 galaxies are visually classified as disks (41 percent), irregulars (23 percent) and bulge-dominated disks (12 percent). However, small PC1 values would indicate group 9 galaxies posses a strong central bulge. Meanwhile, group 1 galaxies much more likely to be visually classified as a pure disk galaxy (72 percent), slightly less likely to be irregular (16 percent) and

CHAPTER 2. PCA MORPHOLOGY

are not bulge-dominated (0 percent). Group 1 galaxies also have higher PC1 values, indicating a weaker bulge component. Using Sérsic index classifications, both groups 1 and 9 have a very large fraction of these galaxies are disk-dominated (85 percent) as opposed to bulge-dominated (15 percent). Groups 1 and 9 would be considered very similar in a Sérsic classification and the differences between these groups are more subtle.

We observe subtle differences between these two groups in many statistics; group 9 galaxies are more asymmetric (0.22 ± 0.10 vs. 0.13 ± 0.11) and have lower M_{20} values (-1.40 ± 0.27 vs. -1.07 ± 0.17) than galaxies found in group 1. Group 9 galaxies are also more concentrated (3.70 ± 0.70 vs. 2.76 ± 0.40). Meanwhile M , I and D statistics all display an increased enhancement in group 1 galaxies because these statistics probe the existence of off-center clumps.

Based on these differences it is possible these two types of galaxies have experienced different formation scenarios or exist at different stages along their evolution. Group 9 galaxies have a large central bulge which could be the result of either a merger or the accretion of many star-forming clumps in the disk. Meanwhile, group 1 galaxies are still clumpy and have small central bulges. Different levels of the amount of violent disk instabilities (VDI; Dekel et al., 2009a; Guo et al., 2015) is a possible explanation for the segregation of groups 1 and 9. Group 9 galaxies have a larger bulge, possibly grown by the migration of clumps to the central galaxy regions following repeated VDIs. Meanwhile, group 1 galaxies, which still have bright clumps in

CHAPTER 2. PCA MORPHOLOGY

the disk (as evidenced by enhanced MID statistics) have yet to experience as many VDIs and thus the central bulge remains smaller.

2.7 Summary

We use a principal component analysis of non-parametric morphology measurements (G , M_{20} , C , A , M , I and D) and agglomerative hierarchical clustering to group galaxies into a more descriptive schema than the traditional spiral, elliptical, and irregular categories. The PCA weights we calculate (Table 2.1) show that non-parametric morphological correlations vary in importance: PC1 is based upon M, I, D, M_{20} and Gini thus it is interpreted as a bulge strength indicator; PC2 is dominated by concentration; and PC3 is dominated by asymmetry; the remaining PCs are less important and difficult to interpret.

The size-mass relation is dependent on PC1 and PC2. Galaxies with high PC1 values (stronger bulges) are generally more compact and quiescent than galaxies with high PC2 values. We determine PC1 is a valid predictor of whether a galaxy is quenched.

We observe segregations of galaxy morphology by group and describe those results as follows:

- **Compact or Bulge-dominated/low PC1, $\sim 57\%$**
 - **Group 6:** Most populated group ($\sim 37\%$ of sample, examples seen in

CHAPTER 2. PCA MORPHOLOGY

Fig. 2.17). Very compact and most massive galaxies; and contains the largest spheroidal (based on Sérsic and visual classifications) and quenched fraction.

- **Group 0:** Large bulge+disk population, has prominent bulge with faint disk component. ($\sim 13\%$, Fig. 2.18). Contains a sizable fraction of massive and quenched galaxies, not to the same extent as group 6 however.
- **Group 9:** Large and massive galaxies with a substantial irregular population. Visually, these galaxies posses tidal tails, bright star-forming knots and a large bulge. ($\sim 6\%$, Fig. 2.19)

- **Bulge+Disk/intermediate PC1, $\sim 20\%$**

- **Group 4:** Smaller and less massive bulge-dominated disk galaxies with high Gini, Sérsic index and concentration values. ($\sim 9\%$, Fig. 2.20)
- **Group 8:** Slightly larger bulge+disk systems. ($\sim 11\%$, Fig. 2.21)

- **Disk-dominated/high PC1, $\sim 22\%$**

- **Group 1:** Large galaxies with prominent (albeit) irregular disks. ($\sim 8\%$, Fig. 2.22)
- **Group 2:** Compact and small disks galaxies. ($\sim 6\%$, Fig. 2.23)
- **Group 5:** Large and low mass disk galaxies with evidence of disturbances and interactions. ($\sim 9\%$, Fig. 2.24)

CHAPTER 2. PCA MORPHOLOGY

- **Group -1:** Low surface brightness galaxies ($\sim 1\%$, Fig. 2.25) with outlier PC values.

The PC classification scheme separates quenched compact galaxies from larger, smooth proto-elliptical systems, and star-forming disk-dominated clumpy galaxies from star-forming bulge-dominated asymmetric galaxies. Additionally, classifications based on PCs provide a different picture from those based on Sérsic- n or visual inspection. A PC classification determines $\sim 51\%$ of galaxies are compact or bulge-dominated (groups 0, 6 and 9) while visual classifications determine $\sim 39\%$ of galaxies are bulge-dominated (either pure spheroids or disk+spheroids) and Sérsic indices classify ~ 20 of galaxies as bulge-dominated ($n > 2.5$).

In the future we will extend our PCA classifications to different redshifts. We will use the classifications defined here to study the evolution of star-formation for a variety of morphological types. Star-formation can be quenched in many ways and with a reliable morphology classification for different epochs we can begin to answer the question: whether star-formation quenching is occurring at the same time as the bulge is forming? A temporal connection between these two could have important consequences on how galaxies have been quenching star-formation.

CHAPTER 2. PCA MORPHOLOGY

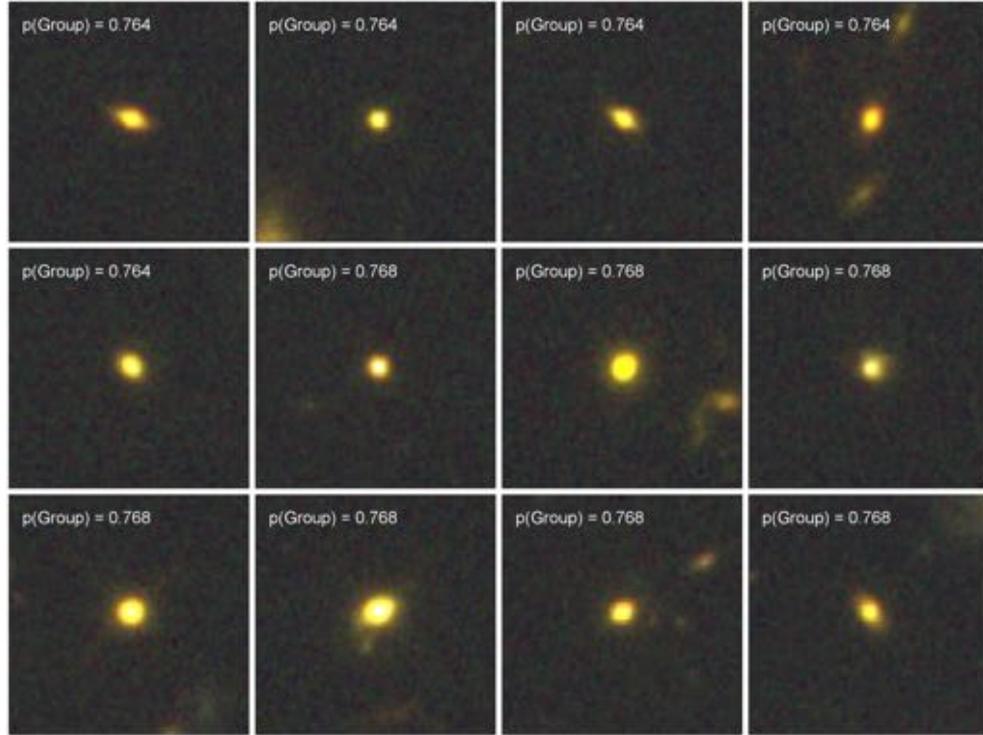


Figure 2.17 Group 6 F125W $1.36 < z < 1.97$ galaxies, shown in F160W/F125W/F814W RGB 6''x6'' postage stamps. $p(\text{Group})$ represents the percentage of times a galaxy is classified into group 6 after the scattering test. Very compact and small spheroidal galaxies. This group contains the largest spheroidal and quenched fraction. Many of these galaxies are barely resolved which leads to their structureless appearance.

CHAPTER 2. PCA MORPHOLOGY

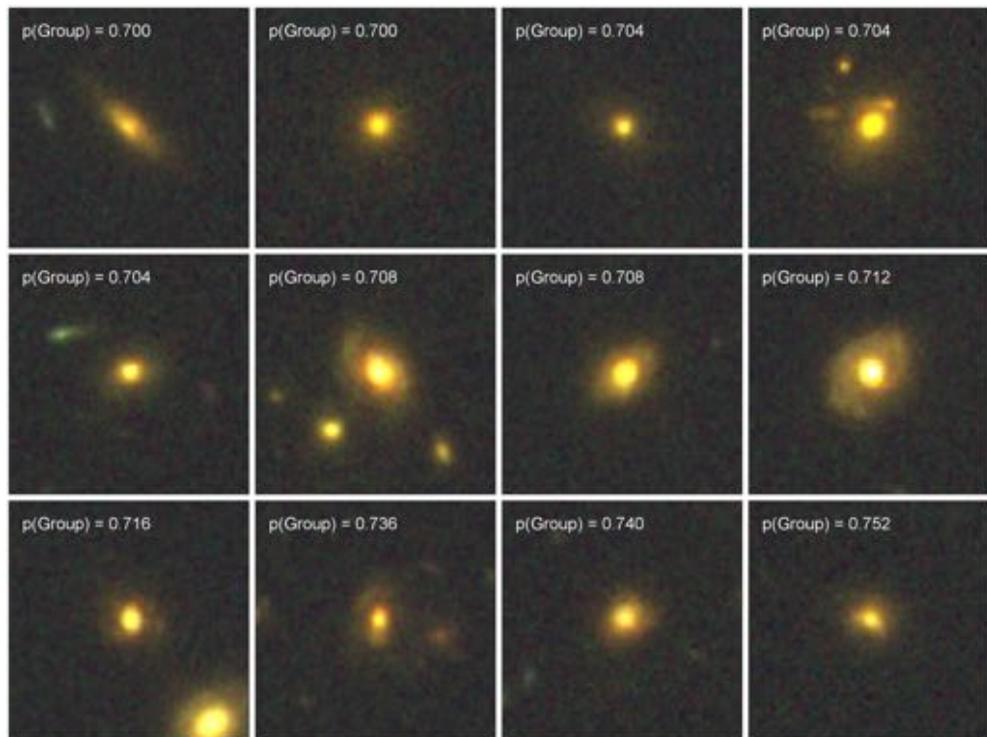


Figure 2.18 Group 0: These galaxies are characterized by a strong bulge component surrounded by a fainter smooth disk.

CHAPTER 2. PCA MORPHOLOGY

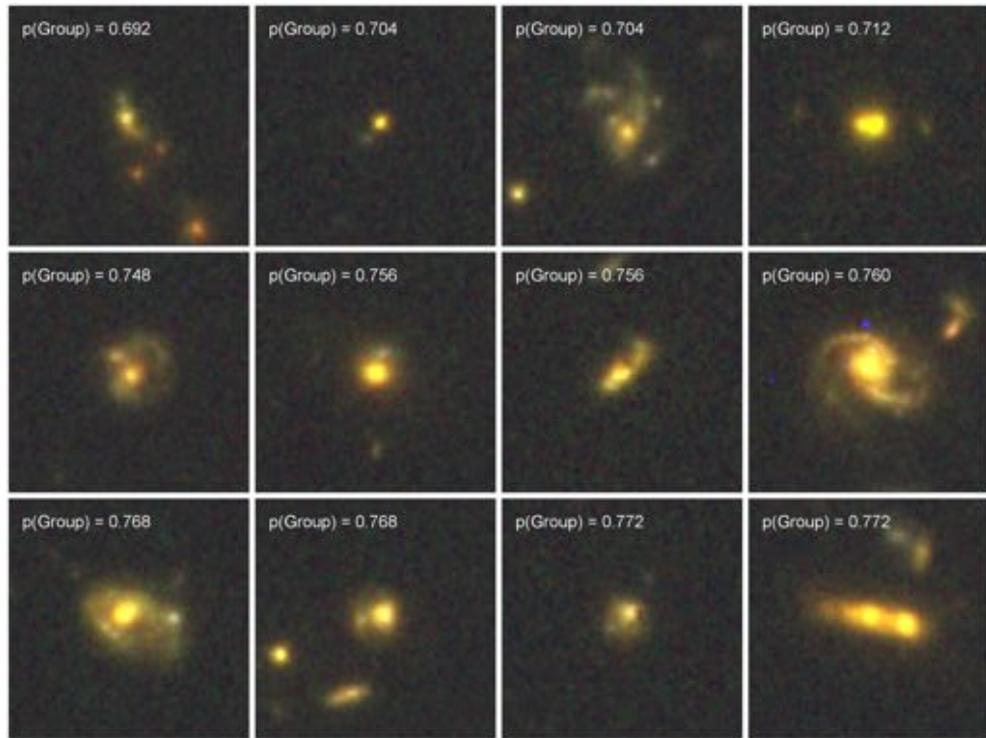


Figure 2.19 Group 9: These galaxies are characterized by their asymmetric, irregular morphologies and strong bulge component.

CHAPTER 2. PCA MORPHOLOGY

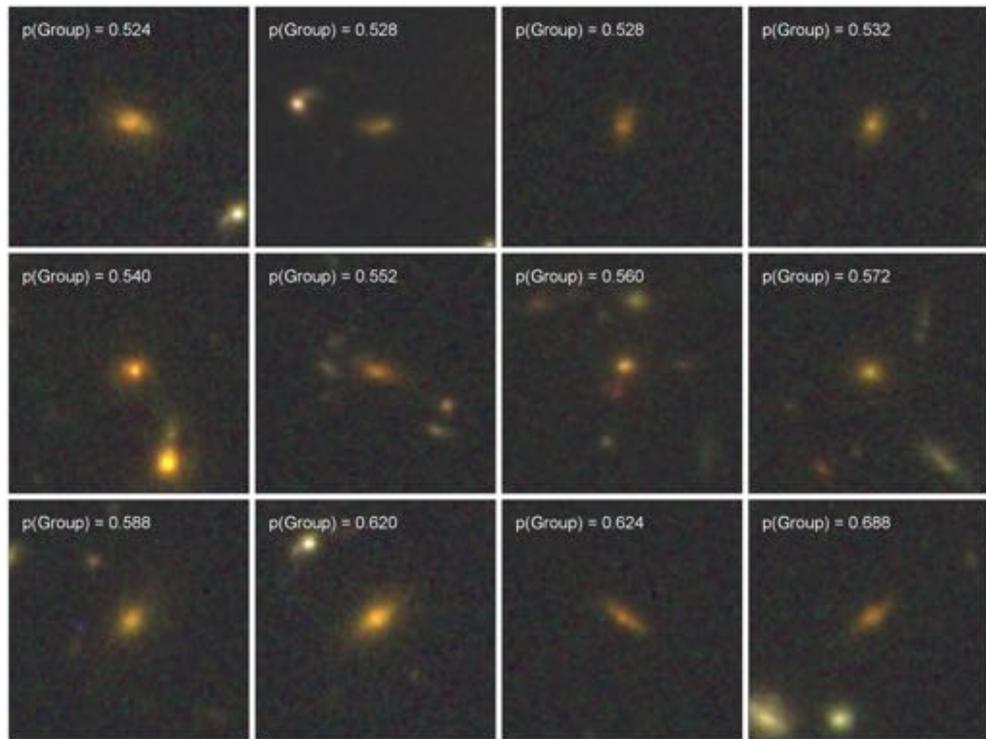


Figure 2.20 Group 4: These galaxies consist of low-mass smooth galaxies with moderate central concentrations.

CHAPTER 2. PCA MORPHOLOGY

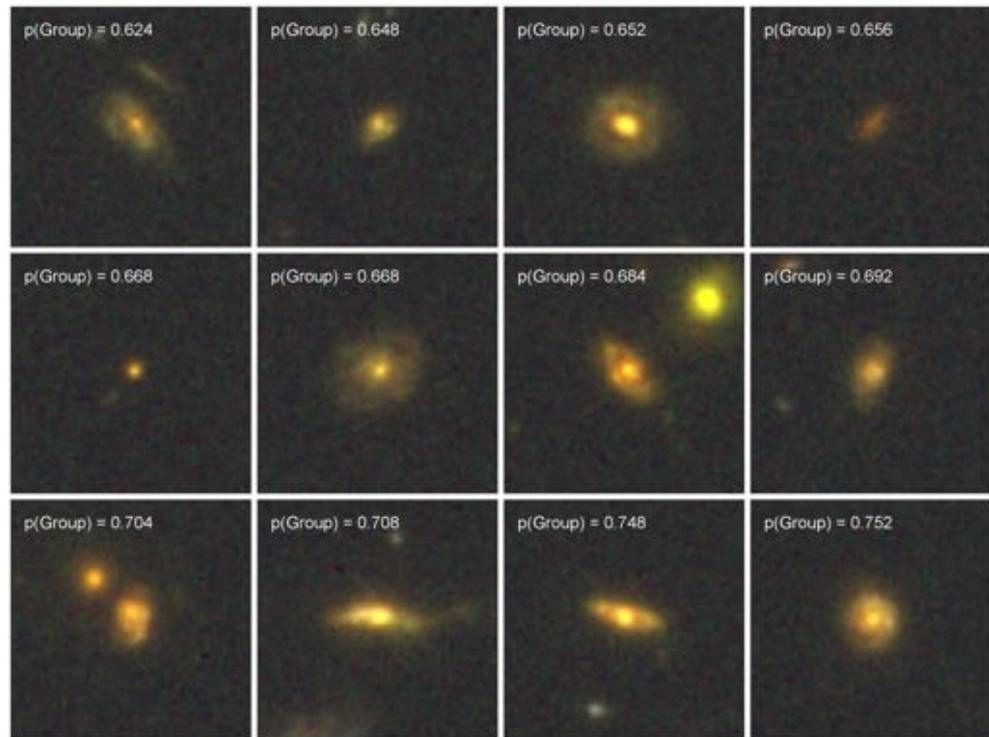


Figure 2.21 Group 8: These galaxies represent class of bulge+disk systems with dominant smooth disks.

CHAPTER 2. PCA MORPHOLOGY

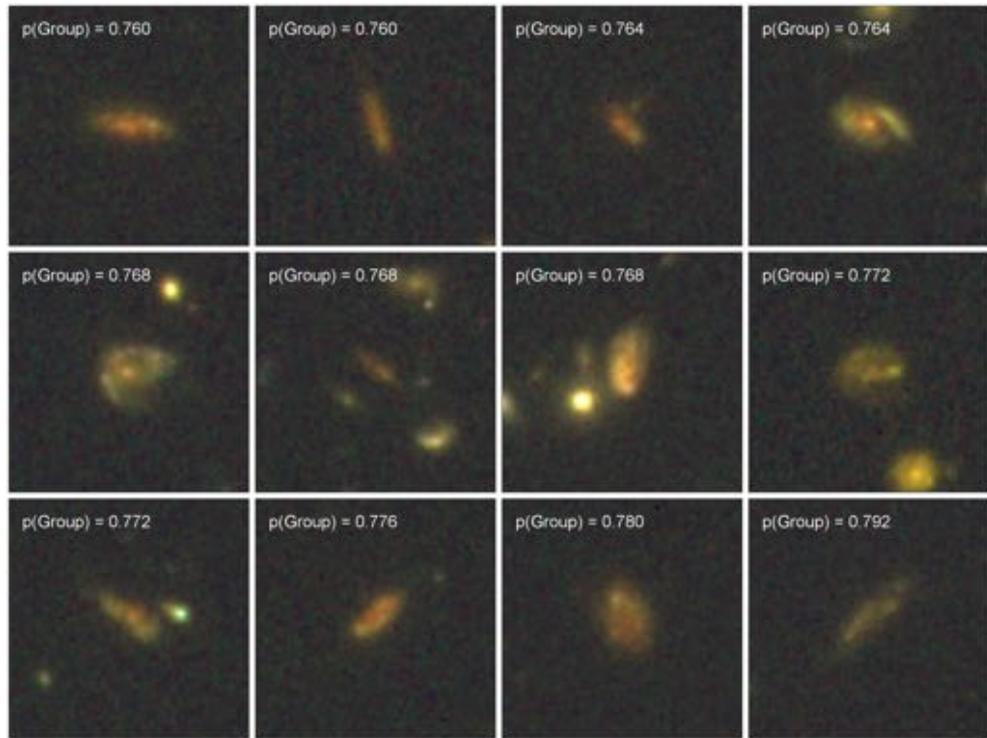


Figure 2.22 Group 1: These galaxies are primarily large disks and irregulars with bright off-center star-forming knots.

CHAPTER 2. PCA MORPHOLOGY

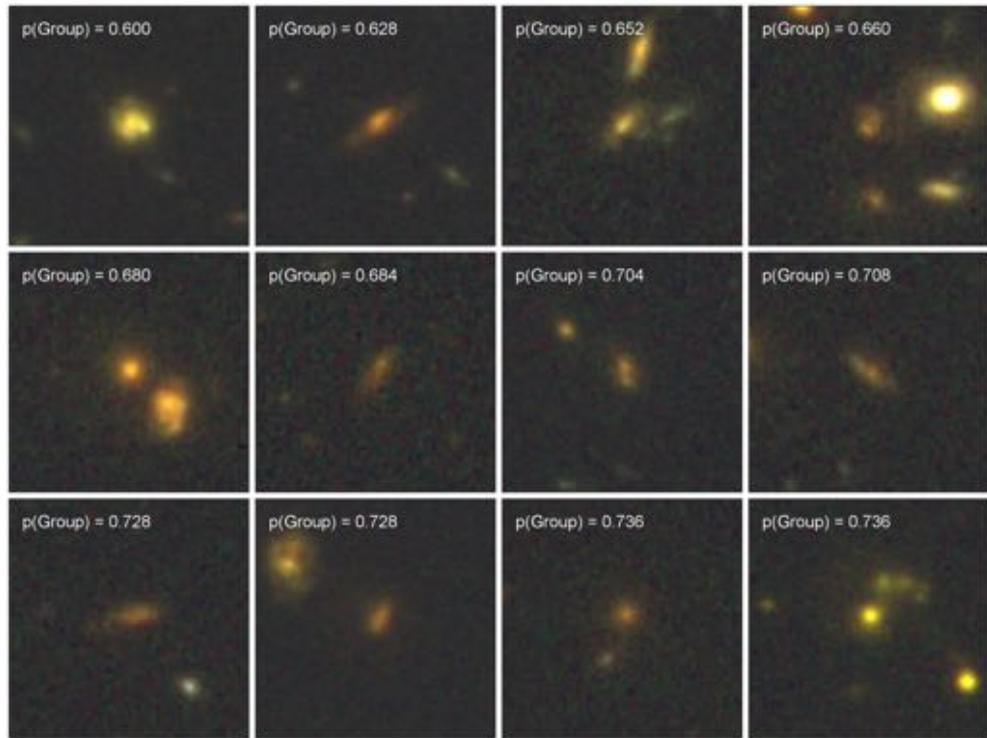


Figure 2.23 Group 2: These galaxies appear to be primarily low-mass star-forming disk galaxies with higher central concentrations and few detected star-forming knots.

CHAPTER 2. PCA MORPHOLOGY

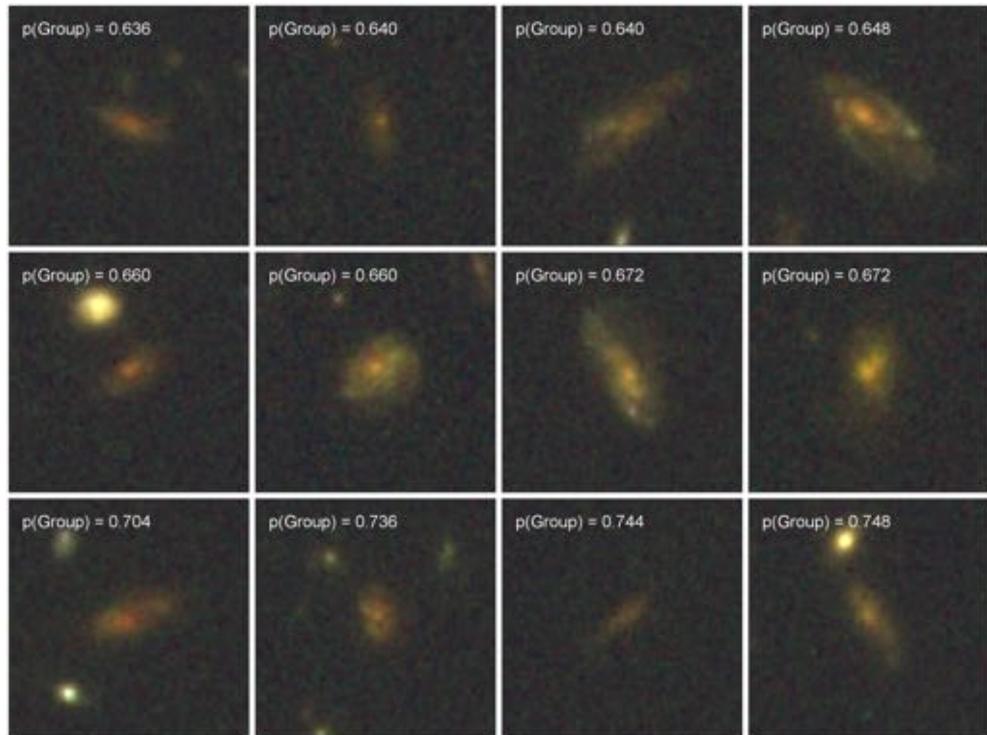


Figure 2.24 Group 5: Many of these galaxies are low-mass extended star forming disk galaxies with weak (if any) bulge components.

CHAPTER 2. PCA MORPHOLOGY

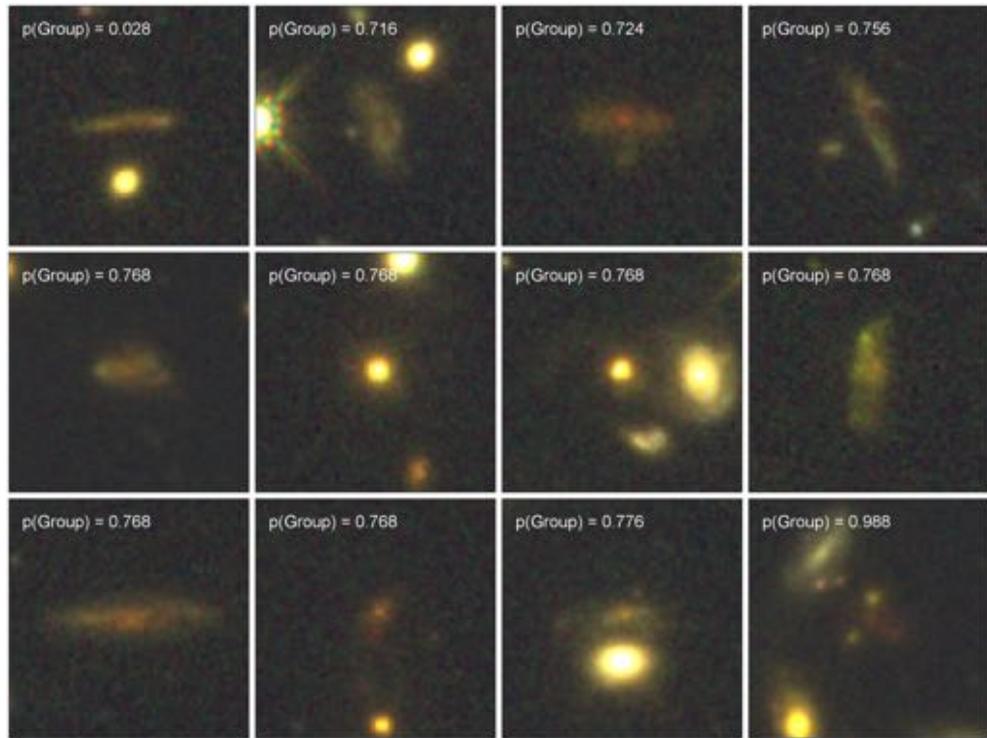


Figure 2.25 Group -1: Low surface brightness galaxies originally in groups 3 and 7, and only have a combined 19 galaxies which are outliers from all other groups.

Chapter 3

Merger Classifications of Pan-STARRS Galaxies Using Random Forest

3.1 Introduction

Morphology is transient and it is difficult to formulate a deterministic model of the physical mechanisms shaping a galaxy. These physical mechanisms include: major/minor mergers (e.g. Hopkins et al., 2009), violent disk instabilities (VDIs; Kereš et al., 2005), cold gas accretion (e.g. Brooks et al., 2009) and AGN/stellar feedback (e.g. Kauffmann & Haehnelt, 2000; Croton et al., 2006; Fabian, 2012). Each of these mechanisms leave behind morphological signatures: mergers can leave

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

behind tidal tails, violent disk instabilities lead to clumpy galaxies, star-forming disks and feedback quenching.

Hierarchical theories of galaxy formation (White & Rees, 1978) are based on a “bottom-up” growth of galaxies, where small proto-galaxies form first and merge with other proto-galaxies until they progressively become large enough to form the visible galaxies in the Universe. Mergers are very important to the formation and structure of the Universe in this hierarchical framework. In particular, mergers are related to the formation and evolution of galaxies (White & Rees, 1978), star formation (Kennicutt et al., 1987; Springel & Hernquist, 2003), nuclear activity (e.g. Engel et al., 2010; Hung et al., 2016)

Mergers come in all shapes and sizes and are defined by their mass ratios (major or minor) and their gas content (gas-rich or “wet” and gas-poor or “dry”). Major mergers (collisions between galaxies of roughly equivalent mass, mass ratio of $\lesssim 1:3$) can destroy disks by the gravitational interactions of the constituent galaxies and eventually reassemble into a relaxed spheroid (Toomre & Toomre, 1972; Barnes & Hernquist, 1996). Meanwhile, minor mergers (which are generally between galaxies with a mass ratio of $> 1:10$) may also disrupt morphologies, just not to the same dramatic extent as major mergers (Mihos & Hernquist, 1994; Lotz et al., 2011; Papovich et al., 2012). Galaxies with significant gas fractions interact leading to peculiar features (such as tidal tails, asymmetries, double nuclei, rings, shells) are typically visible (Toomre & Toomre, 1972). Major gas-rich galaxy mergers rapidly funnel gas

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

into the cores of massive galaxies; forming bulges, which then feeds the supermassive black hole (e.g. Sanders & Mirabel, 1996; Heckman et al., 2004; Naab et al., 2009). Gas-rich mergers provide a supply of star-forming fuel which can lead to starburst activity.

There are many different stages of mergers. A merger begins as two (or more) galaxies start to interact gravitationally. Initially, the the two galaxies are distinct objects but as galaxies approach each other dynamical friction between dark matter halos slows the relative orbital velocities. This orbital velocity decays and as it does tidal forces disrupt the morphology of both galaxies leading to structures known as “tidal arms”. The orbit of the satellite galaxies shrinks until it is totally disrupted or becomes assimilated into the central galaxy as the nuclei merge.

Visual classification studies starting with Hubble (1926) and continuing through more modern work (Lintott et al., 2008b, 2011; Kartaltepe et al., 2015) sought to discriminate disk-dominated, bulge-dominated and irregular galaxies. Citizen science projects such as (Darg et al., 2010a) use the general public to determine which galaxies are mergers. Visual classifications by human annotators are among best source of classifications possible since the human eye can identify very subtle features. However, these visual classifications can take months to complete for a small team of astronomers and days for Galaxy Zoo.

As a result, other studies have sought to use quantitative morphological measurements to identify merging galaxies (e.g. Lotz et al., 2008; Conselice et al., 2003) to

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

speed up the process of classification. However, since there are so many different types of mergers it is necessary to employ multiple different diagnostics to identify them (Conselice et al., 2000; Abraham et al., 1996). The concentration and asymmetry statistics can identify major mergers, while Gini and M_{20} can identify minor and major mergers (Lotz et al., 2004; Abraham et al., 2003). Freeman et al. (2013) has shown that multimode, intensity, and deviation (MID) are even more successful than previous non-parametric statistics recovering visually identified merger remnants (particularly at $z \sim 2$).

The stage of merger also have an effect on how long that particular structural feature (such as a tidal tail) which different diagnostic tools may be sensitive to. For instance, merging galaxies are only sufficiently asymmetric for about a third of the merger life-time (Lotz et al., 2008, 2010a,b)

We can not assume that all mergers result in spheroids , or vice versa (Robertson et al., 2005), so the only way to study if mergers can be an important mechanism in galaxy evolution is to find unambiguous examples of mergers in action. However, mergers in-action are quite rare, as clear signature have cosmically short lifetime (Lotz et al., 2008). Thus they require a large sample of galaxies to draw from. Non-parametric morphological statistics can be a useful way to automatically detect rare mergers. But to understand which statistics are more useful and important for identifying mergers we can not simply rely on a simple data exploration technique. We need to use machine learning techniques to really explore the data because merger

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

signatures are varied and subtle.

Previously, the Sloan Digital Sky Survey has been used to classify galaxies based on morphologies using automated machine learning tools (Ball et al., 2004, 2006, 2007, 2008a,b). Initially, these works made use of spectroscopic data for nearly 500,000 objects to train the entire SDSS DR3, over 143 million objects, to separate galaxies from stars. Later works used machine learning to calculate photometric redshift (and their probabilities) for the entire data set. Machine learning has been shown to perform tasks that humans never could complete on a reasonable timescale.

Machine learning techniques have also been used for higher redshift galaxies(Freeman et al., 2013; Huertas-Company et al., 2015; Kamdar et al., 2016a,b) to differentiate mergers from non-mergers and disks from ellipticals.

In this work, we use a machine learning tool, random forest, to classify local galaxies into mergers and non-mergers. Random forest classification determines the probability for classification and offers insight into the importance of each statistic used in the classification. This can lead to insights into the important physical mechanisms for galaxy evolution.

3.2 Data

3.2.1 Ground Based Surveys: Pan-STARRS and SDSS

The Panoramic Survey Telescope and Rapid Response Systems (Pan-STARRS; Kaiser et al., 2010) is a public survey covering 3π steradians of the Northern hemisphere in 5 optical filters (*grizy*) to a depth of $g = 23.8$, $0.6''$ point spread function full width half-maximum (PSF FWHM) (Kaiser et al., 2010; Tonry et al., 2012). Additionally, there are 10 Medium Deep Survey fields imaged ~ 3 magnitudes deeper than the main survey (Lin et al., 2014). The scientific results from Pan-STARRS are just beginning to be realized and in the next few years the survey will provide the community with exciting science.

The footprint of Pan-STARRS overlaps with the shallower Sloan Digital Sky Survey (SDSS, York et al., 2000b). SDSS covers $11,000 \text{ deg}^2$ of sky to a depth of $r \sim 22.5$ (Abazajian et al., 2009). The SDSS uses a 2.5-m wide-field telescope (Gunn et al., 2006) located at Apache Point Observatory in New Mexico (Abazajian et al., 2009).

Pan-STARRS observations are deeper than SDSS, thus will provide the best resource to measure galaxy morphology. Galaxies from the Pan-STARRS Observed Galaxies Survey (POGS) were selected in two parts: the first set of galaxies represent the sample of POGS galaxies visually identified as mergers sample by Galaxy

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

Zoo (see §3.2.2 for more). The second set of galaxies represent a random selection of Pan-STARRS observed galaxies with all measured morphologies and Galaxy Zoo visual classifications (from SDSS images). The second sample was specifically chosen to not include any merging galaxies or not be contaminated by any foreground stars. Both of these criteria create a pure sample of non-merging galaxies to combine with the merger galaxy sample.

Additionally, we apply random forest criteria to independent sample of Pan-STARRs observations of Mapping Nearby Galaxies at APO (MaNGA) survey galaxies which is one of the core SDSS-IV projects (Bundy et al., 2015). MaNGA is a spectroscopic survey which will measure kinematics of gas and stars for nearly 10,000 nearby galaxies. Spatially resolved spectra will allow for more precise measurements of star-formation in galaxies and the study the growth and assembly of disks and bulges through gas accretion, mergers and secular processes. The combination of deep photometric imaging from Pan-STARRS and the spatially resolved spectral imaging from MaNGA will provide greater ability to relate galaxy morphology with star-formation and the mechanisms influencing both.

3.2.2 Galaxy Zoo

Galaxy Zoo is a citizen-science project that employs online crowdsourcing to classify SDSS galaxies (Lintott et al., 2008b, 2011). Users are asked a series of questions leading them down pre-defined decision pathways concluding in very large sets of

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

visually classified galaxies. Lintott et al. (2008b) defines a weighting factor to lessen the impact of unreliable users. Many other disciplines and media are able to employ crowdsourcing to effectively classify objects. Zooniverse has projects ranging from Shakespeare’s handwritten documents to analyzing videos of chimp behavior in addition to many astronomical projects.

Figure 3.1 from the Galaxy Zoo Data Visualization tool¹ (Willett et al., 2013) shows the decision tree a user is presented with to classify galaxies. The user is shown an image of a galaxy and asked a series of questions following the tree until the end of the tree. Information about all the user choices from the decision tree is retained and analyzed, which eventually lead to the likelihood score a galaxy is a particular morphological type.

There have been multiple projects classifying galaxy morphology (Willett et al., 2013), mergers (Darg et al., 2010a,b), AGN host galaxies (Schawinski et al., 2010) galaxy pairs (Keel et al., 2013), “green pea” compact galaxies (Cardamone et al., 2009) and galactic bars (Hoyle et al., 2011). Depending on the goals of a particular project the classification questions asked of the user are different. In general, the questions begin by asking if galaxies are spirals or spheroids, and subsequently become more specific. These projects use the online users to classify the morphologies of 893,212 SDSS DR6 (spectroscopically and/or photometrically confirmed) galaxies with a Petrosian magnitude $r < 17.7$ (Lintott et al., 2008b).

¹http://data.galaxyzoo.org/gz_trees/gz_trees.html

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

For our study we use the merger catalog of Darg et al. (2010a). In this Galaxy Zoo project the user is asked to classify every galaxy as elliptical, spiral, star/bad image or merger. Every galaxy is classified between 40–80 times and have a weight-corrected classification percentage for each morphological type. In this work, a single value f_m was defined to represent the probability that the galaxy is experiencing a merger. The parameter f_m is from 0 to 1, with 0 representing a galaxy with absolutely no merger-remnant morphology and 1 representing an unmistakable merger remnant. After visually classifying a subset of the sample themselves, Darg et al. (2010a) determines the public was quite conservative with their merger classifications and define $f_m > 0.4$ to represent ‘strongly perturbed’ systems. Their sample is drawn from a volume limited sample of 304,812 SDSS DR6 galaxies with spectroscopic redshift in the range $0.005 < z < 0.1$ and brighter than $M_r < -20.55$ at $z \sim 0.1$ leading to a sample of 4,198 mergers.

They determine \sim 6–9% of all galaxies in their sample are experiencing a merger, and that \sim 2–4% of all galaxies in their sample are experiencing a major merger. Mergers are three times more likely to found in spiral galaxies as ellipticals, potentially mergers are detectable longer in spirals than ellipticals. Correspondingly, galaxies with a high gas fraction are more likely to be disturbed and possess disturbances for longer times since disk galaxies typically have more gas than ellipticals.

The follow-up study (Darg et al., 2010b) found no dependence of merger fraction on environment. However, ellipticals in mergers are generally redder and more massive

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

than spirals in mergers. Merging galaxies in general are more massive than non-mergers. Mergers appear to enhance star-formation only in spirals but do not enhance nuclear activity for any morphology. For a contradictory result see (Ellison et al., 2011).

3.3 Non-parametric Morphology of Pan-STARRS Galaxies

We begin with the “white light” image of a PANSTARRS galaxy (Thilker et al., 2014) which is then fed into SExtractor (Bertin & Arnouts, 1996) which segments the image into constituent parts. The white light image is a composite of g , r , i , z , and y ($400 \text{ nm} \leq \lambda \leq 1000 \text{ nm}$) band images (Kaiser et al., 2010). We choose to perform segmentation of the white light image because these are the deepest images.

We use a carefully tested set of SExtractor inputs to ensure large galaxies are not segmented into many object and obvious stars are not included in the segmentation map of a galaxy. Additionally, background objects are masked out. Once we have an initial segmentation map and SExtractor catalog containing image positions we can use our morphology code to measure the non-parametric statistics. The morphology code uses the SExtractor segmentation map and catalog as a first step in an iterative process to determine the position of the galaxy center and the pixels within the Petrosian radius used to calculate non-parametric structure statistics. While the

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

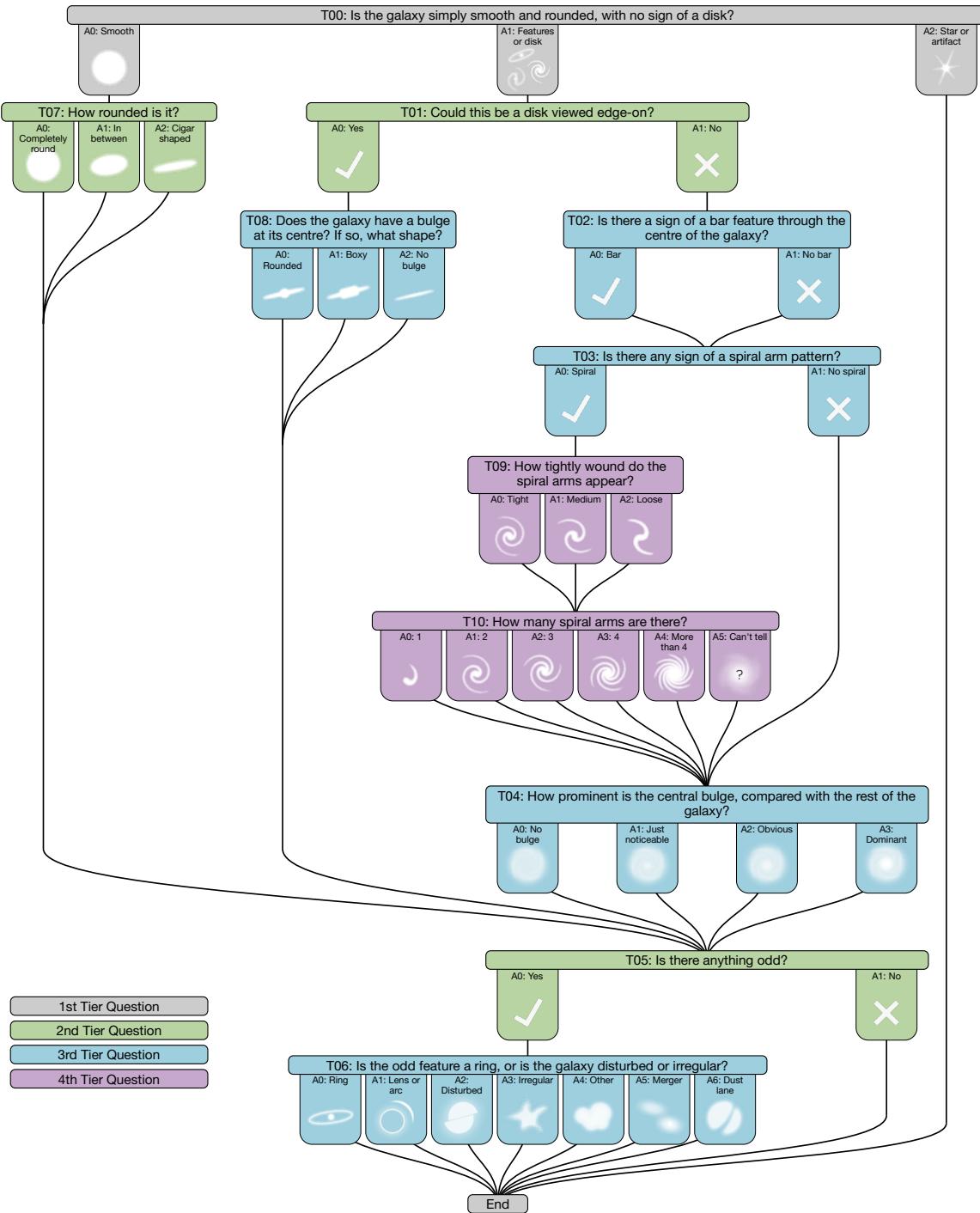


Figure 3.1 Galaxy Zoo Decision Tree for classifications, from Willett et al. (2013).

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

segmentation map used is based on the white light image, we measure morphology upon the g -band image. This ties the local population of galaxies with our higher redshift sample from Chapters 2 and 4, since those redshift ranges were chosen to correspond to rest-frame g -band. Only UV/rest-frame blue light is visible with HST at high redshift.

We use the quantitative morphology code (Lotz et al., 2004, 2008; Peth et al., 2016 and Lotz et. al, in prep) to measure the non-parametric morphological statistics. The SExtractor output catalog and segmentation map provide the initial guess for the extent of a galaxy. The Petrosian radius is measured for increasing elliptical apertures and is determined by the curve of growth within these apertures (Lotz et al., 2004). The flux center and radii determined by the Petrosian radius calculation are then used to calculate the asymmetry value. The Petrosian radius is then recalculated using the center of asymmetry to better capture the true nature of the galaxy. At which point, the asymmetry is recalculated a final time using the updated Petrosian radius. Concentration, Gini and M_{20} are all calculated using the Petrosian radius and asymmetry center. To calculate the multimode, intensity and deviation statistics anew segmentation map is calculated using the algorithm of Freeman et al. (2013).

In Figures 3.2 – 3.4, the non-merger sample is shown in the gray contours and the merger sample is split into merged and pairs. A merged galaxy in this definition is when SExtractor creates only one single unbroken segmentation map for the merger. Meanwhile, a pair is when there are at least two SExtractor segmented galaxies in the

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

merger catalog.

Figures 3.2 - 3.4 shows the distribution of non-merger galaxies (grey contours) and the merger galaxies color coded by the galaxy zoo vote (the percentage of merger votes) for Gini- M_{20} , concentration-asymmetry, and multimode-deviation. The Gini- M_{20} diagnostic has a specificity of 95%, but a completeness of only 24%. Galaxies correctly identified as mergers by Gini- M_{20} have a slightly higher average merger vote (0.63 vs 0.59) than mergers not positively classified by Gini- M_{20} . Meanwhile, the concentration-asymmetry diagnostic has a specificity of 98%, but a completeness of only 15%. Galaxies correctly identified as mergers by concentration-asymmetry have an equal average merger vote (0.63 vs 0.63) compared to mergers not positively classified by concentration-asymmetry. Additionally, the Multimode-deviation diagnostic has a specificity of 71% and completeness of 38%, and correctly classified galaxies are not any more likely to have high merger votes than incorrectly classified mergers (0.60 vs 0.60). All of these indicators are quite successful at creating a pure but incomplete sample of mergers.

3.3.1 Principal Component Analysis of PanSTARRS Galaxies

We apply the Pan-STARRS merger/non-mergers morphologies to our PCA (defined at $z \sim 1.5$) to understand how the PC groups correspond to morphology for different epochs. Figure 3.5 shows a large concentration of galaxies in a portion of PC space not explored at high redshift. These galaxies are grouped into group -1, which was previously the purview of

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

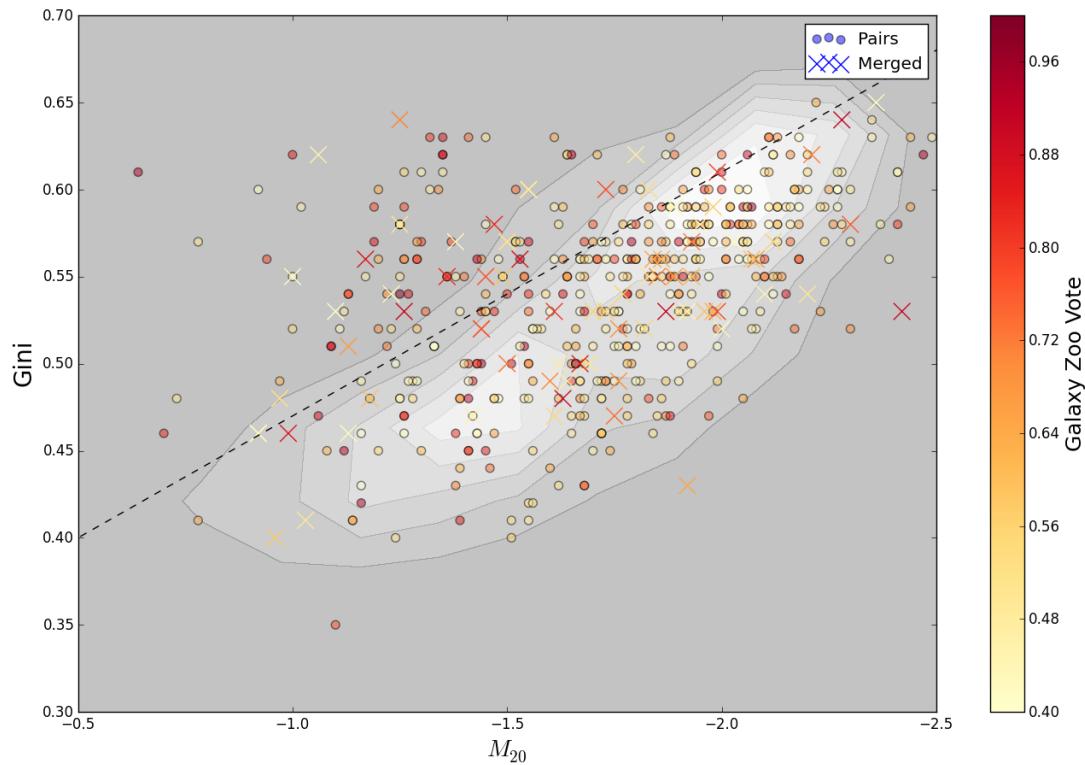


Figure 3.2 Gini - M_{20} diagram for mergers color coded by the Galaxy Zoo merger vote score with filled circles representing galaxies in pairs and x's represent merged systems. The grey contours represent galaxies classified as non-mergers by Galaxy Zoo.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

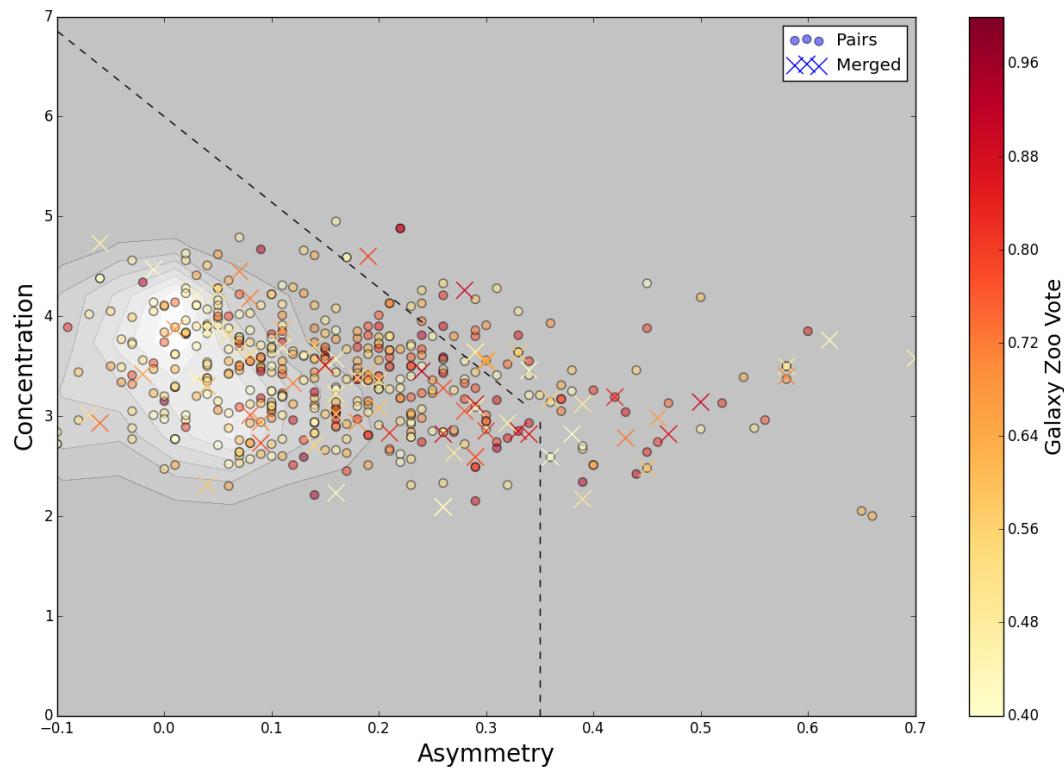


Figure 3.3 Concentration - Asymmetry diagram for mergers color coded by the Galaxy Zoo vote score with filled circles representing galaxies in pairs and x's represent merged systems. The grey contours represent galaxies classified as non-mergers by Galaxy Zoo.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

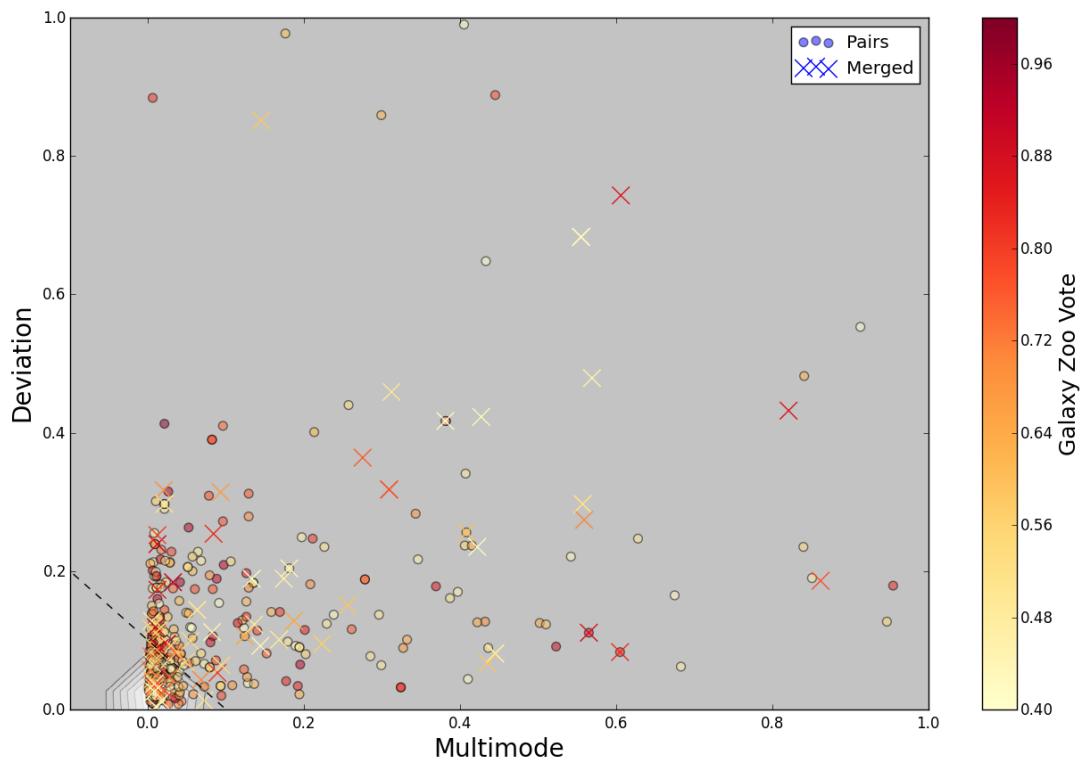


Figure 3.4 Multimode - Deviation diagram for mergers color coded by the Galaxy Zoo vote score with filled circles representing galaxies in pairs and x's represent merged systems. The grey contours represent galaxies classified as non-mergers by Galaxy Zoo.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

very irregular galaxies with faint tidal features. Figure 3.6 shows that low redshift galaxies are not classified as group 6 very often, but rather are more likely to be classified into group 0 or -1. Non-merger galaxies, in particular, are more likely to be classified as group 0 or -1. The transition from group 6 to group 0 as the most populated group is largely a function of increased resolution. In the PanSTARRS galaxies, the faint disk enveloping a central bulge is more easily visible than at high redshift. Since merger pairs are galaxies with burgeoning tidal features or disturbances their similarities to non-mergers should not be totally surprising. Many of the merged galaxies were classified into group 9 at $z \sim 1.5$ appeared to be a transitional stage between bulge and disk-dominated regimes with high disturbances.

There is a large sample of group -1 galaxies which are large disk galaxies. This group represents outliers in PC space based on the definitions of $z \sim 1.5$ galaxies. However, the tightness and proximity of all these data points suggests this outlier class represents a single morphological class. Since these types of galaxies are not present at high redshift we may be witnessing the advent of a new morphological type or surface brightness dimming at high redshift could

PCA can not truly quantify how likely a specific galaxy is a merger. To be able to quantify a merger likelihood a supervised machine learning technique that has been trained to separate mergers from non-mergers is required. In the next section, we define our implementation of the supervised machine learning algorithm, random forest classification, to distinguish mergers from non-mergers.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

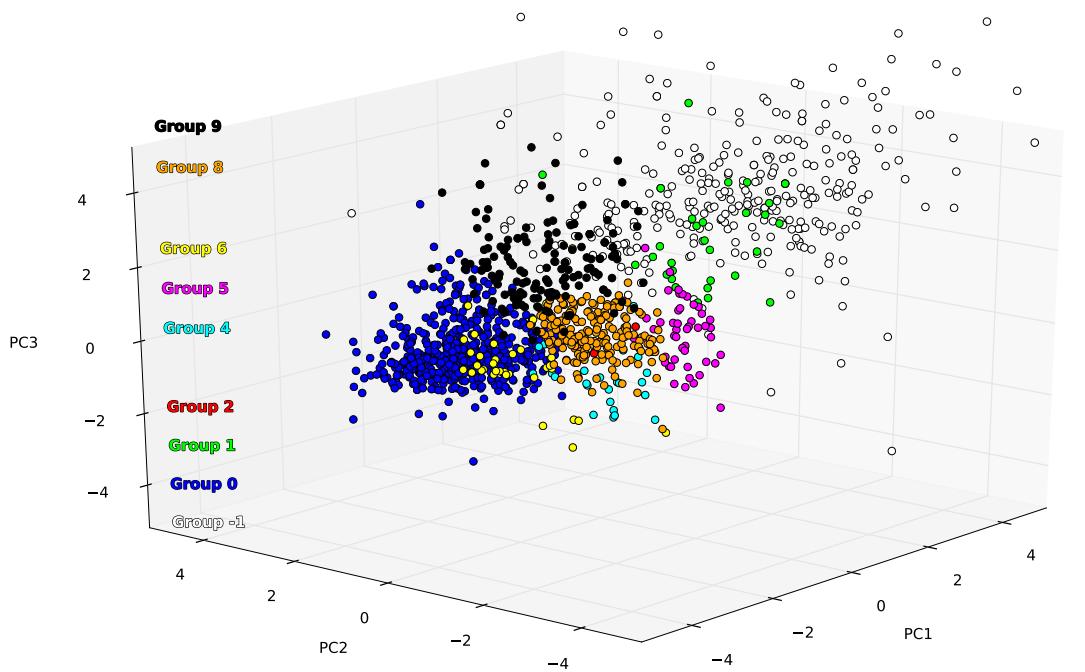


Figure 3.5 PC1-PC2-PC3 plot for PANSTARRS galaxies color coded by PC group.

There are many more examples of group=-1 galaxies, which visually are large disk galaxies. The proximity of group -1 galaxies suggests they are in fact a new group and not merely just the outliers of all other groups.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

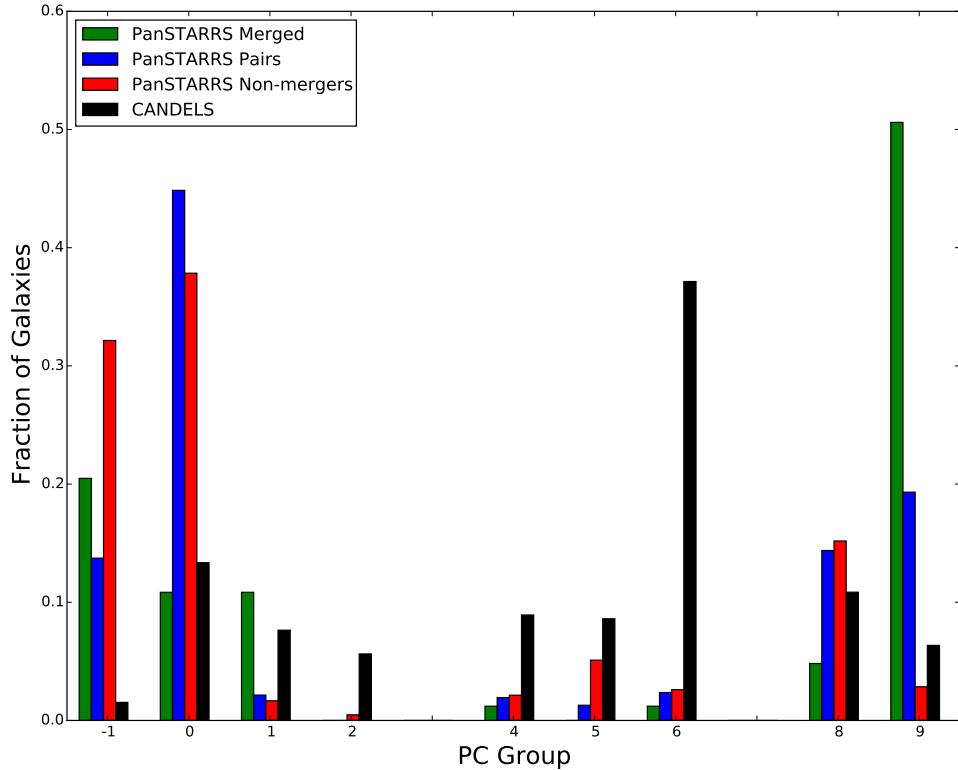


Figure 3.6 Histogram of PC groups for CANDELS and PanSTARSS (merged, merger pairs and non-merger) galaxies. The groups are defined as disk-dominated (1,2 and 5), intermediate (4 and 8) and bulge-dominated (0, 6 and 9). Group 9 is the preferred group for merged system, whereas merger pairs and non-mergers are primarily groups 0 or -1. The vast fraction of galaxies in group -1 suggest an evolution in morphology in galaxies from high to low redshift.

3.4 Random Forest Classifier

The random forest technique was developed by Breiman (2001) as a supervised method for classification. The initial sample of galaxies is divided into a training and a test set, containing 67% and 33% of the original galaxies respectively. A visual representation of the tree nature of the random forest is shown in Figure 3.7. A total number of n trees are created, each containing a random sampling (80%) of the training set that are used to define the splitting characteristics differentiating mergers from non-merger galaxies.

The split best differentiating mergers from non-mergers among the random subset of the features in each node defines the optimal classifiers, an example of which is visualized in Figure 3.8. The decision trees use bootstrap samples of the remaining data and a random selection of features are chosen (in our case 3 features) to create the decision trees at each branch. In the example shown in Figure 3.8 three parameters (G, M, and A) are selected randomly and the change in Gini² impurity value is used to define the splitting criteria for the A statistic. In this context the Gini impurity measures the probability that a merger would be incorrectly classified if the data point and label were both randomly created. This Gini impurity is calculated by Equation 3.1 with p_i representing the probability of finding k data points in a specific class.

This process of determining the best statistic to divide the sample repeats along the branches of the tree until either all galaxies have been classified into a pure sample or the maximum terminal (leaf) node is reached. This process is repeated to create a “forest” of decision trees. The final classification is the average output from each decision tree (Ivezić

²This not to be confused with the Gini coefficient from the morphological measurement.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

et al., 2013; Kamdar et al., 2016a). We use the `scikit-learn` random forest routine for our classifications.

Random forest returns the importance of each input parameter into the classifications as sum of impurity ($\sum \Delta I_m$ for each feature) over all nodes and trees. This provides a degree of importance for each parameter that can be compared to the results from PCA. Previous studies have shown that the M , I , and D statistics are the best statistics at differentiating regular and non-regular galaxies as well as mergers from non-mergers (Freeman et al., 2013).

$$G = \sum_i^k p_i(1 - p_i) \quad (3.1)$$

In addition to classifications, previous works have used random forest to predict galactic parameters from pure dark matter halo properties (Kamdar et al., 2016a,b). These works were able to calculate galactic properties such as stellar mass, metallicity, star-formation rates, etc. simply from properties of the dark matter halo such as dark matter mass, number of dark matter particles, circular velocity, etc. Stellar masses and metallicities were best predicted following the random forest technique.

The random forest classifications are evaluated based on the completeness (Equation 3.2), specificity (Equation 3.3), risk (Equation 3.6), total error (Equation 3.7), positive predictive value (PPV or purity; Equation 3.4), and negative predictive value (NPV) of the classification results when compared to the original labels (e.g. Freeman et al., 2013; Ball & Brunner, 2010) . The goal of any classification scheme is to maximize/minimize these measures, however this is not always feasible. When comparing the random forest classifications to the real classifications there are true positives TP (classification agrees

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

with initial label and in correct class), false positives FP (objects not in a class but have incorrectly classified as such), true negatives TN (classification agrees with initial label and in null class) and false negatives FN (the number of objects in a class incorrectly classified as not belonging to that class).

The importance of completeness and PPV is set based upon the problem being solved (Ball & Brunner, 2010). In our case the purity of our resultant classifications is more important than the completeness (however this is not to say completeness is not extremely necessary).

$$\text{Completeness} = TP / (TP + FN) \quad (3.2)$$

$$\text{Specificity} = TN / (TN + FP) \quad (3.3)$$

$$\text{PPV} = TP / (TP + FP) \quad (3.4)$$

$$\text{NPV} = TN / (TN + FN) \quad (3.5)$$

$$\text{Risk} = 1 - \text{completeness} + 1 - \text{specificity} \quad (3.6)$$

$$\text{Total Error} = (FN + FP) / N \quad (3.7)$$

During classification the out-of-bag error for each data point is recorded and averaged over the forest (Breiman, 2001). The more often a feature is used at each node to split

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

points of a tree the more important that feature is. The importance of a feature after training is measured when the values of the feature are randomized among the training data and the out-of-bag error is again computed on these new data sets. The importance score for a feature is computed by averaging the difference in OOB errors before and after the randomization over all trees. Features with large importance values are more important than features with small values.

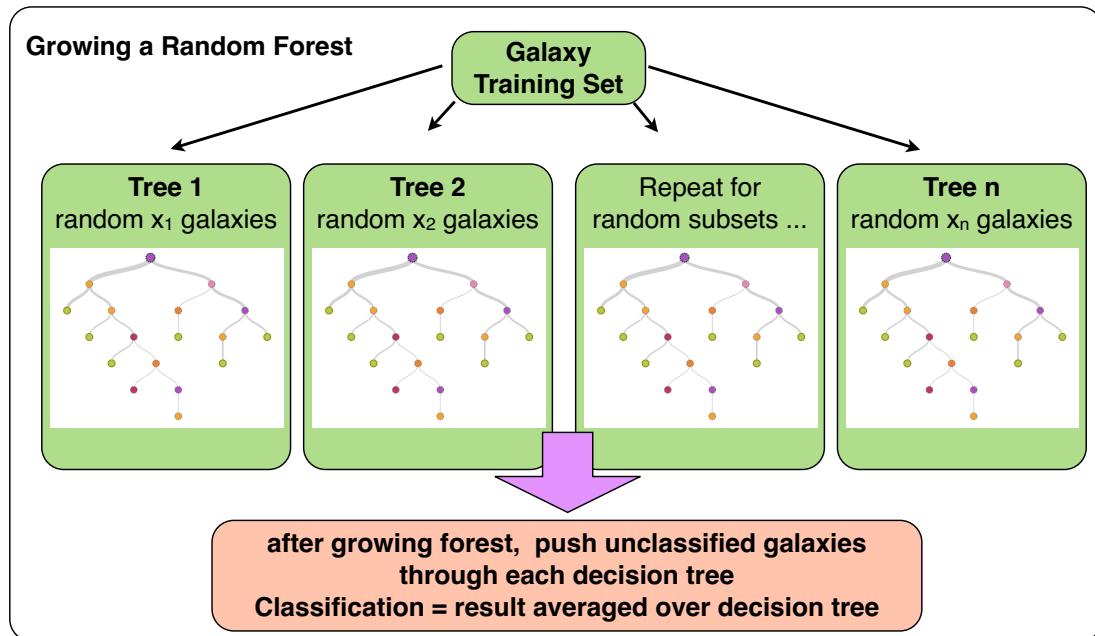


Figure 3.7 Visualization of how the training set is utilized to create a “forest” of decision trees that are used for classification (Lotz et al., in prep).

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

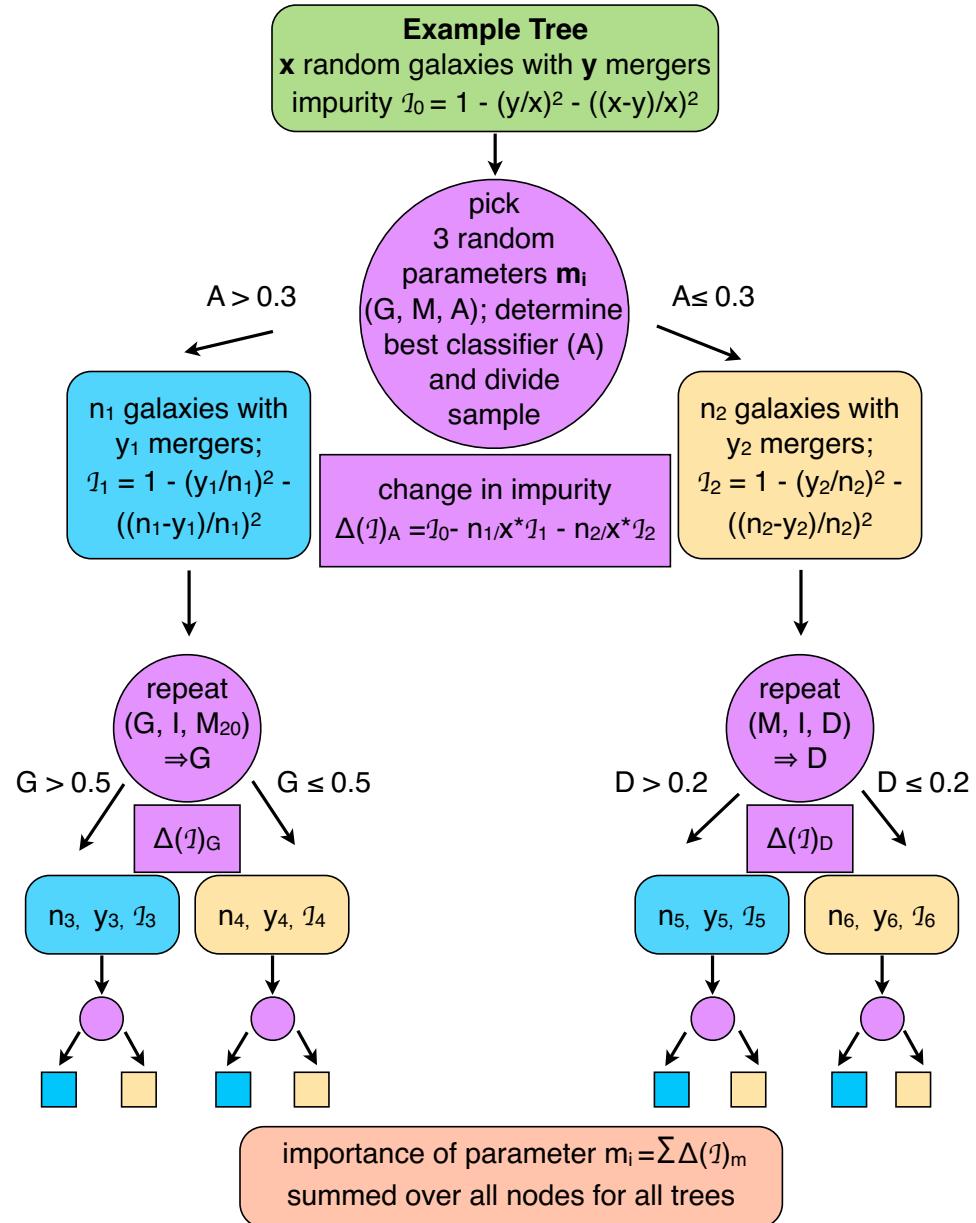


Figure 3.8 Visualization of how each tree in the random forest decides how to divide the sample of galaxies for training purposes (Lotz et al., in prep.).

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

3.4.1 Random Forest Inputs

using a Conroy et al. 2009 dust model)

We use all 7 PCs, concentration, asymmetry, the Gini coefficient, M_{20} , $d(G, M_{20})$, $F(G, M_{20})$, multimode, intensity, deviation along with $g-r$ rest-frame color, the specific star-formation rates and stellar masses from the SDSS MPA-JHU DR7 spectroscopic value added catalog³, the merger/non-merger classification from Galaxy Zoo SDSS/Pan-STARRS galaxies as our input to train the random forest and predict a merger/non-merger classification. Our random forest is grown to include a maximum of 100 leaf nodes, 500 trees, and use 3 features per branch for decision purposes. We tested the effect the number of estimators and maximum leaf nodes have on the summary statistics and OOB (out-of-bag) score. See §3.5.1.1 for more on these tests.

The $d(G, M_{20})$ statistic measures the distance in $G-M_{20}$ space of a galaxy to the merger/non-merger dividing line, and is defined in Equation 3.8 (Snyder et al., 2015b). This statistic has been shown to be a good indicator of merger activity.

$$d(G, M_{20}) = \frac{| -0.14M_{20} - G + 0.33 |}{0.14} \quad (3.8)$$

We also grow a forest using only using the most basic non-parametric morphological measurements, concentration, asymmetry, the Gini coefficient, M_{20} , $d(G, M_{20})$ and $F(G, M_{20})$ (see Equation 2.15), to test the amount of information and classification strength would be lost from an implementation of random forest without *MID* or PCs.

³http://www.sdss3.org/dr9/algorithms/galaxy_mpa_jhu.php

3.5 Results

3.5.1 Random Forest Classifications

Table 3.1 shows the feature importances for the full sample. The most important feature to determine merger classification is, unsurprisingly, asymmetry. For the full sample the next most important features are PC7, deviation, $d(G, M_{20})$, PC5, PC3 and PC2. Deviation as a statistic is analogous to asymmetry and PC3 is heavily dependent on asymmetry, so the fact that all of these statistics are important to differentiate mergers from non-merger galaxies is not entirely surprising (see Chapter 2, Table 2.1). However, the heavy importance of PC7 on the classifications is. PC7 is the least important principal component at $z \sim 2$ (and only captures 5% of the total variance of the data), but has such a strong effect upon the merger classification. PC7 is dominated by the correlation between Gini, M_{20} and multimode which are anti-correlated with concentration, intensity and deviation.

We tested the results of using a smaller set of morphological statistics, in case a user only had access to concentration, asymmetry, Gini, M_{20} , $d(G, M_{20})$ and $F(G, M_{20})$. We wanted to understand how much specificity and completeness of the classifications would be lost by not measuring the *MID* statistics or PC values. Table 3.2 shows that even when we use a condensed number of features, the importance of said features are nearly identical to a full run. Asymmetry is still by far the most important statistic, and the $d(G, M_{20})$ is a distant second. The other non-parametric morphology statistics are still not very important.

Meanwhile, Figures 3.16 – 3.18 show the comparison of feature importances dependent on the sample used. Error bars are determined by the standard deviation of importances

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

after 1000 random forest runs. We previously noticed that asymmetry is by far the most important feature for classifying mergers. Figure 3.16 shows that the samples of red and blue galaxies (divided by $g - r > 1.5$) show that features have different importances depending on the nature of a sample. Blue galaxy merger classifications are more dependent on PC7 and PC2 than red galaxies. Meanwhile red galaxy merger classifications are more dependent upon PC3, deviation, PC5 and M_{20} .

Figure 3.17 shows the comparison of feature importances between the merged and pair galaxy samples. Merged galaxies are far more dependent upon the multimode statistic than pairs. The multimode statistic measures the size (in pixels) of the two brightest regions. A merged galaxy will be visible because of the numerous bright regions within a single segmentation map. The remainder of the features (except for asymmetry and intensity) are slightly more important for merging pairs than merged galaxies.

Figure 3.18 shows that the importance in classifying blue galaxies or merging pairs is very similar to the full sample. Most galaxies in the full sample are blue and/or merging pairs, so the agreement between these should not be surprising.

Figure 3.9 shows the Gini– M_{20} but now color coded by the fraction of times a galaxy is classified as a merger following a random forest. The correlation between classification fraction and Gini– M_{20} is a function of random forest itself. Random forest finds the subspaces and divides those spaces to create classifications. Similarly in Figure 3.11 the concentration-asymmetry subspace is correlated with the classification fraction for the same reason. The random forest classifications in both Gini– M_{20} and concentration-asymmetry subspaces show that the previously defined merger/non-merger dividing lines need to be

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

adapted.

Figure 3.10 shows random forest merger/non-merger classification confusion matrix in Gini– M_{20} subspace. Galaxy zoo classified mergers are classified correctly above the Gini– M_{20} relation line, and incorrectly below, and vice versa for non-mergers. Similarly, Figure 3.12 shows random forest merger/non-merger classification confusion matrix in C–A subspace. The Gini– M_{20} merger dividing line has a 3% false positive rate but a 30% false negative rate. The C–A merger dividing line has a 1% false positive rate but a 36% false negative rate. The Gini– M_{20} and C–A merger diagnostics are known to minimize the number of false positives but not for minimizing false negatives (Lotz et al., 2008). The training set is not a statistical representative of the Universe, since we include a disproportionate amount of merging galaxies.

3.5.1.1 Random Forest Input Parameter Tests

At every node in a tree 80% of the training set is divided into merger and non-merger categories depending on 3 randomly selected statistics. The remaining portion of the training set is then classified. The amount of error in classification is captured in an out-of-bag estimator (OOB, James et al., 2014). This process is repeated as you traverse the tree downwards until either the final nodes are pure or the maximum number of nodes has been reached. We test the effects of different number trees in the forest (also known as estimators) and maximum leaf nodes and show the results on the OOB score in Figure 3.13 and the results on the summary statistics (completeness, specificity, etc.) in Figure 3.14. The OOB score decreases with increasing maximum leaf nodes but past 100 leaf nodes does

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

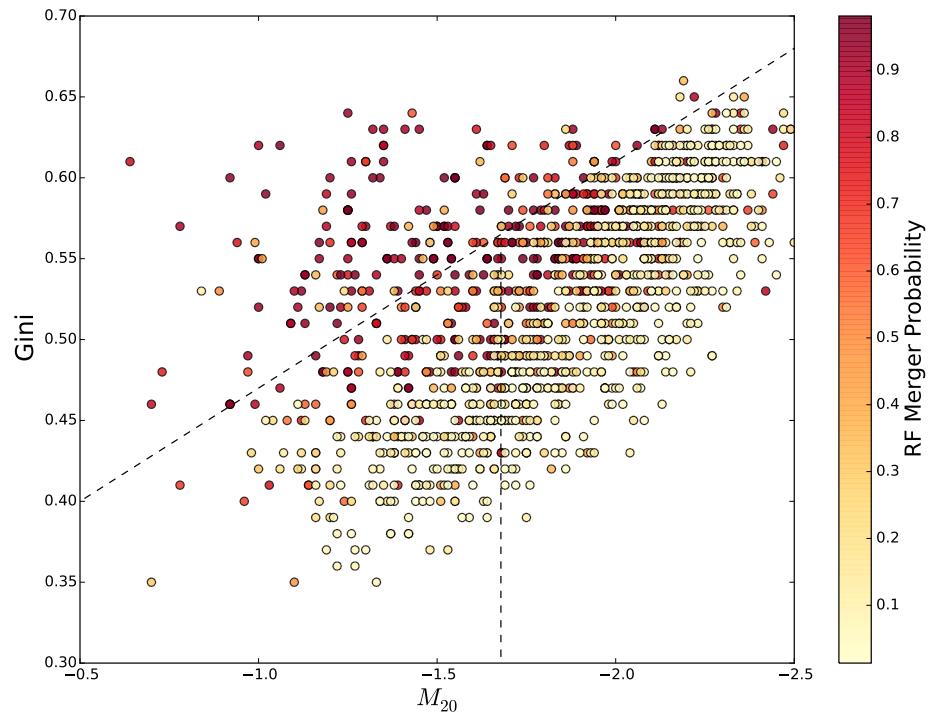


Figure 3.9 Gini- M_{20} color-coded by average RF Classification Probabilities following 1000 iterations of random forest. The dividing line of (Lotz et al., 2004) of mergers/non-mergers appears to need to be changed to accommodate this sample.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

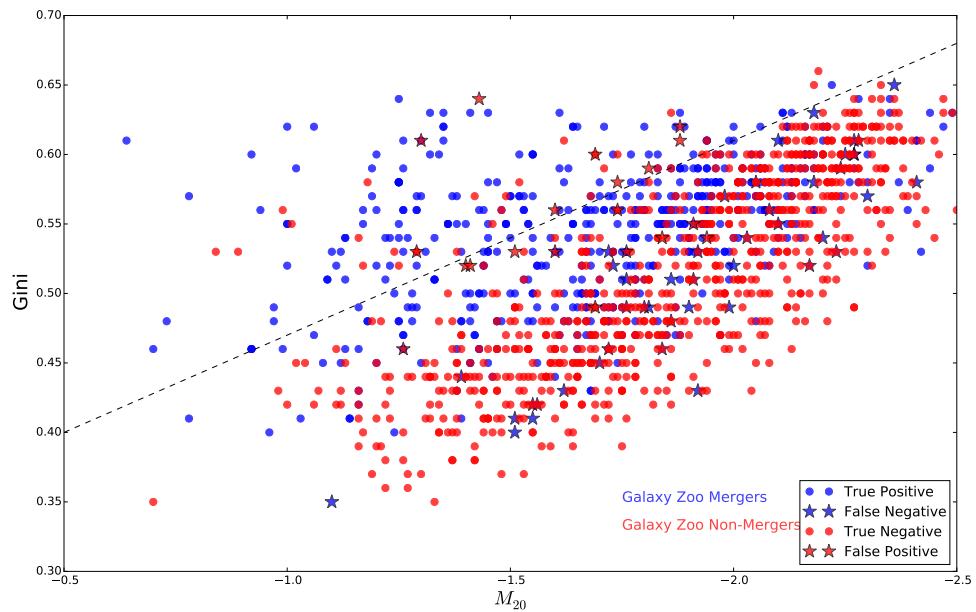


Figure 3.10 Gini- M_{20} color-coded by galaxy zoo merger classification and symbol coded by the confusion matrix. Unsurprisingly galaxies with G- M_{20} values on the outskirts of their classes distributions were classified incorrectly.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

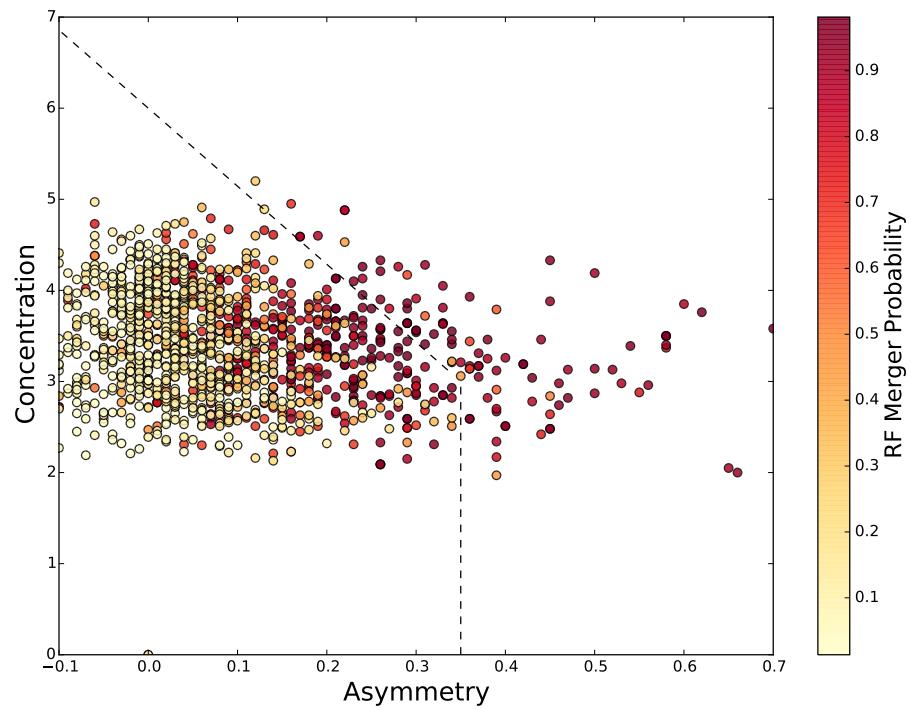


Figure 3.11 Concentration-Asymmetry color-coded by average RF Classification Probabilities following 1000 iterations of random forest. The dividing line of (Lotz et al., 2004) of mergers/non-mergers appears to need to be changed to accommodate this sample.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

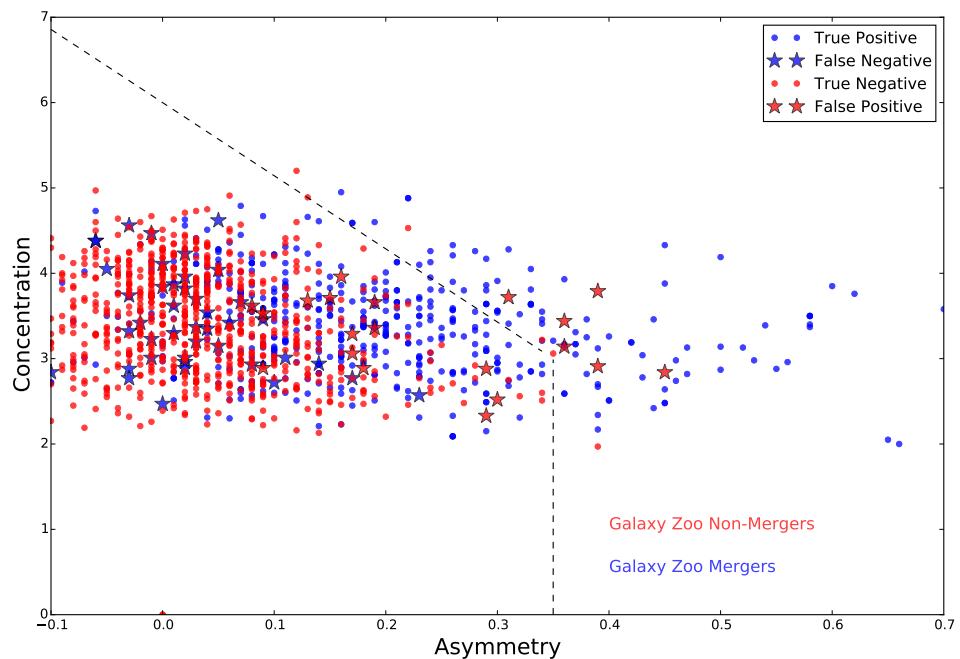


Figure 3.12 Concentration-Asymmetry color-coded by galaxy zoo merger classification and symbol coded by the confusion matrix.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

	Full Sample	Blue	Red	Merged	Pairs
PC1	0.33 ± 0.02	0.32 ± 0.02	0.40 ± 0.03	0.26 ± 0.03	0.37 ± 0.02
PC2	0.50 ± 0.03	0.52 ± 0.04	0.38 ± 0.03	0.36 ± 0.05	0.51 ± 0.03
PC3	0.56 ± 0.04	0.51 ± 0.04	0.69 ± 0.07	0.52 ± 0.05	0.53 ± 0.03
PC4	0.34 ± 0.02	0.35 ± 0.02	0.30 ± 0.02	0.31 ± 0.03	0.34 ± 0.02
PC5	0.53 ± 0.04	0.43 ± 0.03	0.69 ± 0.07	0.37 ± 0.04	0.52 ± 0.03
PC6	0.34 ± 0.02	0.34 ± 0.03	0.34 ± 0.03	0.34 ± 0.04	0.37 ± 0.02
PC7	0.82 ± 0.05	0.83 ± 0.07	0.59 ± 0.06	0.67 ± 0.07	0.82 ± 0.06
Gini	0.30 ± 0.02	0.36 ± 0.03	0.20 ± 0.02	0.25 ± 0.03	0.30 ± 0.02
M_{20}	0.30 ± 0.02	0.31 ± 0.02	0.45 ± 0.05	0.22 ± 0.02	0.31 ± 0.02
Multimode	0.27 ± 0.02	0.28 ± 0.02	0.29 ± 0.03	0.91 ± 0.10	0.26 ± 0.01
Intensity	0.25 ± 0.01	0.27 ± 0.02	0.24 ± 0.02	0.40 ± 0.05	0.28 ± 0.02
Deviation	0.56 ± 0.04	0.47 ± 0.04	0.72 ± 0.08	0.44 ± 0.04	0.54 ± 0.04
Asymmetry	1.00 ± 0.06	0.94 ± 0.08	1.02 ± 0.09	1.16 ± 0.12	0.92 ± 0.06
Concentration	0.26 ± 0.01	0.27 ± 0.02	0.28 ± 0.02	0.20 ± 0.02	0.28 ± 0.01
$g - r$ Color	0.37 ± 0.03	0.32 ± 0.03	0.39 ± 0.05	0.33 ± 0.05	0.39 ± 0.03
$\log M_*$	0.26 ± 0.01	0.28 ± 0.02	0.26 ± 0.02	0.26 ± 0.03	0.27 ± 0.02
sSFR	0.30 ± 0.02	0.33 ± 0.03	0.22 ± 0.02	0.24 ± 0.03	0.32 ± 0.02
$F(G, M_{20})$	0.28 ± 0.01	0.29 ± 0.02	0.33 ± 0.03	0.23 ± 0.02	0.29 ± 0.01
d(G, M_{20})	0.54 ± 0.04	0.70 ± 0.06	0.35 ± 0.03	0.64 ± 0.07	0.48 ± 0.04

Table 3.1 Feature Importances of Random Forest classifications. The importances are scaled by the largest importance (Asymmetry). Blue/Red division based on $g - r > 1.5$ and Merged/Pairs based on number of neighbors.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

	Full Sample	Blue	Red	Merged	Pairs
Asymmetry	1.00 ± 0.04	0.88 ± 0.05	1.03 ± 0.05	1.16 ± 0.08	0.93 ± 0.04
Concentration	0.34 ± 0.01	0.35 ± 0.02	0.35 ± 0.02	0.27 ± 0.03	0.37 ± 0.02
Gini	0.27 ± 0.02	0.29 ± 0.02	0.20 ± 0.02	0.23 ± 0.02	0.28 ± 0.02
M_{20}	0.36 ± 0.01	0.34 ± 0.02	0.51 ± 0.03	0.30 ± 0.03	0.36 ± 0.02
$f(G, M_{20})$	0.33 ± 0.01	0.35 ± 0.02	0.37 ± 0.02	0.29 ± 0.03	0.36 ± 0.01
$d(G, M_{20})$	0.55 ± 0.03	0.64 ± 0.04	0.39 ± 0.03	0.60 ± 0.04	0.54 ± 0.03

Table 3.2 Feature Importances of Random Forest classifications. The importances are scaled by the feature importance of Asymmetry. Even when fewer morphological statistics used the relative feature importances are very similar to the random forest with the full set of features. Blue/Red division based on $g-r > 1.5$ and Merged/Pairs based on number of neighbors.

not substantially decrease the OOB error any further. The number of estimators has no appreciable effect on the OOB score. Figure 3.14 shows that the number of estimators and the number of maximum leaf nodes (above 10) have no effect on the summary statistics. The PPV, completeness, and risk statistics are substantially worse below 10 maximum leaf nodes. These two figures show that our decision to use 500 estimators and a maximum of 100 leaf nodes represent good input parameters to use in the random forest classifications.

Figure 3.15 is a receiver operating characteristic (ROC) curve that demonstrates the performance of a classification when the threshold for classification is changed. In our case, we define galaxies with $> 40\%$ chance of being classified by random forest as a merger, which corresponds to the equilibrium on the ROC curve between maximizing the true positive rate while minimizing the false positive rate. The ROC curve for our classification scheme is particularly good. A perfect ROC curve would have a 100% true positive rate for all false positive values. We can see that a random forest determined merger probability of 0.4 is

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

the best maximizing the true positive rate while minimizing the false positive rate.

3.5.1.2 Comparisons of RF on Different Subsamples

Table 3.3 shows the summary statistics (and 1σ errors) for random forest classifications using a number of subsamples. The entire sample is divided in blue and red colors based on $g - r > 1.5$ rest-frame colors. Meanwhile, for the merged/pairs division the merged galaxies are combined with the full non-merger sample and likewise for the sample of pair galaxies. The random forest classifications are repeated 1000 times to determine the mean and 1σ error for the summary statistics. The sample used has a very significant effect on the summary statistics. For instance, the sample containing only merged galaxies and non-mergers has classifications which are the most specific and has the lowest total error of any sample. However, the strength of these summary statistics is balanced with the low completeness score. When comparing the completeness and specificity of every subsample with the full sample the better a subsample does in specificity it lacks in completeness and vice versa. It appears that there is no single random forest framework which can improve upon using the full sample in terms of both of these statistics.

Figures 3.16 and 3.18 show that certain morphological statistics are more or less important for classifying mergers depending on the subsamples investigated. For every subsample asymmetry was the most important statistic. A comparison between red galaxies and blue galaxies shows PC3, PC5 and deviation are more important for red galaxies and PC7, D(G, M_{20}), and PC2 are more important for blue galaxies. The statistics important to red galaxies are more dependent on the irregularity of the galaxy, which for a red galaxy would

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

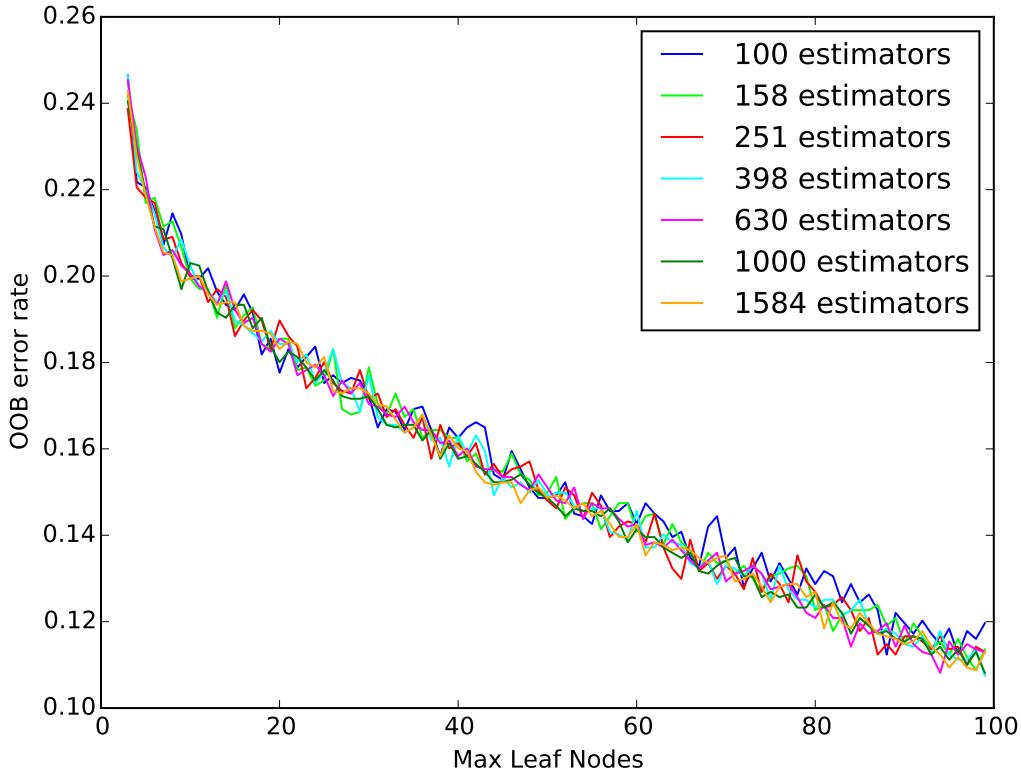


Figure 3.13 OOB (out-of-bag) errors for Random Forest classifications using different numbers of maximum feaf nodes and different numbers of estimators (AKA trees in the random forest). The OOB error estimates the fraction of misclassified data by using the a portion of the training set as the test set. We observe the OOB error continually decreases with an increasing number of maximum leaf nodes. However, the number of estimators (trees) has no effect on the OOB error. For consistency with Freeman et al. (2013) we use 500 estimators.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

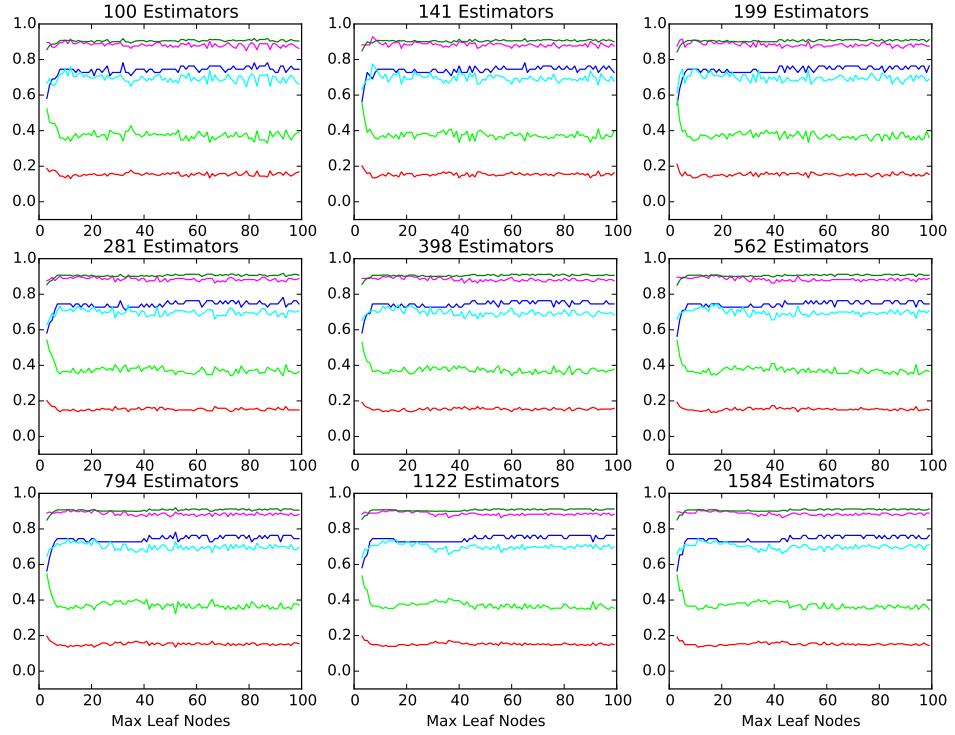


Figure 3.14 Summary statistics (specificity = *magenta*, completeness = *blue*, risk = *neon green*, total error = *red*, NPV = *forest green*, PPV = *cyan*) for Random Forest Classifications Using Different maximum feaf nodes and different numbers of estimators. None of the summary statistics are improved by an increase in either the number of estimators (trees in random forest) or the maximum number of leaf nodes (past 10 leaf nodes) which shows that we are free to choose any value we want for number of estimators of leaf nodes.

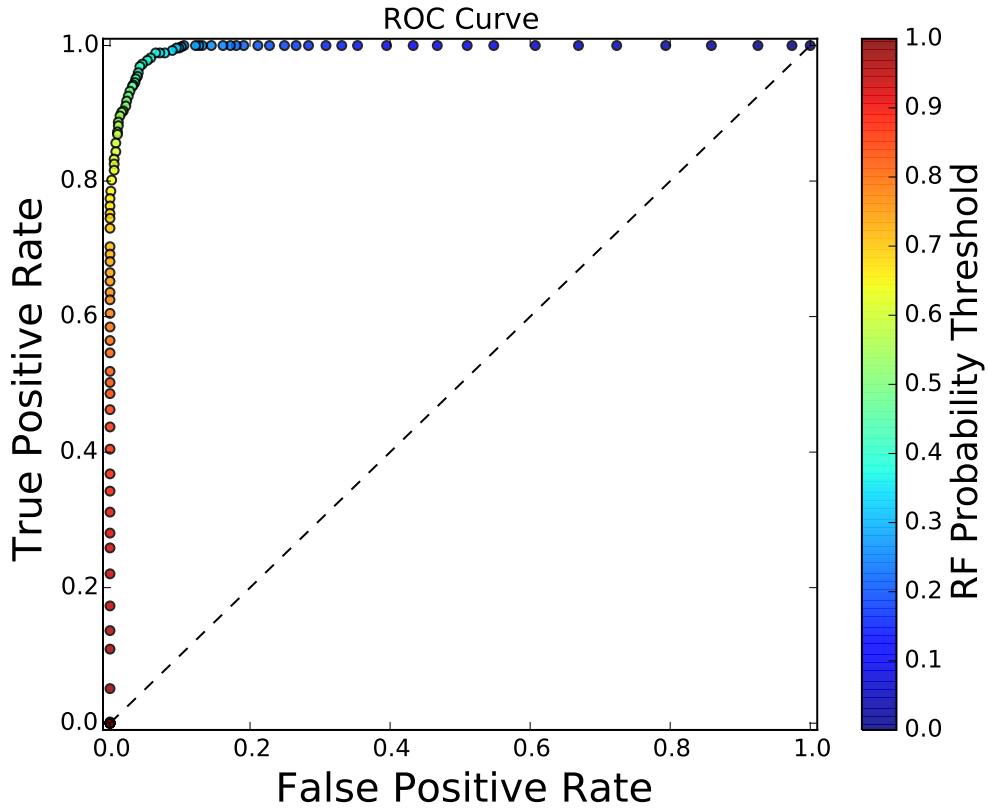


Figure 3.15 Receiver Operating Characteristic (ROC) curve shows the performance of the random forest classifications as the threshold for classification is changed. A perfect classification would have an ROC curve that is a right angle. This particular ROC curve shows that our classification is very good. The true/false positive rates are the number of true/false positives at a specific threshold divided by the total number of true positives and true negatives. The threshold used is the fraction of times a galaxy is classified as a merger following random forest.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

be a rarity. Likewise, the statistics important for blue galaxies are dependent on bulge strength which would differentiate blue galaxies more than red galaxies. Meanwhile, multimode is designed to find double nuclei and thus would be more suited to finding merged galaxies than pairs or the full sample.

If a user decides that they wish to find a very pure sample of mergers, but do not care about the completeness of their classifications they would be wise to use a random forest trained on the merged sample as opposed to the full merged and pairs sample. The problem of what factors are important or unimportant in a classification scheme is always up to the user to define.

3.6 Discussion

3.6.1 Random Forest Classifications of MaNGA Galaxies

MaNGA Galaxies

The Mapping Nearby Galaxies at APO (MaNGA) survey represents observations from SDSS-IV (Bundy et al., 2015). MaNGA used tightly packed optical fibers, which allowed spectral measurements of \sim 10,000 galaxies. The MaNGA Sample was drawn from a flat stellar mass distribution with $\log M_* > 9 M_{odot}$ and a wide range of environments.

We investigate a subset of MaNGA galaxies (\sim 900), measuring non-parametric morphological statistics and classify the results using a random forest. These galaxies also possess

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

	Full Sample	Blue Galaxies	Red Galaxies	Merged Sample	Pairs Sample
Completeness	93.29 ± 1.11	92.14 ± 1.68	94.70 ± 1.48	90.55 ± 3.25	92.20 ± 1.26
Specificity	93.17 ± 0.91	94.54 ± 1.05	92.13 ± 1.55	99.33 ± 0.30	94.04 ± 0.86
Risk	13.53 ± 1.04	13.32 ± 1.53	13.17 ± 1.61	10.13 ± 3.19	13.76 ± 1.12
Total Error	6.78 ± 0.51	6.28 ± 0.67	6.68 ± 0.82	1.46 ± 0.34	6.62 ± 0.51
PPV	89.92 ± 1.16	89.85 ± 1.69	91.25 ± 1.53	93.07 ± 2.81	89.56 ± 1.29
NPV	95.53 ± 0.69	95.85 ± 0.83	95.30 ± 1.22	99.07 ± 0.32	95.62 ± 0.66

Table 3.3 Summary statistics of Random Forest Classifications based on Using Different Subsamples. Blue/Red division based on $g - r > 1.5$ and Merged/Pairs based on number of neighbors.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

	Full Sample	Blue Galaxies	Red Galaxies	Merged Sample	Pairs Sample
Completeness	89.09 ± 1.26	90.63 ± 1.90	93.82 ± 1.46	88.15 ± 3.51	87.53 ± 1.38
Specificity	90.45 ± 1.01	93.34 ± 1.12	91.46 ± 1.69	98.96 ± 0.39	91.45 ± 0.96
Risk	20.46 ± 1.03	16.03 ± 1.67	14.73 ± 1.73	12.89 ± 3.38	21.02 ± 1.11
Total Error	10.08 ± 0.50	7.59 ± 0.71	7.45 ± 0.88	2.01 ± 0.37	9.95 ± 0.50
PPV	85.90 ± 1.19	87.71 ± 1.73	90.49 ± 1.65	89.45 ± 3.39	85.01 ± 1.32
NPV	92.73 ± 0.73	95.03 ± 0.93	94.51 ± 1.19	98.84 ± 0.34	93.00 ± 0.68

Table 3.4 Summary statistics of Random Forest Classifications based on Using Different Subsamples. Blue/Red division based on $g - r > 1.5$ and Merged/Pairs based on number of neighbors. Random forest classifications using only C, A, G, M20, d(G,M20) and F(G,M20).

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

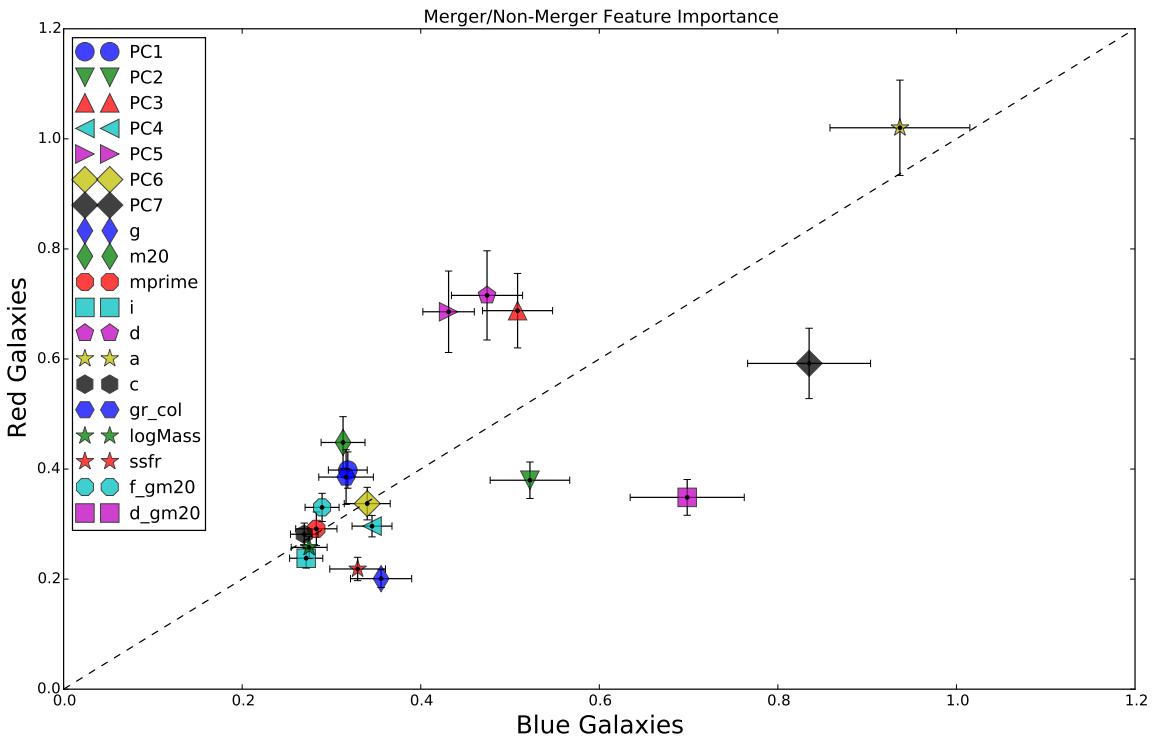


Figure 3.16 Feature importance comparisons between blue and red Galaxies, which are separated by $g - r > 1.5$. Asymmetry, PC3, PC5, PC7 and deviation are the most important features. Features in the upper half are more important for red galaxies (such as PC3, PC5 and deviation) are more dependent on the galaxy irregularity. Meanwhile features in the lower half are more important for blue galaxies (such as PC7, D(G, M_{20}), and PC2) are more dependent on the bulge strength.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

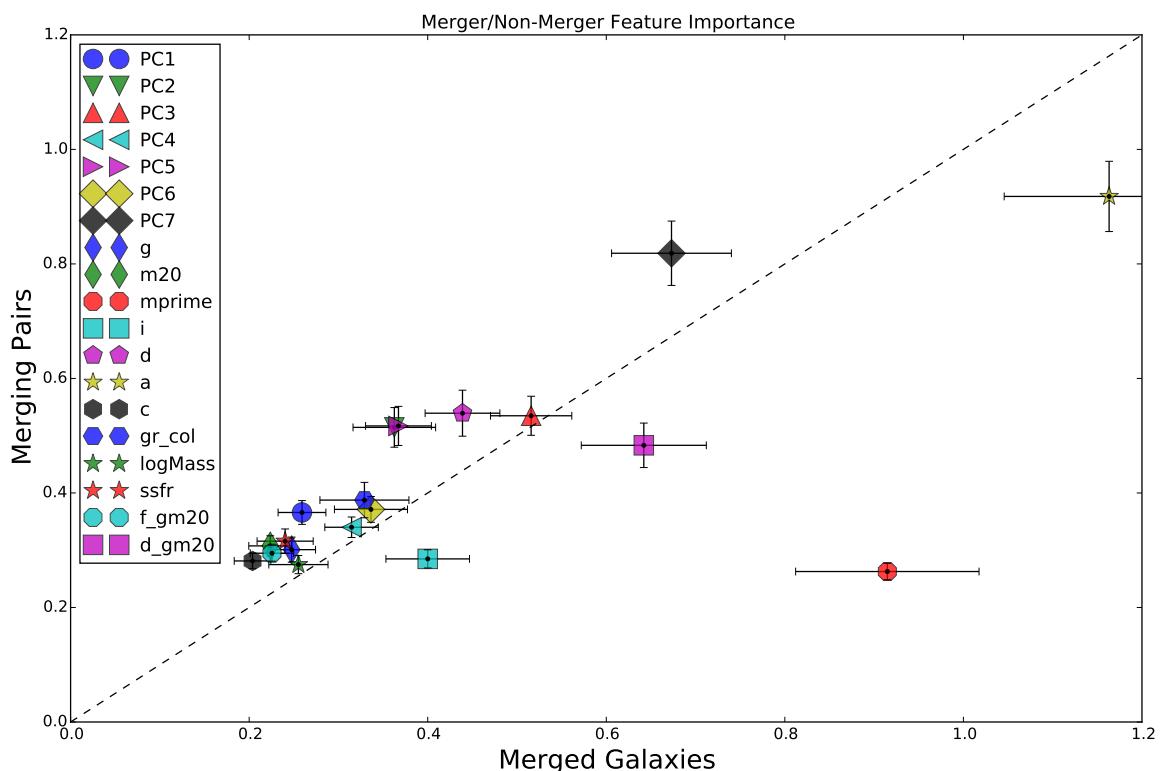


Figure 3.17 Feature importance comparisons between merged galaxies and merging pairs. Multi-mode is much more important to identifying merged galaxies rather than pairs.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

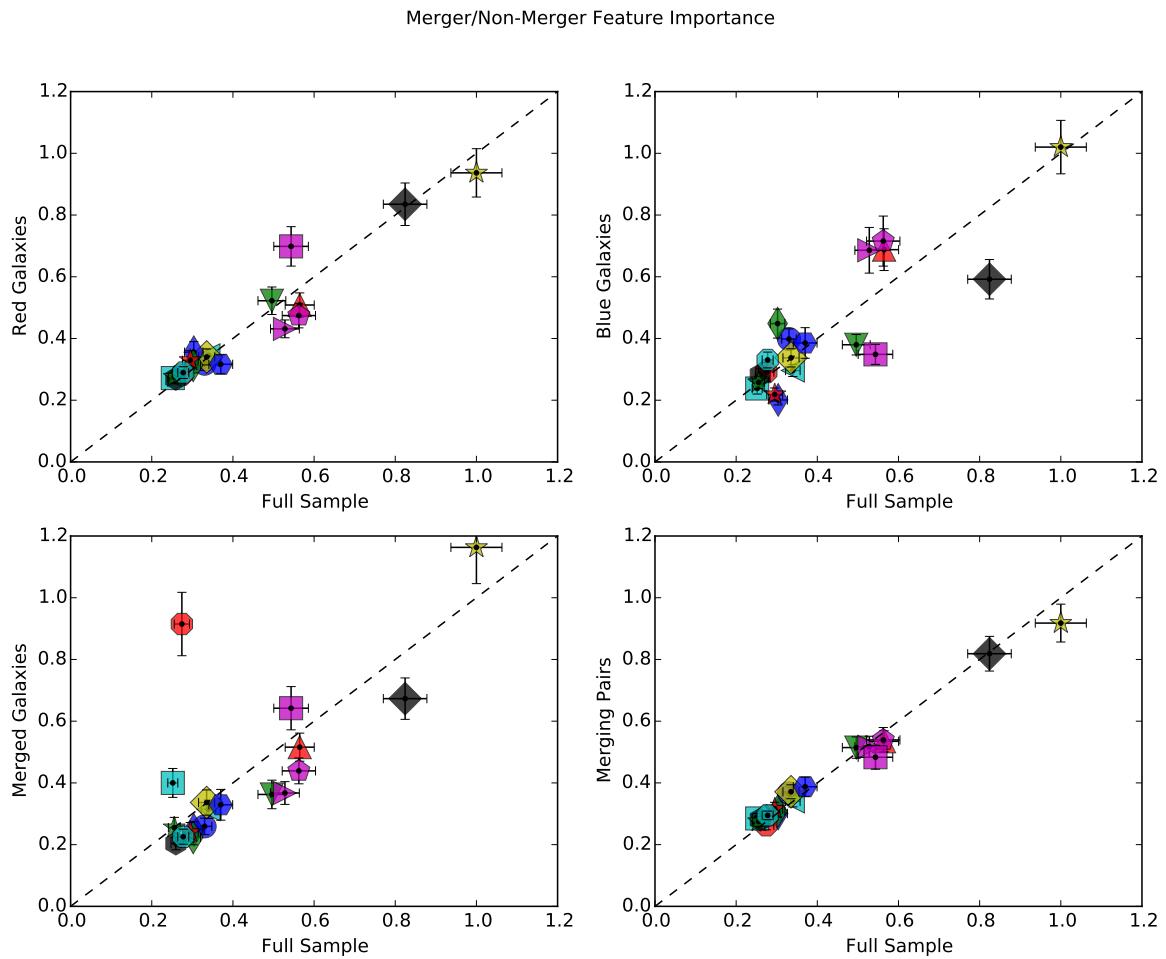


Figure 3.18 Feature importance comparisons between the full sample and blue, red, merged and merging pair galaxies.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

visual classifications from Galaxy Zoo for comparison. We select only galaxies with $z < 0.05$ as higher redshift galaxies will not be sufficiently resolved in the MaNGA/Pan-STARRS sample, leaving us with (~ 650) galaxies.

Results

Random forest merger classification results show that the local ($z < 0.05$) MaNGA merger sample is 100% complete and 78% specific. However, the amount of false positives is much greater than the number of true positives (142 vs 5). This test can select what might be a merger but additional criteria will be needed to confirm.

Figures 3.19 and 3.20 show the white light images and segmentation maps (from SExtractor) for MaNGA galaxies falsely classified as mergers by random forest. These figures only represent the 37 false positives with random forest merger probabilities > 0.6 . These galaxies represent the morphologies that are most misclassified by random forest and their misclassifications can inform future studies as to the potential pitfalls of merger classifications. Quite a few of these galaxies appear disturbed and could potentially be merger remnants. A few of the galaxies are large disk galaxies with bright star formation knots or have a foreground star in the segmentation map. Only a few galaxies have segmentation maps that do not seem to match the white light image.

The SDSS Sky Server 12 Database is used to view these merger false positive galaxies and to determine if SDSS finds other photometric and/or spectroscopic objects within the SExtractor galaxy segmentation map. Table 3.5 shows examples and counts for each of the types of false positives: bright foreground star, crowded field, nearby neighbor, asymmetric

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

or disturbed, tidal arms and unknown. Only three galaxies have a spectroscopically defined star within the galaxy image. In total, there are 14 galaxies with a bright star that is either inside or outside the segmentation map of a galaxy, either would significantly disrupt morphological measurements. One galaxy is in a very crowded field and a poorly defined segmentation map is disrupting morphological measurements. There are a number of potential merger or interaction remnant. These include, 13 asymmetric or disturbed spiral morphology, 3 galaxies with tidal arms and 3 with a nearby satellite galaxy. There are an additional 3 galaxies without any apparent morphological disturbance or bright foreground star. The SDSS images may not be deep enough to view very subtle and faint merger traits.

Even though the random forest classifications found many more false positives than true positives (as defined by Galaxy Zoo), these results are encouraging. The majority of merger false positives have foreground stars contaminating the segmentation maps. The morphological statistics are finding the bright region of off-center light and calling the galaxy a merger as it should. However, in these few cases the off-center light is in actuality a star. Edge-on galaxies are identified due to their dust lanes, while face-on galaxies can have clumpy star-formation that resembles merger remnants.

3.7 Summary and Conclusions

With 93% completeness and 93% specificity random forest is able to distinguish mergers from non-merger galaxies in PanSTARRS imaging using a variety of input features (PCs, non-parametric morphologies, sSFR, M_* , rest-frame color). The galaxies were initially visually classified by users of Galaxy Zoo and further analysis by Darg et al. (2010a) created

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

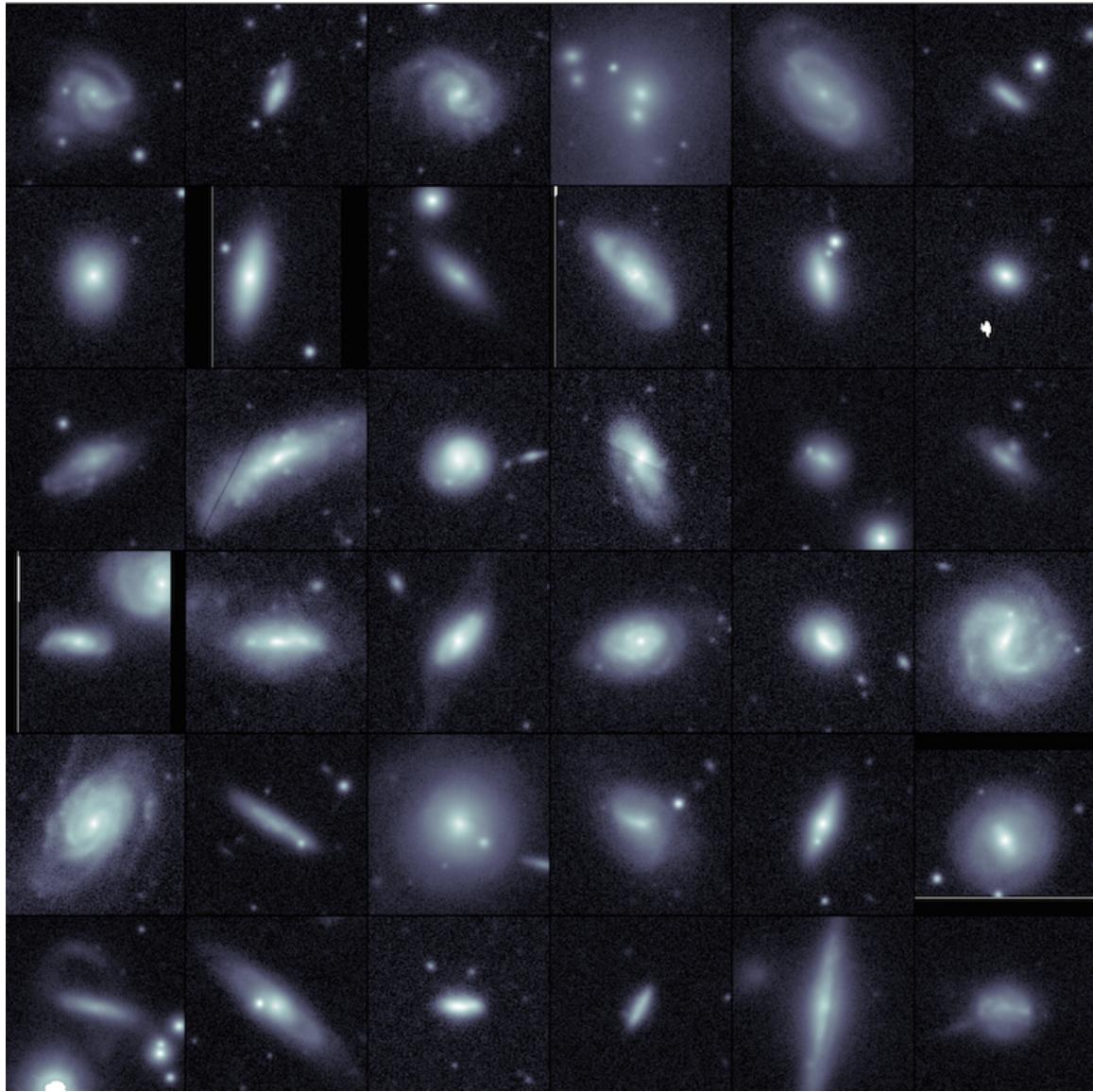


Figure 3.19 White light images of local ($z < 0.05$) MaNGA galaxies falsely identified as mergers by random forest. Some galaxies appear to be merger remnants overlooked by Galaxy Zoo users, while other galaxies have foreground stars contaminating the morphological statistics.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

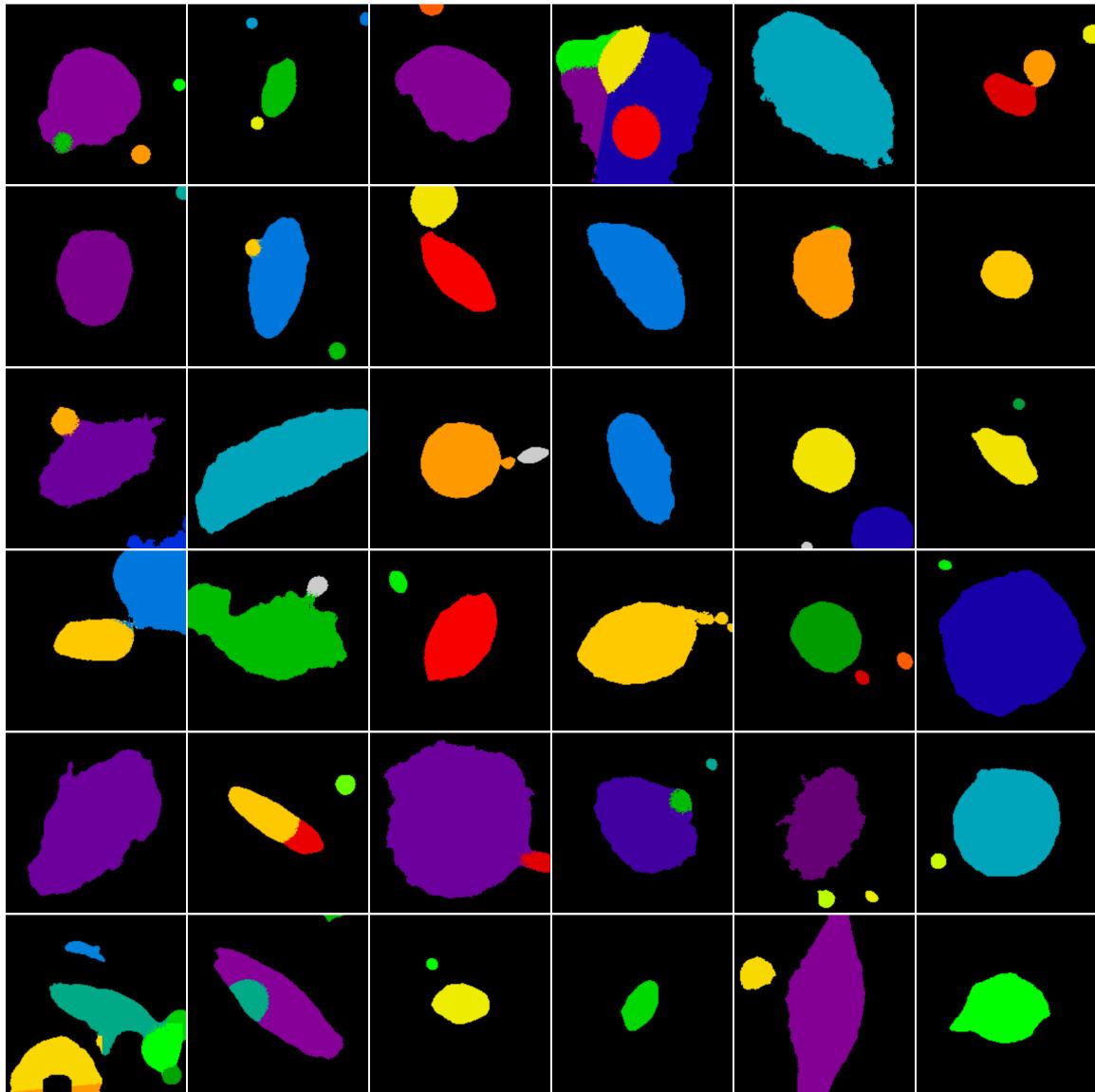


Figure 3.20 Segmentation maps (from SExtractor) of local ($z < 0.05$) MaNGA galaxies falsely identified as mergers by random forest. There are only a few instances of segmentation maps that do not appear to follow the white light image and in these cases a foreground star overlaid on a galaxy contaminates the segmentation process.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES
USING RANDOM FOREST

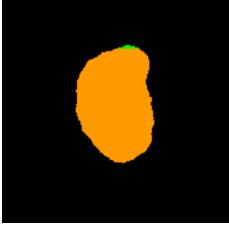
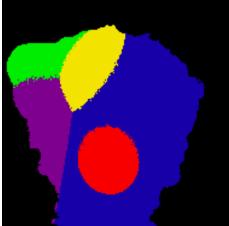
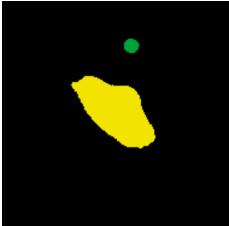
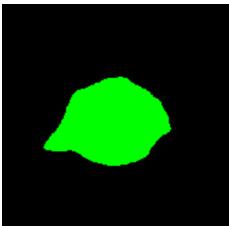
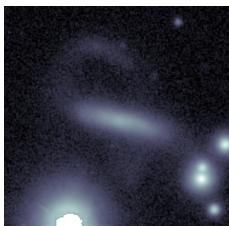
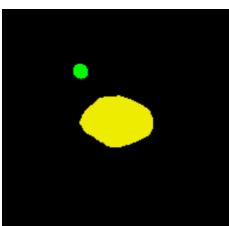
Type of False Positive	# of examples	Ex. Image	Ex. Segmap
Bright foreground star	14		
Crowded Image	1		
Nearby Neighbor	3		
Asymmetric/ Disturbed	13		
Tidal features	3		
Unknown issue	3		

Table 3.5 Examples and counts for different types of merger false positives.

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

a sample of expert vetted merging galaxies. These merging galaxies include galaxies that have already merged, and those in interacting pairs with visible tidal disruptions that will one day merge. We have determined that asymmetry is by far the most important indicator of whether a galaxy is experiencing a merger. The next most important features include: PC7, PC5, PC3, deviation and $d(G, M_{20})$. The importance of PC7 represents a very interesting result because PC7 is the least important PC but plays a huge role in determining whether a galaxy is a merger. If PC7 were a merger indicator it would be reasonable that it is not very important for a large diverse sample of high redshift galaxies, since mergers are not very common. Our sample of PanSTARRS represents a small subset of the non-merger galaxies at low redshift.

A random forest classification scheme using only concentration, asymmetry, Gini, M_{20} , $F(G, M_{20})$ and $d(G, M_{20})$ is nearly as successful at classifying mergers (with 89% completeness and 90% specificity) as a random forest using a much larger set of data. The importances of these features is very similar to the importances of these sample features in the full random forest classification set. Asymmetry is still by far the most important statistic, with $d(G, M_{20})$ the next most important.

A random forest using only the bare essential non-parametric morphological statistics could provide a very decent classification of mergers. However, if the highest levels of accuracy and completeness are most important to a study then more features can help to improve the classification results.

For a local ($z < 0.05$) sample of MaNGA galaxies random forest classifications trained upon the Galaxy Zoo and Pan-STARRS sample is 100% complete and 78% specific, albeit

CHAPTER 3. MERGER CLASSIFICATIONS OF PAN-STARRS GALAXIES USING RANDOM FOREST

with a greater number of false positives than true positives (142 vs 5). Even though random forest finds a large number of false positives, further analysis shows that up to 50% of these galaxies are potentially mergers missed by Galaxy Zoo classifiers. Other false positives have foreground stars contaminating galaxy flux but otherwise resemble off center clumps of merger remnants. The number of false positives is still much smaller than the number of true negatives, and represent a very manageable sample to be visually classified by a single user.

The reliability of random forest classifications of mergers will only increase as Pan-STARRS obtains more images which will lead to more robust merger studies. The usage of random forest and other supervised learning based classifications will only become more important as LSST and other large surveys come online that will observe far too many galaxies to be quickly classified by humans. Better image segmentation and star separation methods will also improve random forest classification purity.

Simulations could provide another possible training set. In simulations the exact moments of mergers would be known through merger tree analysis. However, it is important to understand that a morphology study of simulated galaxies to work there need a prescription for mapping the results of an N-body or hydrodynamical to a realistic looking galaxy. We would also ideally have access to multiple viewing angles of the mergers, which would eliminate some of the ambiguity of how to define a merger visually. The next chapter deals with the VELA simulations which could eventually function as the random forest training set, however more galaxies are required to match the same levels of completeness and specificity as seen with using the Galaxy Zoo and Pan-STARRS training set.

Chapter 4

VELA Simulation Galaxy

Morphologies

The VELA hydrodynamical simulation of 10 CANDELized¹ galaxies provide an avenue to study how morphology evolves over time. A time series cross-correlation between morphology (PC1, PC2 and PC3) and physical parameters (sSFR, f_{gas} , \dot{f}_{gas} , ex-situ- M_*/M_* , ex-situ- \dot{M}_*/M_* , M_{dm}/M_* , \dot{M}_{dm}/M_*) for both the inner kpc and entire galaxy determines the strength of correlation and Δt between a morphology and physical parameter time series. PC1 (inverse bulge strength) is most correlated with sSFR, f_{gas} , and ex-situ- M_*/M_* . This correlation implies that as the bulge grows stronger gas fraction and ex-situ stellar mass decrease, and star-formation quenches. PC2 (concentration) is not very well correlated with any physical parameter. The

¹Images have been processed to resemble real galaxies observed by CANDELS

CHAPTER 4. VELA SIMULATIONS

anti-correlation implies as galaxies become more compact, star-formation quenches while gas and ex-situ mass decrease. PC3 (asymmetry) is correlated with sSFR and f_{gas} , reaffirming that star-forming galaxies have more gas and are more asymmetric than quenched galaxies.

Additionally, the VELA simulations are classified into the PC groups defined in chapter 2. Each galaxy is followed through time to discover that the vast majority of simulated galaxies become bulge-dominated by $z=1$. Only one galaxy, VELA27, becomes disk-dominated by the end of the simulation. Major mergers are found to cause bulge-dominated systems to become more irregular, and grow disk and also cause disk-dominated systems to become temporarily irregular before settling into a bulge-dominated galaxy. Minor mergers have a comparably minor and transient effect on a galaxy's morphology. The VELA simulations represent a very small sample of galaxies and future iterations of the morphological dataset will allow for greater diversity of morphology. With increased diversity of galaxy formation conditions we will be able to better compare simulated galaxies to the real observed galaxies of surveys such as CANDELS.

4.1 Introduction

Well known relationships exists between morphological evolution and galaxy stellar mass or SFR (Kauffmann et al., 2003; Wuyts et al., 2011), the bulge-SMBH

CHAPTER 4. VELA SIMULATIONS

co-evolution (e.g. Elbaz & Cesarsky, 2003) and of course the fundamental plane for elliptical galaxies (Terlevich et al., 1981). However, such studies rely on our ability to imply relationships between possible progenitor and descendant galaxies, since we can not directly examine galaxies evolving in observations. In this context, the exact mechanisms causing galactic evolution are not known.

These studies can not follow the evolution of a single galaxy through time. Observational studies can identify galaxies at different epochs and make assumptions about which high redshift galaxies represent the progenitors of lower redshift galaxies. Analytic tools such as number counts and luminosity functions (e.g. Faber et al., 2007) can be used to study the evolution of galaxies and are binned by redshift and parameters such as mass, luminosity, number density and morphology. The shape and magnitude of luminosity functions vary depending on galaxy properties such as morphology, mass, colors and spectral types (Sandage et al., 1985; Faber et al., 2007 and references therein).

We can identify the effects of galaxy evolution by observing how the number (or luminosity) density for various parameters changes over time. For instance, by binning galaxies by mass and star-formation, luminosity function analysis led to conclusions such as: more massive galaxies experience star-formation earlier than lower mass galaxies (Cowie et al., 1996) and that late type galaxies were more common in the past than today, while early types have remained constant in frequency (Lin et al., 1999).

CHAPTER 4. VELA SIMULATIONS

The binning is a necessary step to create a sufficient sample size but can have consequences. Luminosity functions suffer from an inability to disentangle whether changes are inherent to the characteristics of such a class or whether galaxies are just scattering amongst bins. For instance, if the number density of a particular bin remains constant is this an example of no evolution occurring? Or is there an equal number of galaxies scattering in and out of the particular bin over time? Without knowing how many galaxies “should” be in the bin, it is hard to distinguish between these two scenarios. Additionally, size and surface brightness limits of surveys leads to uncertainty in how complete and representative a sample may be.

On the other hand, hydrodynamical simulations allow us to “watch” galaxies “grow up” into their final evolved forms (Lotz et al., 2008; Snyder et al., 2013). With hydrodynamical simulations we have direct access to the star-formation rates, stellar mass, gas mass, dark matter mass and ex-situ stellar mass values all as a function of time which can all be used to directly prove the relationships observations only were able to very suggestively imply. We can use the hydrosimulations to directly witness “downsizing”, do the simulations show that less massive galaxies form stars after more massive galaxies?

Hydrodynamical simulations have already found that galaxy interactions can trigger either bulge or disk formation (Snyder et al., 2015b), massive galaxies quench earlier and faster (Zolotov et al., 2015), and supernova feedback are effectively maintaining prolate galaxy shapes (Tomassetti et al., 2016). Each of these results are

CHAPTER 4. VELA SIMULATIONS

possible because we can track galaxies and their properties through time.

Previously, simulations were only able to study isolated individual galaxies (Noguchi, 1999; Immeli et al., 2004a,b; Bournaud et al., 2007; Elmegreen et al., 2008; Bournaud & Elmegreen, 2009; Bournaud et al., 2009). The next step placed galaxies into hydrodynamical cosmological simulations but these initially suffered from low mass and spatial resolutions along with poorly constrained gas-physics models (Springel & Hernquist, 2003; Kereš et al., 2005; Governato et al., 2007; Dekel et al., 2009b). Improvements to sub-grid models include the addition of SNe feedback, SF regulation (Agertz et al., 2011) and SN outflows (Governato et al., 2012). Newer examples of modeling galaxy formation in a cosmological context have enhanced resolution and better gas-physics models (e.g. Ceverino et al., 2014).

High resolution zoom-in cosmological simulations can bridge this gap of knowledge. Simulations, such as VELA (Ceverino et al., 2010a; Ceverino & Klypin, 2009; Ceverino et al., 2012; Dekel et al., 2013; Ceverino et al., 2014), can track the mass, size, star-formation rate, gas accretion rate, morphology, etc. of a galaxy throughout time. VELA is an adaptive mesh-refinement simulation with a maximum resolution of 25 pc. (Zolotov et al., 2015). Simulations can provide unprecedented insight into the types of physical mechanisms affecting the star-formation rate and morphology. For instance, halos provide torques capable of inducing galaxy elongation, eventually leading to puffed up inner stellar orbits aligned with the gas disk (Tomassetti et al., 2016). A majority of simulated high redshift galaxies are elongated and not disks

CHAPTER 4. VELA SIMULATIONS

or spheroids (Ceverino et al., 2015). Additionally, simulations point toward evidence suggesting inside-out quenching arises through the manifestation of a star-forming ring surrounding the inner central region (Tacchella et al., 2016) and gas metallicities of clumps is evidence for fast gas accretion from the cosmological inflow of metal-poor gas (Ceverino et al., 2016). The quality of simulations is determined by the resolution and reliability of sub-grid models, which is continually evolving and improving (Torrey et al., 2014).

Zolotov et al. (2015) describes galaxy evolution with gas mass and the inflow/outflow rate of gas. In their evolutionary framework, the gas mass growth rate begins to increase steeply triggering the beginning of a compaction phase. This compaction is the increase of stellar mass in the inner kpc. The galaxy need not necessarily shrink in terms of total size (both physical and in mass). The stellar mass growth is slower than the gas mass growth. The total SFR follows the gas mass for both the inner kpc and the total galaxy. The gas inflow rate is greater than the SFR implying a wet compaction² is occurring due to the preponderance of gas. Wet compaction is also evidenced by the large fraction of stars formed in the inner bulge. For stars to have formed in the bulge there must be a large reservoir of gas to draw from. Starbursts in the bulge would be evidence of a “blue-nugget” phase of galaxy evolution (Dekel & Burkert, 2014) similar to observational evidence of compact, SF galaxies (Barro et al., 2013, 2014a; Bruce et al., 2014a,b; Williams et al., 2014).

²“Wet” referring to an overabundance of gas, as opposed to a “dry merger” where there is little to no gas present.

CHAPTER 4. VELA SIMULATIONS

We investigate the similarities and differences between galaxies observed in the real universe by CANDELS and the simulated galaxies of VELA. By comparing the simulated galaxies to the observed galaxies we can relate what we know about the physical mechanisms shaping the simulated galaxies (such as gas accretion, merger activity, etc.) to similar real galaxies. In this way, we can provide an explanation of the hows and whys of the evolution of galaxy morphology in terms of star formation and physics.

However, simulated galaxies and observed galaxies are not always visually similar. The raw images of high-resolution simulated galaxies are too highly resolved to resemble Hubble images of a real galaxy. For this reason, we need a method to blur high-resolution simulated images to match the seeing of HST (a process known as CANDELization Mozena, 2013; Moody et al., 2014; Snyder et al., 2015b). With this procedure we transform simulated galaxies into analogs of real observed galaxies while including observed wavelength and line-of-sight dependences. We can analyze images of simulated galaxies as if they were observed by HST. Previous studies of VELA galaxy morphology (Snyder et al., 2015b) have found that the VELA simulated galaxies are becoming more bulge-dominated over time, but that galaxies that become bulge dominated may not always remain that way. Galaxy interactions may either trigger bulge or disk growth. Other examples of morphological studies using simulations (Scannapieco et al., 2010) compare Sérsic indices, disk-to-total ratio, colors of simulated galaxies to the real observed galaxies of Gadotti (2009) and find simulated bulges

CHAPTER 4. VELA SIMULATIONS

resemble pseudo-bulges rather than observed bulges.

What the VELA simulated galaxy sample lacks in overall number of galaxies is compensated by the number of time steps and camera angles each galaxy is observed from. There are only 10 individual VELA simulations but a sample of over 2500 galaxies because each galaxy is observed at 6 different viewing angles and between 20-40 time steps. We can use these numerous time steps to follow individual galaxies through time to understand how mergers or gas accretion influence galaxy morphology. The multiple viewing angles will allow us to determine what affect (if any) viewing angle has on correlations between morphology and physical mechanisms.

In the previous chapter we used random forest to classify galaxies into mergers and non-mergers but in this chapter we use a much different technique (time series cross-correlations) to understand how morphology is related to physical assembly. Both techniques quantify the relationship between non-parametric morphological statistics.

The random forest classifications define the statistical subspaces explaining mergers but initially only for galaxies that have been visually identified as either mergers or non-mergers. This analysis allows us to determine which statistics are the most important for making a merger/non-merger distinction and allows us to speculate on how physical mechanisms affect the visual morphology. The random forest training assumes the visual identification of mergers are completely objective. However, this might be the case. Human classifiers all bring their own biases into the classification.

Instead, the VELA simulations represent a sample where everything about the sys-

CHAPTER 4. VELA SIMULATIONS

tem (from stellar mass to the times mergers occur) is known. We can directly measure the effects that physical assembly mechanisms have upon the galaxy morphology. We can measure the time a galaxy needs to change its morphology following a physical mechanism or vice versa. The primary drawbacks, however, include the small number of simulated galaxies and the possibility the models of galaxy evolution/formation are not completely accurate.

4.2 Data

4.2.1 VELA Simulation

The VELA simulations are a suite of zoom-in hydro-cosmological simulations of moderately massive galaxies calculated using Eulerian gas dynamics and an N-body Adaptive Refinement tree (ART, Kravtsov et al., 1997; Kravtsov, 2003). The simulations adopt the standard WMAP5 Λ CDM cosmology with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.045$ and $h = 0.7$ (Komatsu et al., 2009). The VELA simulations are described in depth by Ceverino et al. (2010a); Ceverino & Klypin (2009); Ceverino et al. (2012); Dekel et al. (2013); Ceverino et al. (2014).

Numerous sub-grid models are incorporated to model physical processes on scales below the simulation resolution (Ceverino & Klypin, 2009). These models account for UV-background photoionization, gas and metal cooling, stochastic star formation, gas recycling, metal enrichment, supernovae feedback and the feedback due to the

CHAPTER 4. VELA SIMULATIONS

radiation pressure and radiative heating of young stars (Ceverino et al., 2010b, 2012, 2014). AGN/super massive black hole feedback is not accounted for. However, for the halo masses investigated ($10^{11} - 10^{12} M_{\odot}$) AGN feedback is not believed to be a dominant effect (Moody et al., 2014).

The assumed uniform UV background is based on the redshift dependent of Haardt & Madau (1996) model. However in regions of dense gas ($> 0.1 \text{ cm}^{-3}$) a suppressed UV background is used to recreate the self-shielding of dense gas and allows gas to cool to $T \sim 300 \text{ K}$, which is necessary to form stars (Zolotov et al., 2015). Gas and metal cooling rates are calculated for a given gas density, temperature, metallicity, and UV background based on the CLOUDY code (version 96b4; Ferland et al., 1998). The stochastic star formation model follows the observed Kennicutt-Schmidt law (Kennicutt, 1998; Mozena, 2013) and forms stars in regions of dense ($> 1 \text{ cm}^{-3}$), cool gas ($T \sim 300 - 1000 \text{ K}$). The scale of the mesh refinement is between 17 - 35 pc which leads to star particles of $M \sim 10^5 M_{\odot}$, sufficiently small to resolve minor stellar clusters (Moody et al., 2014; Snyder et al., 2015b). Winds from stars and supernovae are emitted at a constant rate of 40 Myrs following star-formation (Moody et al., 2014; Zolotov et al., 2015). The sub-grid radiation pressure model adds a non-thermal pressure from the ionizing photons of massive stars to the total gas pressure (Ceverino et al., 2014) and the feedback from radiation pressure helps to quench star formation which produces realistic star formation histories for lower mass galaxies (Trujillo-Gomez et al., 2015; Moody et al., 2014). Future simulations will likely incorporate

CHAPTER 4. VELA SIMULATIONS

AGN feedback but for now it is important to remember that a potentially important source of feedback is not incorporated.

Haloes with virial masses $10^{11} < M/M_\odot < 10^{12}$ and not undergoing a major merger at $z = 1$ were selected randomly from the N-body simulation. These haloes were re-sampled and re-simulated with full hydrodynamics at high resolution and full physics to $z \sim 1$. The ROCKSTAR halo finder (Behroozi et al., 2013) was used to calculate masses and physical sizes (Moody et al., 2014). The simulated galaxies have $9.3 < \log M_*/M_\odot < 10.7$ and are available in increments of roughly 120 Myrs (Moody et al., 2014).

4.2.2 Image Processing and CANDELization

The post-processing method converts a simulated galaxy into raw mock images using the dust radiative transfer (RT) code **SUNRISE** (Jonsson, 2006; Jonsson et al., 2010). Every star particle is assigned a spectral energy distribution (SED) based on mass, age and metallicity using **STARBURST99** stellar population models and a Kroupa (2001) initial mass function (IMF) (Snyder et al., 2015b). The emitted light is followed through surrounding regions of gas and dust which leads to scattering, absorption, and dust re-emission (Mozena, 2013). The dust density is assumed to follow the metal density calculated in the VELA simulations (Snyder et al., 2015b). A Milky Way dust-to-metals mass ratio model is assumed (Cardelli et al., 1989; Gordon et al., 2003; Snyder et al., 2015b; Mozena, 2013).

CHAPTER 4. VELA SIMULATIONS

Each output galaxy is observed from 6 different viewing angles. Of these six different viewing angles, one is edge-on and one is face-on. These views are determined from the angular momentum vector. The other 4 angles are randomly selected (Snyder et al., 2015b). However, the “randomness” of these angles is currently under debate.

Our analysis focuses on 9 galaxies with images that have been prepared to match realistic seeing (Moody et al., 2014; Snyder et al., 2015b). This process, known as “CANDELization”, simulates the noise and seeing of Hubble Space Telescope WFC3/IR images.

The `SUNRISE` output images are convolved with a PSF for each of these filters, binned to a pixel scale of 0.06 arcsec and have noise added to match the noise of typical CANDELS observations (Mozena, 2013; Grogin et al., 2011; Snyder et al., 2015b). Now, we essentially have a library of galaxy images across cosmic time and waveband that we can directly compare to observed galaxies in the CANDELS survey. Additionally, the Gini/ $M_{20}/A/M/I/D$ structural statistics have been measured in Snyder et al. (2015b). We use our PCA-based technique described in Chapter 2 to group and classify the simulated galaxies.

Physical parameters such as gas mass, ex-situ stellar mass and dark matter mass are derived from the dark matter halo simulations and are not the result of fitting to an SED model.

4.3 PCA-Morphology Groups

Snyder et al. (2015b) showed that the evolution of morphology is not uniform. Galaxies at $z \sim 2$ are typically compact, with potentially unresolved star-forming disks. The bulge and disk components of these galaxies grow in both size and mass between $z \sim 1.5$ and $z \sim 1$. The amount of large disk galaxies (as characterized by Gini and M_{20} values) forming stars is correlated with stellar mass. This could be the result of increasing star-formation efficiency since $z \sim 2$ (at least until $z \sim 1$). Additionally, the most massive galaxies possess more ordered motion (implying rotating disks) than less massive galaxies (Kassin et al., 2012).

Simulated galaxies bifurcate into low Gini coefficients and higher Gini coefficients at higher redshift. Both sets of galaxies appear to grow a bulge (increasing Gini and decreasing M_{20} values) by $z \sim 1$. The Petrosian radii follow a $(1 + z)^{1.5}$ relation. However, there is a wide scatter in this evolution where some galaxies remain constant in physical size while others grow greatly (Snyder et al., 2015b). The estimated half-light radius is measured at a rest-frame B -band but does not incorporate changes to the PSFs which might account for the 2–4x difference when compared to the 3D half-mass radius (Zolotov et al., 2015).

We study 7 structural measurements (Gini, M_{20} , C , A , M , I , D) for every galaxy at every time step and viewing angle. We then project the morphological data from the VELA simulations into the PC basis defined in Chapter 2 and Peth et al. (2016). We can use the convex hull method (of chapter 2 and Peth et al. 2016) to group

CHAPTER 4. VELA SIMULATIONS

the simulated galaxies into the same groups as the observed CANDELS galaxies. We observe the simulated galaxies over a significant amount of time and observe evolutionary stages of the PC groups.

We also investigate the relationship between morphology, as represented by the PC group classification, and major/minor mergers, with an additional emphasis on the gas fraction of the galaxy. We find that major mergers are (unsurprisingly) capable of completely transforming the morphology of a galaxy. Major mergers can not only turn disk-dominated galaxies into spheroids but can also help trigger disk growth in bulge-dominated systems. Minor mergers have more limited impact on the overall galaxy morphology. Any change caused by a minor merger is short lived and transient. When mergers are infrequent secular processes can explain the regrowth of disks in previously bulge-dominated galaxies (when no merger has occurred for at least a Gyr or more).

4.3.1 PC Group Demographics

Figure 4.1 shows the histogram of PC group classifications at all time steps and viewing angles for the 10 VELA galaxies with morphological measurements. The PC group classifications of CANDELS galaxies are included for comparison. The simulated galaxies are most commonly classified into group 6, similar to observed galaxies. Group 6 is defined by a compact, small size and the apparent lack of disk structure. In contrast with the CANDELS sample, many fewer galaxies are classified

CHAPTER 4. VELA SIMULATIONS

into the disk-dominated groups 1, 2 and 5 and the bulge-dominated asymmetric group 9. This indicates the limit of the ability of the simulation to create disk galaxies to the same extent as the real Universe. The small sample of VELA simulations may not be representative of the Universe which could explain the lack of disk galaxies. However, this work is intended to investigate the morphologies of the simulated galaxies and not to investigate the veracity of the simulations themselves. The dearth of group 9 galaxies indicates VELA galaxies do not experience many mergers and visible evidence of merger activity is short lived. VELA galaxies experienced frequent mergers prior to $z \sim 3$ and the beginning of morphological measurements. Additionally, if disturbed morphologies were shorter lived than the 100 Myr time steps it would be possible for a galaxy to experience a merger and settle without displaying any evidence (Snyder et al., 2015b).

Table 4.2 and Figure 4.1 show the distribution of groups by viewing angle. The demographics of the PC groups are remarkably similar across viewing angles except for a few dissimilarities present in Camera 1 (edge-on view). The Camera 1 viewing angle identifies an excess of group 4 galaxies and a slight deficiency of group 6 galaxies compared to the other viewing angles. Group 4 is likely the most susceptible to viewing angle changes and are thus more likely to be edge-on systems. Cameras 5-8 are random orientations and thus it is not surprising the PC group demographics are quite consistent.

Table 4.1 shows the group classification distribution for each VELA galaxy. VELA02

CHAPTER 4. VELA SIMULATIONS

and 03 spend a considerable amount of time classified as group 6. Meanwhile, VELA12, 15, and 27 are rarely classified into group 6 but rather are mostly classified in groups 4 and 8 (the intermediate morphological stage between bulge-dominated and disk-dominated). VELA 04, 05, 14, 26 and 28 display evidence of an outer envelope and are often classified as group 0. VELA14 spends an excess fraction of time as group 9 (compared to other simulations), which is a possible indicator of tidal tails or other merger remnants.

Figure 4.2 is the Gini– M_{20} diagram for all the VELA galaxies differentiated by PC group and binned by redshift with $z \geq 2$ (red) and $z < 2$ (blue). The white contours represent the location of CANDELS galaxies defining each PC group.

Figure 4.3 is the Concentration-Asymmetry diagram for VELA galaxies, differentiated by PC group and binned by redshift with $z \geq 2$ (red) and $z < 2$ (blue). The white contours represent the location of CANDELS galaxies defining each PC group. The black dashed line separates galaxies with $A > 0$ and $A < 0$. As explained in Chapter 2 $A < 0$ values can be thought of as $A = 0$. However, the sheer number of galaxies with $A < 0$ leads us to believe there is an offset in asymmetry values of Snyder et al. (2015b). Overall the galaxies are not very asymmetrical. The lone exception is group 9, but this is one of the defining characteristics of the group and would explain the overall lack of group 9 galaxies.

Correcting the distribution of asymmetry values to match the distribution of the observed CANDELS galaxies has only a minimal effect on PC group classifications.

CHAPTER 4. VELA SIMULATIONS

We use the average asymmetry of CANDELS group 6 galaxies to define the likely zero-point for asymmetry, as these galaxies are not visibly asymmetric. The PC values are recalculated and group classifications of the VELA galaxies are repeated. Figure 4.4 shows the differences in group classification following the asymmetric correction. Group 4 galaxies become reclassified as groups 8 or 9. Since the change in asymmetry only affects the PC group classification for a small amount of galaxies we will not correct asymmetry values.

PC Group Classifications of Minor and Major Mergers

Determining the existence of a merger is an important aspect of research, as it allows us to understand the relationship between physical interactions and morphology. Chapter 3 dealt with identifying mergers through the use of visually classified morphology alone. In the VELA simulations we have information on the amount of stellar mass formed outside of a galaxy (ex-situ stellar mass) which is an indicator of the strength of a merger.

An increase of ex-situ stellar mass between 10-30% represents a minor merger and an increase of $> 30\%$ represents a major merger (Zolotov et al., 2015). These values correspond to the standard definition of 1:10 and 1:3 mass ratios between galaxies experiencing minor and major mergers. Figure 4.5 shows the PC groups that minor and major mergers are classified into. Many minor mergers are classified as groups 0, 6 and 8. In the CANDELS sample groups 1 and 9 possess the most visually classified

CHAPTER 4. VELA SIMULATIONS

irregular galaxies and are most likely merger remnants. Major mergers are most commonly classified into groups 4 and 6. Major mergers are more likely than minor mergers to be classified into the disk-dominated groups 1, 2 and 5.

Since most minor mergers are classified into bulge-dominated groups, minor mergers will lead to bulge growth and the “compactification” detailed in Dekel et al. (2009a); Zolotov et al. (2015). The demographics of minor mergers closely resembles the total VELA sample implying minor mergers do not affect the morphology of a galaxy very much. Meanwhile, major mergers are related to disturbed morphologies because of the overabundance of major merger remnants classified into irregular/disk-dominated PC groups (groups 1, 2, 4 and 5).

As we will investigate in the next section, the morphological change of a galaxy may be delayed from infall of ex-situ stellar mass, gas or dark matter.

4.3.2 PC Group Flow

How exactly does morphology evolve with time? The VELA simulation suite provides a test bed to directly witness morphological changes, and corresponding merger events.

Figure 4.6 shows the PC group classification for all VELA galaxies over cosmic time, color coded by the gas fraction (redder is low gas fraction and bluer are higher gas fraction). The size of the marker represents the fraction of viewing angles classified into a specific group at a specific time. Red dashed lines show the occurrence

CHAPTER 4. VELA SIMULATIONS

Group	-1	0	1	2	4	5	6	8	9
VELA02MRP	0.8	16.9	1.1	0.0	7.9	1.6	62.9	3.0	5.7
VELA03MRP	0.0	1.9	1.5	2.7	10.2	1.0	78.8	2.9	1.0
VELA04MRP	1.0	34.9	0.0	0.0	12.3	0.3	45.9	4.8	0.7
VELA05MRP	0.4	42.8	4.2	0.8	12.7	3.4	28.0	7.2	0.4
VELA12MRP	0.6	8.4	0.0	0.6	47.2	2.8	10.1	30.3	0.0
VELA14MRP	0.0	30.4	3.8	0.0	10.1	1.3	30.4	19.0	5.1
VELA15MRP	0.0	12.7	4.2	0.8	47.5	5.0	9.3	19.7	0.8
VELA26MRP	0.5	37.7	0.0	0.0	12.1	0.2	24.9	23.6	1.0
VELA27MRP	0.4	14.8	3.1	0.0	10.9	11.7	9.4	48.4	1.2
VELA28MRP	0.0	64.5	0.7	0.0	8.7	0.0	22.7	1.3	2.0
CANDELS	1.5	13.3	7.6	5.6	8.9	8.6	37.1	10.9	6.4

Table 4.1 VELA PCA Group Demographics at all viewing angles. The CANDELS sample is shown for comparison.

CHAPTER 4. VELA SIMULATIONS

Group	-1	0	1	2	4	5	6	8	9
CAMERA0	0.6	24.2	1.9	1.1	12.6	3.6	41.0	13.2	1.9
CAMERA1	0.7	24.3	3.3	0.2	24.1	2.6	31.6	11.4	1.8
CAMERA5	0.2	29.8	1.1	0.9	14.0	3.4	35.3	14.3	1.1
CAMERA6	0.2	24.5	1.7	0.2	18.6	2.2	36.4	14.7	1.5
CAMERA7	0.6	26.1	0.4	1.1	12.7	1.9	37.5	17.2	2.4
CAMERA8	0.0	28.4	1.1	0.0	16.3	1.1	36.7	14.9	1.5
CANDELS	1.5	13.3	7.6	5.6	8.9	8.6	37.1	10.9	6.4

Table 4.2 VELA PCA Group Demographics by Camera Angle. CAMERA0 represents the face-on view and CAMERA1 represents the edge-on view. The remaining cameras are from random angles. The CANDELS sample is shown for comparison.

CHAPTER 4. VELA SIMULATIONS

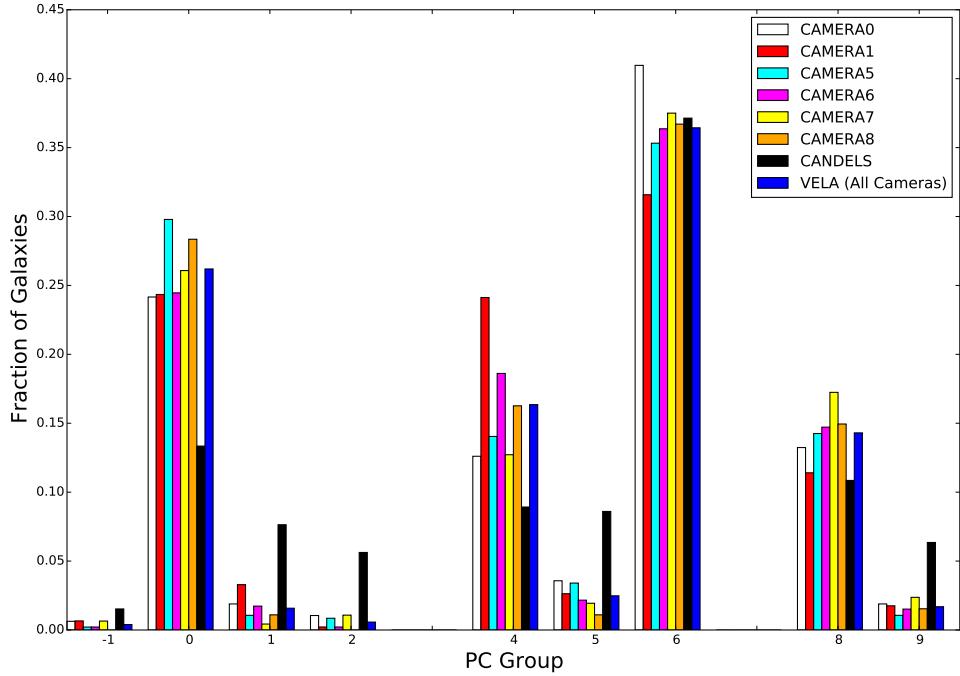


Figure 4.1 Histogram of VELA galaxies (for all time steps and camera angles) and the CANDELS sample binned by PC group. Similar to observed galaxies there is a overabundance of group 6 galaxies. These are the galaxies noticeable for their small, compact size and their lack of disk features. Overall VELA galaxies are far less disk dominated than the CANDELS sample. Many fewer VELA galaxies are classified as group 1, 2 or 5 than CANDELS galaxies.

CHAPTER 4. VELA SIMULATIONS

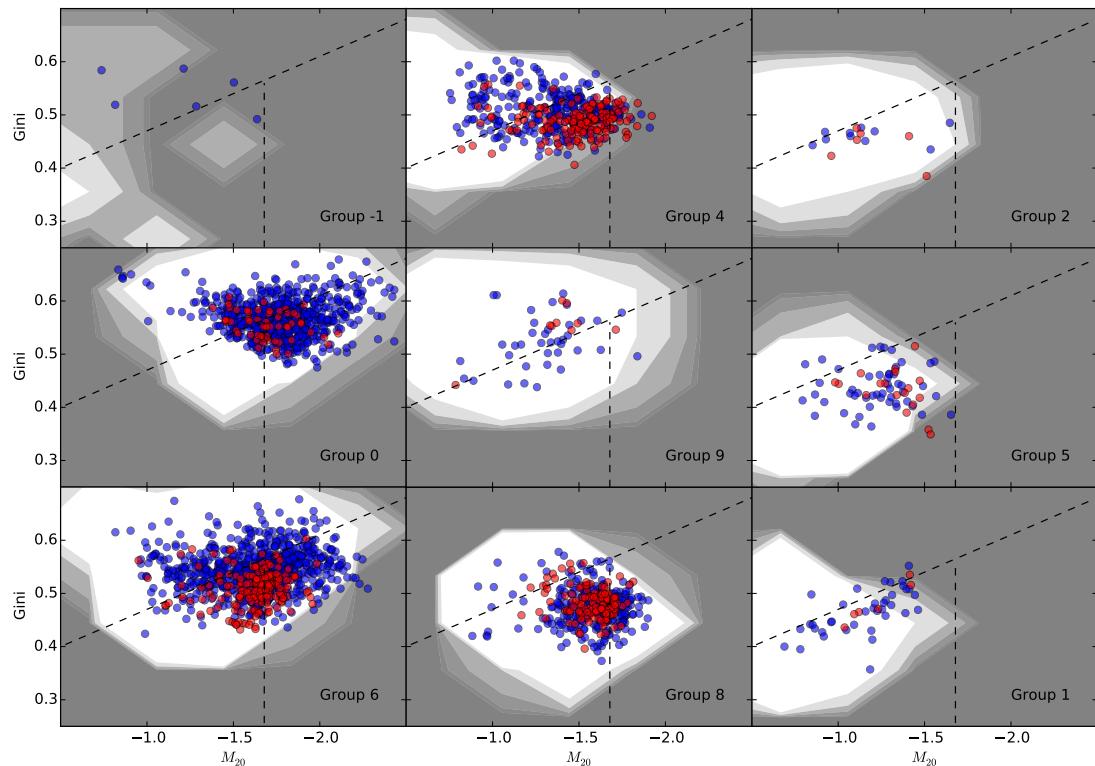


Figure 4.2 Gini– M_{20} plot for all VELA galaxies binned by redshift $z \geq 2$ (red) and $z < 2$ (blue). White contours represent the location of CANDELS galaxies defining the specific PC groups.

CHAPTER 4. VELA SIMULATIONS

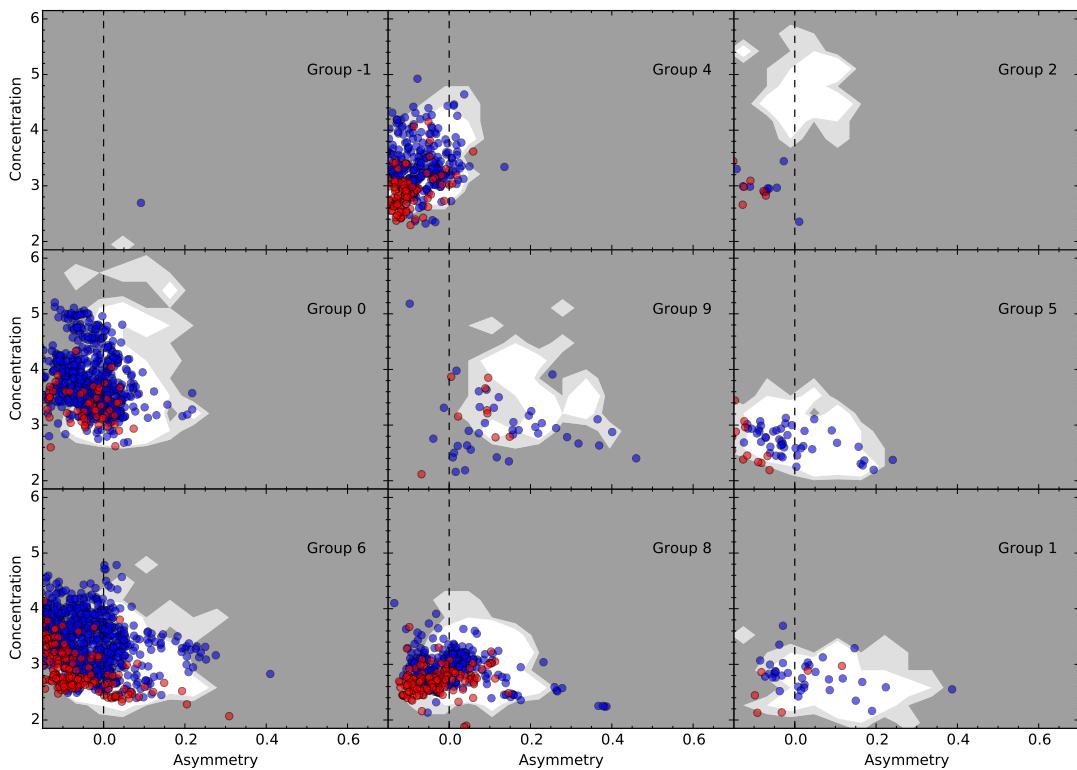


Figure 4.3 Concentration–Asymmetry plot for all VELA galaxies binned by redshift $z \geq 2$ (red) and $z < 2$ (blue). White contours represent the location of CANDELS galaxies defining the specific PC groups.

CHAPTER 4. VELA SIMULATIONS

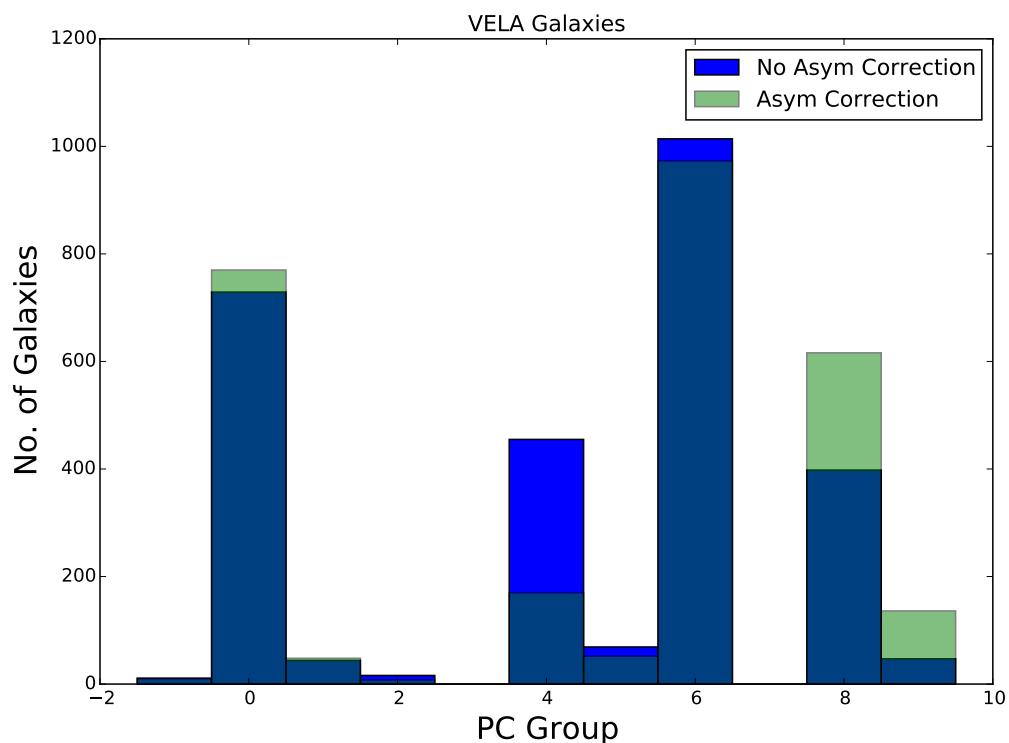


Figure 4.4 Histogram of VELA galaxies (for all time steps and camera angles) binned by PC group following an Asymmetry correction. Corrections to the asymmetry statistic change the classifications of group 4 galaxies to either group 8 or 9.

CHAPTER 4. VELA SIMULATIONS

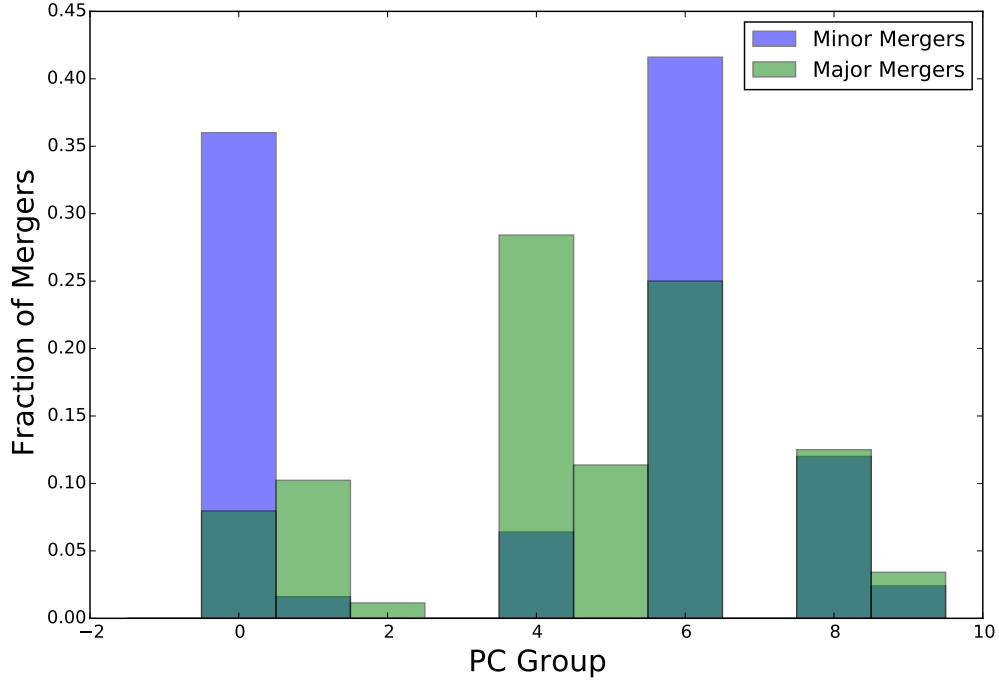


Figure 4.5 Histogram of PC group classifications at the exact time step of minor mergers ($\left(\frac{\text{ex-situ } M_{*(t+1)}}{\text{ex-situ } M_{*(t)}} > 1.1 \right)$) and major mergers ($\left(\frac{\text{ex-situ } M_{*(t+1)}}{\text{ex-situ } M_{*(t)}} > 1.3 \right)$) by PC group classification. The majority of minor mergers are classified into groups 0 and 6, while major mergers are mainly groups 4 and 6. Major mergers are also disproportionately classified into groups 1 and 5. Since much of the analysis of this chapter deals with time delays between morphology and physical parameters we will also investigate the group classifications of galaxies in time steps preceding and following the instance of a merger when determined via ex-situ M_* .

CHAPTER 4. VELA SIMULATIONS

of a minor merger and blue dashed lines show the occurrence of a major merger. We notice that VELA02 starts as a disk-dominated group 1 galaxy but quickly transforms into a bulge-dominated group 6 galaxy. For the majority of its life the galaxy alternates between group 0 and 6 designations. The vertical dashed lines represent minor mergers (red) and major mergers (blue). Following a major merger the galaxy becomes more disturbed as shown in the transition from group 6 to group 9 or group 4. Group 9 is characterized by the amount of bulge-dominated irregular galaxies seen visually. VELA02 experiences group 9 status following both major merger events. We notice that major mergers presage a morphological and gas mass change. The first major merger disturbs the galaxy while simultaneously causing a decrease in the gas fraction. The merger in effect “uses up” the available gas. The second merger leads to a short lived more gas rich disk+bulge system that quickly settles back into a gas poor bulge dominated system. The minor mergers do not have a significant impact on the gas fraction or morphological structure of the galaxy. The galaxy alternates from group 0 to group 6 and back again meaning a small disk component is forming and dissipating, which could be the result of the minor mergers, but the evidence is not overwhelmingly strong either way.

For VELA27, in figure 4.6, amidst $1.5 < z < 2.2$ the galaxy alternates between a bulge-dominated and intermediate morphology. At lower redshift $z < 1.5$ the disk of the galaxy begins to dominate the morphology. VELA27 is notable for being the galaxy that is disk-dominated for a significant amount of time at lower red-

CHAPTER 4. VELA SIMULATIONS

shift. VELA27 experiences a major merger and subsequently becomes more bulge-dominated and more compact. Following the major merger the gas fraction decreases but it is hard to ascertain whether this decrease is due to the merger or simply a secular process of gas gradually turning into stars or being expelled from the galaxy.

No matter the initial morphology, following a major merger a galaxy experiences a strong transformation. When bulge-dominated VELA galaxies experience a major merger the bulge becomes disturbed and we see group 6 galaxies transformed into more disk-dominated morphological classes. This could be an example of a primary galaxy accreting the merger-disrupted satellite galaxy. We see evidence for this in Figure 4.6. Following the accretion of satellite by the primary a short time period exists where the bulge-dominated galaxy has a visible disk component before the disk is consumed by the bulge. Whereas when a disk-dominated VELA galaxy experiences a major merger the bulge becomes stronger, the galaxy becomes more compact and galaxies become naked bulge-dominated group 6 galaxies. These occurrences are rare since only a VELA galaxies are observed as disk-dominated.

Minor mergers (shown by red dashed lines in Figure 4.6) have less impact on the morphology of the total galaxy. Minor mergers may or may not cause the morphology of the galaxy to change. If the galaxy morphology does change following a minor merger the change is usually short lived and the galaxy will revert to the pre-minor merger morphology.

Figure 4.7 inverts the relationship shown in Figure 4.6 to show gas fraction as a

CHAPTER 4. VELA SIMULATIONS

function of lookback time, color coded by PC group. Bulge dominated groups (such as group 6,0 and 9 are red, and groups 1, 2, and 5 are blue, intermediate groups 4 and 8 are shades of blue/red-ish white) dominate the overall morphology of our galaxy sample. Most simulations end up with no gas by $z = 1$ and by which time the galaxy has become bulge-dominated (except for VELA27). VELA27 is unique because even as the gas fraction decreases the disk continues to grow and become more dominant. The loss of gas appears to be related to a major merger occurring right as the gas fraction reaches a maximum.

In many instances major mergers (and to a lesser extent minor mergers) occur at a local maximum of gas fraction followed by a steady decline until the gas reservoir is empty.

4.4 Time Series Cross-Correlations

Gas accretion and other phenomena are believed to be an important driver of galaxy assembly (Snyder et al., 2015a,b). For this reason it is important to investigate relationship between the rate of numerous physical features (such as the rate of inflowing gas mass or ex-situ stellar mass) and the galaxy morphology. The VELA simulations provide quantitative morphological and physical galactic measurements that can be analyzed to determine if morphology causes physical processes (such as sSFR or gas fraction) or vice versa.

CHAPTER 4. VELA SIMULATIONS

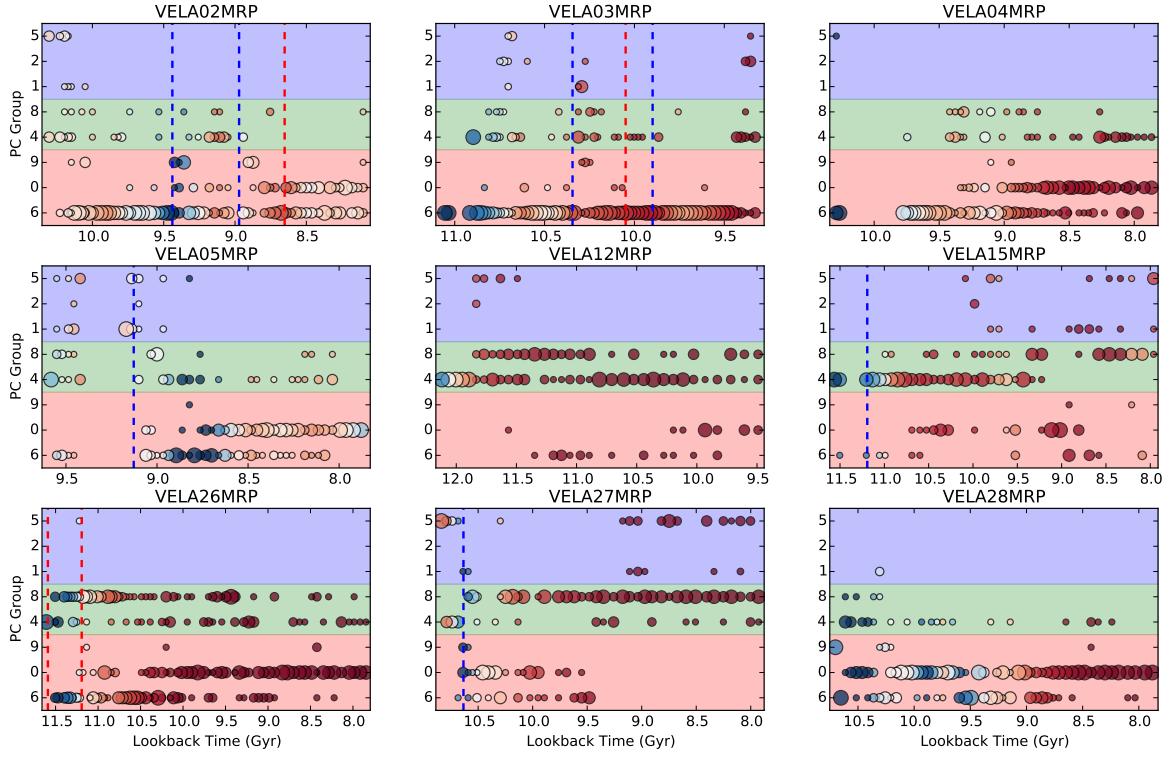


Figure 4.6 The PC group as a function of redshift for all VELA galaxies at all viewing angles. The size of the marker demonstrates the number of viewing angles classified into the group. The shaded regions represents groups defined by their: bulge-dominated appearance (*red*), disk-dominated appearance (*blue*), intermediate appearance (*green*). Each time step is colored based on the gas fraction (red is lower gas fraction, white is intermediate gas fraction and blue is high gas fraction). Red-dashed lines represent minor mergers (ex-situ M_* increases between 10-30%). Blue-dashed lines represent major mergers (ex-situ M_* increases $> 30\%$).

CHAPTER 4. VELA SIMULATIONS

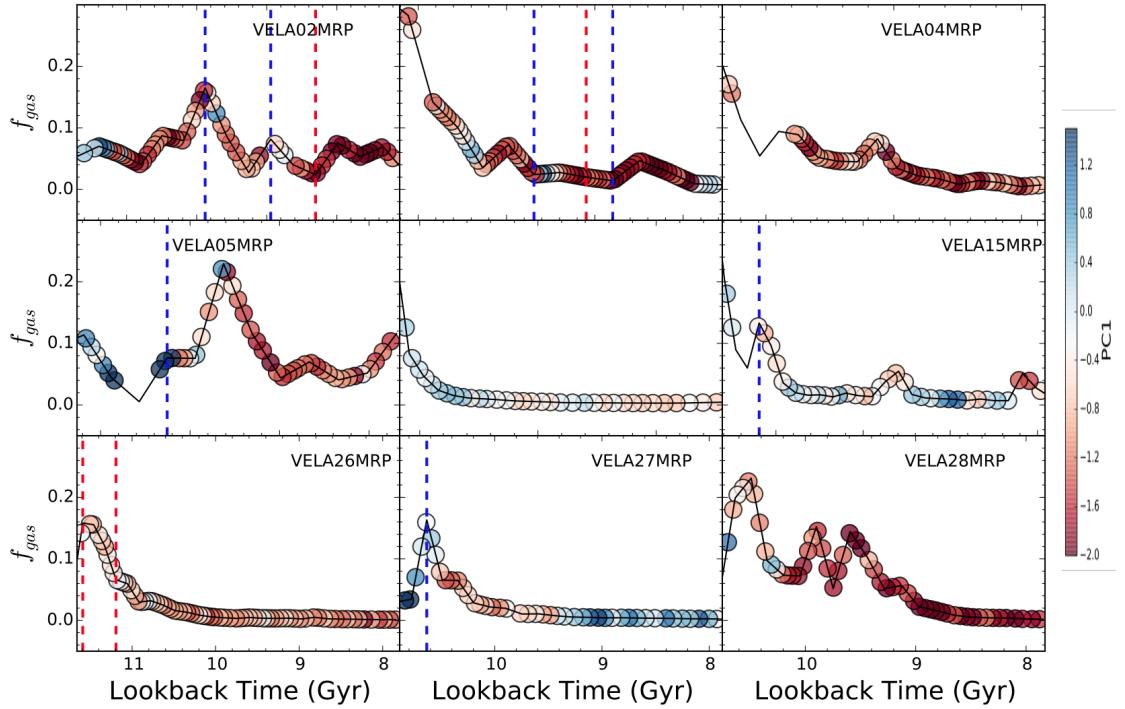


Figure 4.7 The gas fraction as a function of lookback time for all VELA galaxies color coded by the average PC1 of all viewing angles (redder = more prominent bulge). Red-dashed lines represent minor mergers (ex-situ M_* increases between 10-30%). Blue-dashed lines represent major mergers (ex-situ M_* increases $> 30\%$).

CHAPTER 4. VELA SIMULATIONS

Since indicators of mergers and gas accretion have been shown to be short lived phenomena (Lotz et al., 2008; Snyder et al., 2015b) it is important to have a multitude of observed time steps and viewing angles. Morphological measurements such as PC1 and PC2 can quantify the strength and concentration of the central bulge. PC3 indicates the level of asymmetry in the galaxy. The VELA simulation suite follows individual galaxies as they evolve with measurements of the gas fraction, star-formation rate, ex-situ mass and dark matter mass. Ex-situ stellar mass can be used as an indicator of merger activity (Zolotov et al., 2015). The gas fraction represents the amount of fuel remaining to form stars. The star-formation rate can be used to indicate the evolutionary stage of a galaxy (Barro et al., 2013). The dark matter mass rate of change represents the mass accretion history of the dark matter halo which in Λ CDM can be used to understand formation scenarios (White & Rees, 1978; Zhao et al., 2003). By comparing the morphological statistics to physical features we can study the causality between these phenomena and the amount of time needed to influence one another.

The time series analysis of morphology and physical galactic properties can help further prove the strong relationship between SFR and morphology (e.g. Wuyts et al., 2011; Lee et al., 2013), the effects of mergers upon morphology (e.g. Lotz et al., 2011), and the connection between gas fraction and morphology (e.g. Lotz et al., 2010a; Huertas-Company et al., 2015). We can compare the time dependence of physical parameters and use that information to delve into the causation of such

relationships.

4.4.1 Discrete Correlation Function

We use time series analysis between two 1D signals to determine the amount of correlation and the time delay. We use the discrete correlation function (DCF; Edelson & Krolik, 1988) because our time series data are irregularly sampled. The physical gas and stellar parameters are sampled much more sparingly than the non-parametric morphology. However, the physical parameters are measured since $z \sim 7$, whereas morphology is only measured since $z \sim 3$. We use the python code developed by Robertson et al. (2015) to calculate the discrete correlation function. The DCF corrects for the spurious correlations between time series (Edelson & Krolik, 1988).

Calculating the DCF first requires an unbinned DCF (UDCF, equation 4.1). a_i and b_j are the respective time series with mean values of \bar{a} and \bar{b} and variances of σ_a^2 and σ_b^2 . The next step is to bin all time delays $\Delta t_{ij} = t_{a,i} - t_{b,j}$ in the range $\tau - \frac{\Delta\tau}{2} \leq t_{ij} \leq \tau + \frac{\Delta\tau}{2}$ and divided by the number of data points (N) in the bin (equation 4.2).

Positive values of the DCF indicate either both a_i and b_j are increasing or both are decreasing. Meanwhile, a negative DCF value indicates one time series increases while the other decreases. If the correlation (either positive or negative DCF) peaks at $-\Delta t_{ij}$ (or equivalently³ $-t_{delay}$) values we say that a_i **leads** b_j , whereas a correlation

³We use Δt_{ij} and t_{delay} terminology interchangeably

CHAPTER 4. VELA SIMULATIONS

peaking at $+\Delta t_{ij}$ ($+t_{delay}$) values we say that a_i **lags** b_j . The width of these peaks is essential to determine the importance of the correlations. If the peaks in time delay are too large (>1 Gyr) then the time series are likely not correlated.

$$UDCF_{ij} = \frac{(a_i - \bar{a})(b_j - \bar{b})}{\sqrt{\sigma_a^2 \sigma_b^2}} \quad (4.1)$$

$$DCF(\tau) = \frac{1}{N} \sum UDCF_{ij} \quad (4.2)$$

Discrete correlation functions have been used in: AGN reverberation mapping studies to determine the time lag between the visibility of different emission lines (Haas et al., 2011), the variability of blazars (Agarwal et al., 2015) and spectral variability of stars (e.g. Gaur et al., 2015).

Significance testing is based on MC sampling of cross correlations between one parameter (in our case a PC) and a random distribution with the same average and standard deviation as another parameter (such as stellar mass, gas fraction,etc.). The $n\sigma$ values are determined from a two-sided Student's t distribution as shown in Eq. 4.3. The A_n value depends on the the degrees of freedom in the Student's t distribution which we define as the number of data points in each corresponding bin of the DCF and the level of significance achieved ($1\sigma, 2\sigma$ and 3σ corresponding to 68.7%, 95%, and 99.5% importance). Correlations stronger than 3σ are significant.

$$n\sigma = A_n \frac{\sqrt{variance}}{\sqrt{N}} \quad (4.3)$$

CHAPTER 4. VELA SIMULATIONS

For the sake of this analysis, we consider correlations that are stronger than 3σ for at least 100 Myrs but not more than 1 Gyr to be important indicators of the causation between morphology and galactic physical parameters.

In many of the cross-correlations there is a primary peak and multiple subsequently less important peaks. These secondary peaks are likely the result of the multiply peaked (e.g. PC1 or gas fraction) time series. The cross-correlation matches secondary peaks of one time series with the primary peak of the other time series leading to less important cross-correlation peaks. Analysis in subsequent sections will mainly pertain to the primary peak. In rare cases multiple peaks are nearly equally important and are addressed in turn.

We tested the relationship between morphology and physical properties of the simulation (such as ex-situ stellar mass fraction, dark matter mass fraction, gas fraction and star-formation rate) by cross-correlating the time series of each property with the PC results. We then stack the results of the discrete cross-correlation functions from all the VELA galaxies available. The peaks in the cross-correlation function correspond to time lags or time leads between the time series of PCs and the time series for a physical parameter. The time leads and/or lags help us determine the cause and effect demonstrated between morphology and physical properties. All the Δt values are in relation to the physical parameter time series leading or lagging the PC time series.

PC1 correlates most strongly with the, ex-situ stellar mass fraction, sSFR and

CHAPTER 4. VELA SIMULATIONS

gas fraction. PC2 anti-correlates very weakly with gas fraction, ex-situ stellar mass and dark matter mass fraction rate of change. PC3 correlates most strongly with sSFR and gas fraction. These relationships are also dependent upon the region of the galaxy investigated. Some relationships are stronger in the central kpc (indicating an influence on compaction), while other relationships are more important for the entire galaxy.

The morphological measurements of the VELA simulations were initially performed only on galaxies brighter than 24.5 magnitude. Correspondingly, only morphological measurements from the lower end of the redshift range exist, when the galaxies were sufficiently massive and bright. When the galaxy is too small or insufficiently massive the non-parametric morphology measurements are not reliable (Lotz et al., 2004; Grogin et al., 2011; Peth et al., 2016).

Figure 4.8 shows the PC1 (disk growth/bulge weakening) values for the VELA galaxies as a function of cosmic time. Larger values of PC1 indicate disk growth while smaller values of PC1 indicate stronger bulges. The error bars are determined from the standard deviation of PC values from all viewing angles. Figures 4.9 - 4.16 show PC2, PC3 and physical parameters as functions of time. In general, PC values are consistent across viewing angle, only a few outliers are noticeable (particularly PC2 and PC3 for VELA04). The star-formation rates and ex-situ stellar mass fractions for each VELA galaxy are quite distinct amongst one another.

The next sections show an example of a cross-correlation between the morphology

CHAPTER 4. VELA SIMULATIONS

and physical properties of a VELA galaxy. In particular, the focus is on the cross-correlation amongst PC1 and physical properties for VELA02. VELA02 experiences a few mergers and gas accretion which provide an opportunity to observe how these mechanisms are related to morphology.

The Evolution of VELA02

VELA02 represents an interesting galaxy simulation to focus on because between $1 \lesssim z \lesssim 1.8$ there were two minor mergers, 1 major merger and a continual stochastic accretion of gas onto the galaxy. At $z \sim 1.8$ the galaxy is fairly disk-dominated since the bulge is weak. Gas is accreted fairly strongly until the first minor merger occurs which trigger bursts of star-formation contributing to initial visible disturbances. Meanwhile, the stars that are formed during the burst lose angular momentum and accreted gas help to build the bulge. The gas reservoir eventually becomes tapped and star-formation begins to quench. Soon thereafter, a major merger completely disrupts the visual morphology (Snyder et al., 2015b) which leads to final state for the galaxy as bulge-dominated spheroid.

PC1 - sSFR

Zolotov et al. (2015) proposes peak SFR occurs at peak compaction. Immediately prior to peak sSFR, the sSFR increases more steeply in the inner kpc than for the total galaxy. After peak compaction and SF are reached the SFR declines slower

CHAPTER 4. VELA SIMULATIONS

in the 10 kpc radius as compared to the inner kpc. We investigate the relationship between sSFR and galaxy morphology. The structure of a galaxy is often tied to the star-formation characteristics of the system. The first step towards investigated the hypothesis of a sSFR-compaction connection in the context of morphology is to see how the SFR and structure of the galaxy are related in time.

In figure 4.17 the top panel shows the specific star-formation rate of VELA02 (as seen face-on) as a function of time, the middle panel shows PC1 of VELA02 (as seen face-on) as a function of time and the bottom panel shows the cross-correlation between sSFR and PC1. The bottom panel (also known as a correlogram) shows the correlation as a function of delay time between the two time series.

The correlation is multiply peaked. There are peaks of -1 Gyrs, -400 Myrs, 100 Myrs, 800 Myrs and 1.28 Myrs. The multiple peaks are probably the result of the multiple peaks in PC1. The strongest positive correlation occurs with sSFR lagging PC1 \sim 800 Myrs. This correlation barely rises above the 3σ threshold, and thus the importance of this correlation is not very strong. The multiple peaks sSFR indicate VELA02 is compacting then growing then re-compacting numerous times. At least for VELA02 the relationship between bulge strength/disk growth and star-formation rate is quite complicated.

PC1 - f_{gas} and \dot{f}_{gas}

Zolotov et al. (2015); Dekel et al. (2013) refer to the rapid influx of gas into the

CHAPTER 4. VELA SIMULATIONS

central kpc as the “compactification” of the galaxy. At this time the galaxy begins experiences a peak in star-formation followed by a gradual quenching.

In figure 4.18 the top panel shows the gas fraction ($M_{gas}/(M_* + M_{gas})$) in the central 1 kpc of VELA02 (as seen face-on) as a function of time. The middle panel shows the PC1 of VELA02 (as seen face-on) as a function of time, and the bottom panel shows the cross-correlation between f_{gas} and PC1. For VELA02 f_{gas} lags PC1 by ~ 800 Myrs with a positive correlation and an anti-correlation with gas fraction leading PC1 by ~ 1 Gyr. However, both of these correlations are just below the 3σ threshold for significance. There are three additional local maxima at -1.3, -0.5 and +1.3 Gyrs that are stronger than 2σ significance. Much like the correlation between PC1 and sSFR, the correlation between PC1 and gas fraction is multiply peaked. The multiple peaks indicate the multiple episodes of bulge growth and decay coexist with increasing and decreasing gas fractions.

Instead of finding a smoothly increasing function of f_{gas} with time as imagined in Zolotov et al. (2015) we find that these galaxies are much more complicated. We calculate the time derivative of the gas fraction (\dot{f}_{gas}) with the simple `python numpy` gradient function. We compare our results to the gas “inflow” as defined in Zolotov et al. (2015). In their model, gas inflows more rapidly at the central kpc than into the entire galaxy but once compaction and peak SFR are achieved the gas inflow into the central kpc drops off more steeply than for the entire galaxy. The inflow rate of gas becomes overwhelmed by the combined star formation rates and gas outflow rates

CHAPTER 4. VELA SIMULATIONS

that are increasingly depleting the gas reservoir.

In figure 4.19 the top panel shows the rate of change of gas fraction into the central 1 kpc of VELA02 (as seen face-on) as a function of time. The middle panel shows the PC1 of VELA02 (as seen face-on) as a function of time, and the bottom panel shows the cross-correlation between \dot{f}_{gas} and PC1. We find a strong positive correlation with \dot{f}_{gas} lagging PC1 by ~ 940 Myrs. This correlation only barely surpasses the 3σ importance threshold, but is the only correlation to be stronger than 2σ at any time delay. The relative strength of the correlation could be proven further by stacking results from other VELA simulation galaxies. In this case, the maximum \dot{f}_{gas} occurs 900 after the galaxy has grown a disk.

PC1 - ex-situ- M_*/M_* and ex-situ- \dot{M}_*/M_*

Ex-situ stellar mass refers to the stellar mass accreted by the central galaxy during a merger or tidal disruption event. The amount of ex-situ stellar mass tells us the strength of the merger, which we define as a major merger if the increase of ex-situ stellar mass is $> 30\%$ and a minor merger if the increase is between 10-30%. Understanding the time delay between when a merger occurs and when the morphology indicates one happened is a very important piece of information telling us how long a merger may be visible morphologically and thus allow us to understand how important mergers are to galaxy evolution.

In figure 4.20 the top panel shows the ex-situ stellar mass/ M_* in the central kpc

CHAPTER 4. VELA SIMULATIONS

of VELA02 (as seen face-on) as a function of time. The middle panel shows the PC1 of VELA02 (as seen face-on) as a function of time, and the bottom panel shows the cross-correlation between ex-situ- M_*/M_* and PC1. We find a strong positive correlation with ex-situ stellar mass lagging PC1 by ~ 800 Myrs. Counter intuitively in this case, the bulge weakens *before* the maximum amount of ex-situ stellar mass is accreted by the galaxy. In this case the total amount of ex-situ stellar mass might not be as important as how quickly the central galaxy is accreting the satellite. There is a moderately strong and wide correlation peaked at no time lag between PC1 and ex-situ stellar mass which indicates the possibility bulge strength is influenced (relatively) instantly by a merger.

In figure 4.21 the top panel shows the inflow rate of ex-situ stellar mass into the central 1 kpc of VELA02 as a function of time, the middle panel shows PC1 of VELA02 (as seen face-on) as a function of time, and the bottom panel shows the cross-correlation between ex-situ- \dot{M}_*/M_* and PC1. We find the potential for two peaks in the correlation function, with either ex-situ- \dot{M}_*/M_* leading PC1 by ~ 400 Myrs or ex-situ- \dot{M}_*/M_* lagging PC1 by ~ 940 Myrs . However, neither of these correlations are much stronger than 3σ . The first of these peaks could be evidence that the bulge of a galaxy takes 400 Myrs to react to a merger event. The second of these peaks shows the disk of the galaxy is most dominant before the merger, which could suggest VELA02 is slowly converted from a disk-dominated to bulge-dominated galaxy while the merger occurs.

CHAPTER 4. VELA SIMULATIONS

PC1 - \dot{M}_{dm}/M_*

An increase in the dark matter mass is another important indicator of merger activity. We do not study the amount of dark matter and correlate this with morphology because dark matter is always increasing for the galaxies in the simulations and thus a cross-correlation would not be a very meaningful statistic. However, the rate at which dark matter is accreted by the galaxy can be stochastic and thus a cross-correlation is a valid measurement.

In figure 4.22 the top panel shows the rate of dark matter mass in the central 1 kpc of VELA02 (as seen face-on) as a function of time. The middle panel shows the PC1 of VELA02 (as seen face-on) as a function of time, and the bottom panel shows the cross-correlation between \dot{M}_{dm} and PC1. We find \dot{M}_{dm} leads PC1 either by ~ 40 Myrs or ~ 650 Myrs or \dot{M}_{dm} lags PC1 ~ 650 Myrs. None of these correlations are much stronger than a 3σ importance so no strong statements of causality between PC1 and dark matter mass rate can be determined.

4.4.2 Stacks of PCs and Physical Parameters Correlations

Are the correlations observed in VELA02 present for other VELA galaxies as well? If they are, how important are the cross-correlations and do these correlations occur at the same time delay which can be used to define causation? To answer these

CHAPTER 4. VELA SIMULATIONS

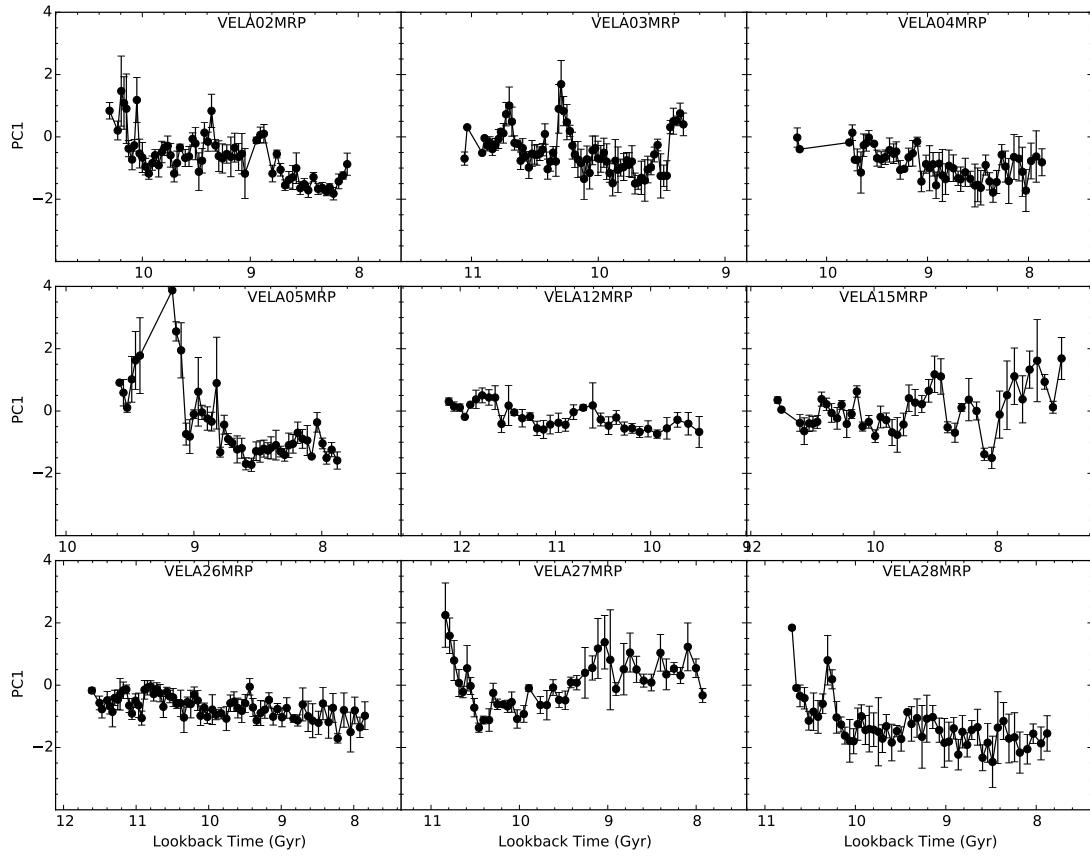


Figure 4.8 PC1 for all VELA galaxies a function of time. The error bars are determined from the standard deviation of PC1 values at each redshift for all viewing angles. There exist only a few extreme differences which could be the result of a poorly measured statistic (such as M) or a merger.

CHAPTER 4. VELA SIMULATIONS

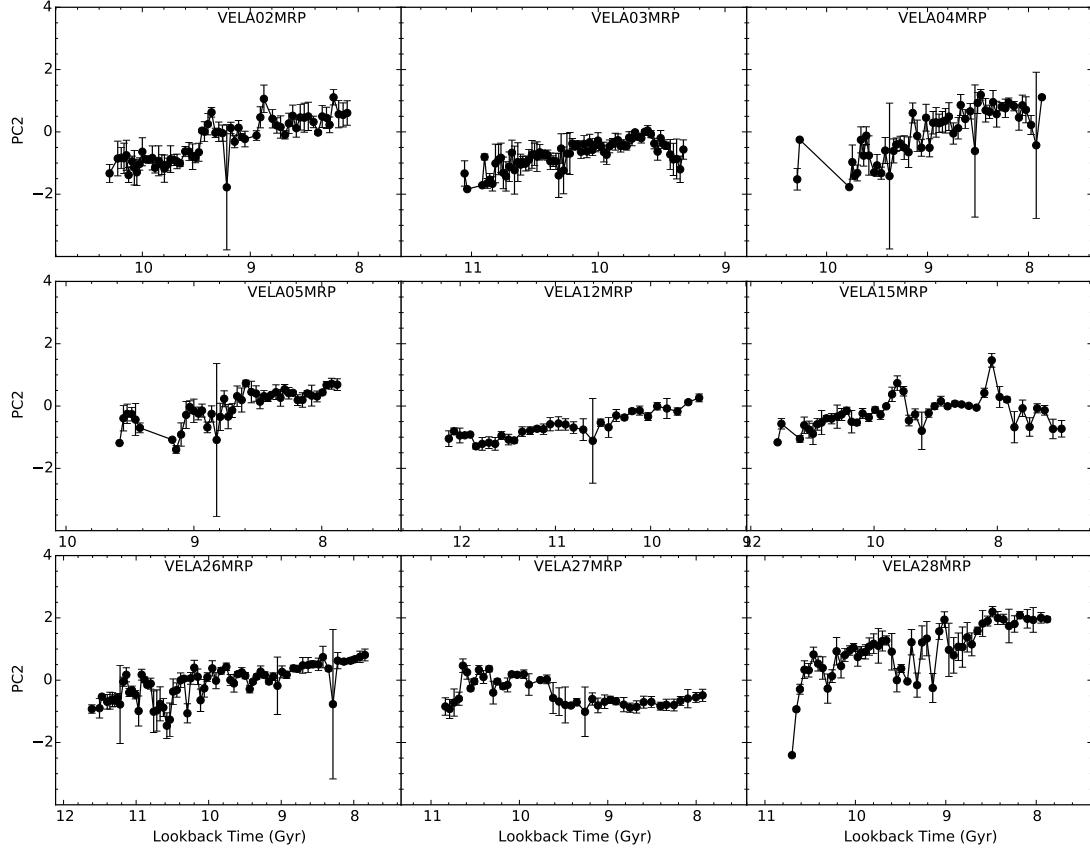


Figure 4.9 PC2 for all VELA galaxies a function of time. The error bars are determined from the standard deviation of PC2 values at each redshift for all viewing angles. There exist only a few extreme differences which could be the result of a poorly measured statistic (such as M) or a merger.

CHAPTER 4. VELA SIMULATIONS

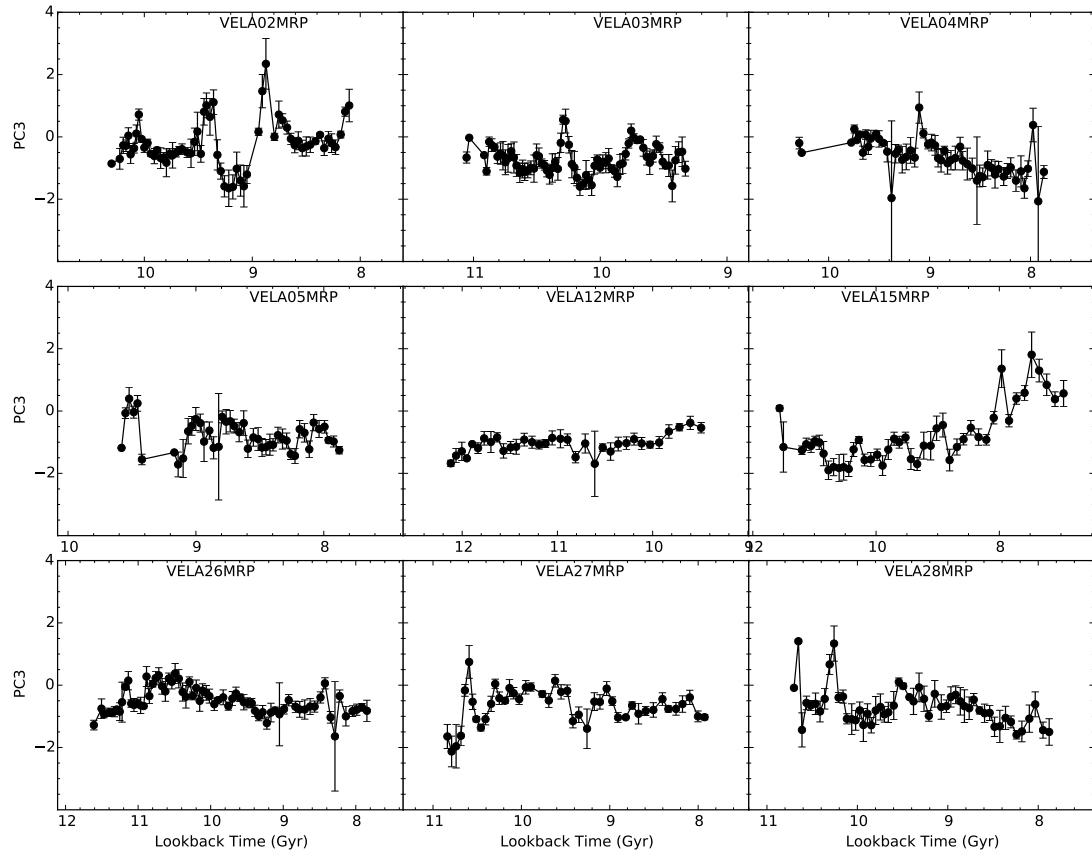


Figure 4.10 PC3 for all VELA galaxies a function of time. The error bars are determined from the standard deviation of PC3 values at each redshift for all viewing angles. There exist only a few extreme differences which could be the result of a poorly measured statistic (such as M) or a merger.

CHAPTER 4. VELA SIMULATIONS

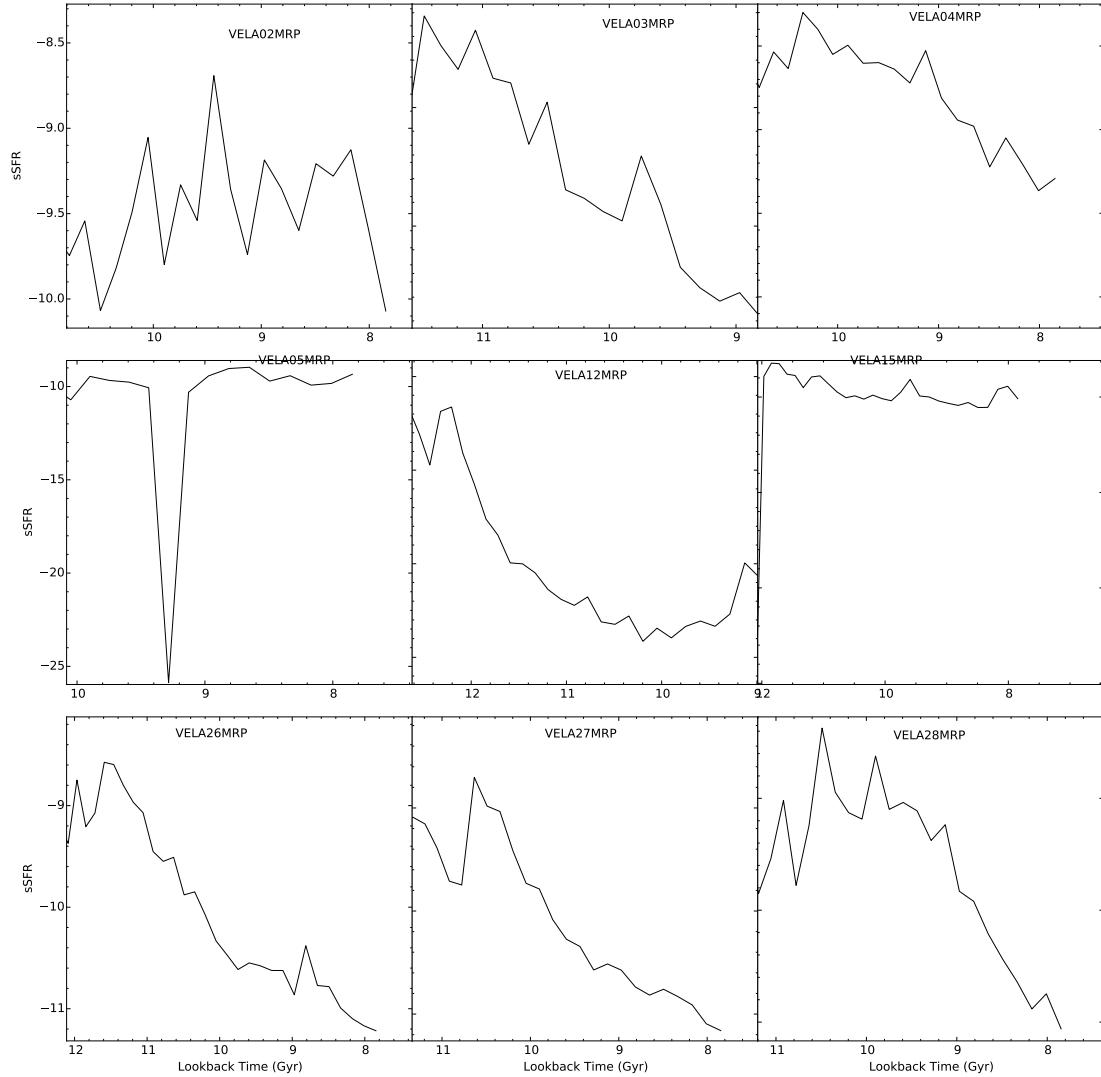


Figure 4.11 sSFR for all VELA galaxies a function of time.

CHAPTER 4. VELA SIMULATIONS

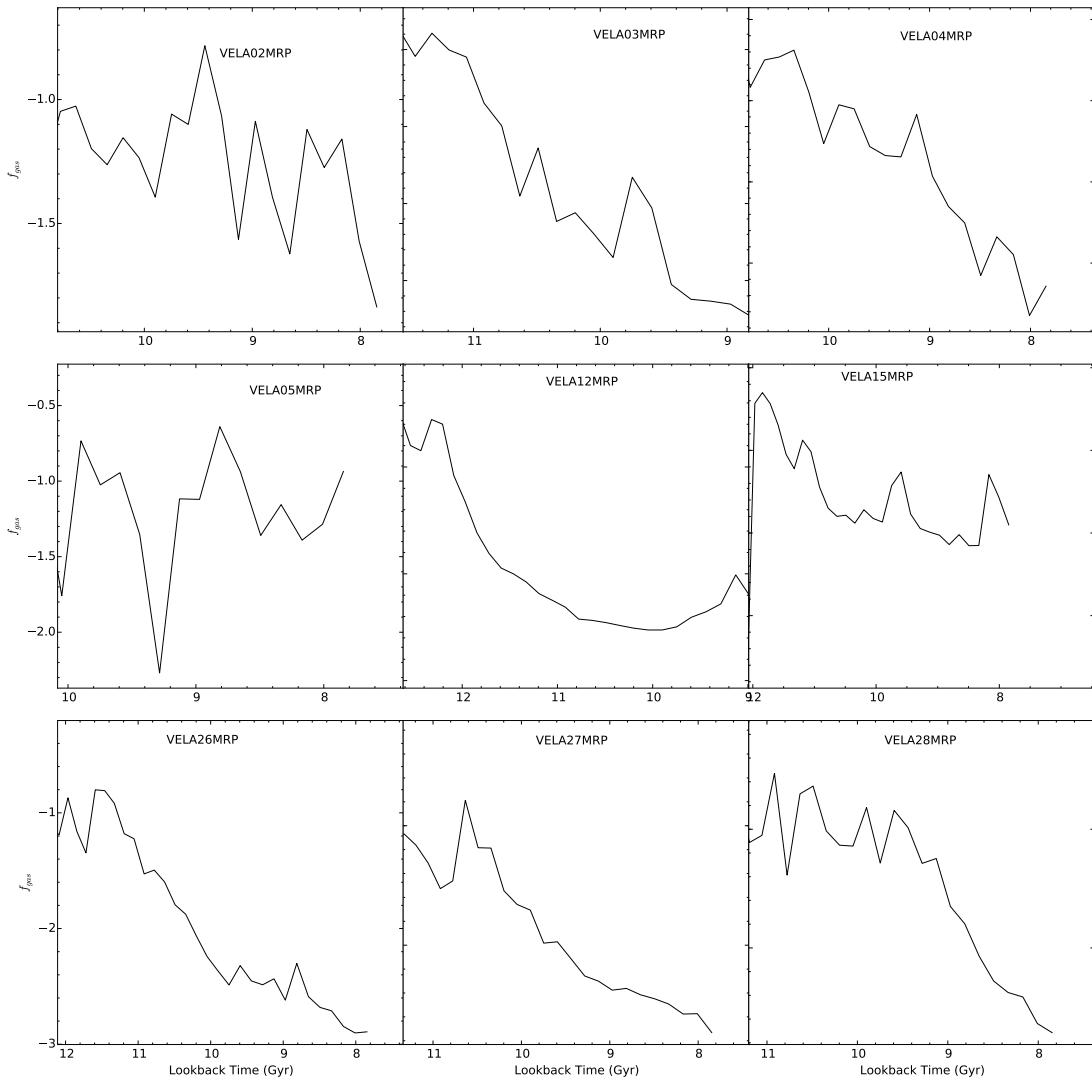


Figure 4.12 f_{gas} for all VELA galaxies a function of time.

CHAPTER 4. VELA SIMULATIONS

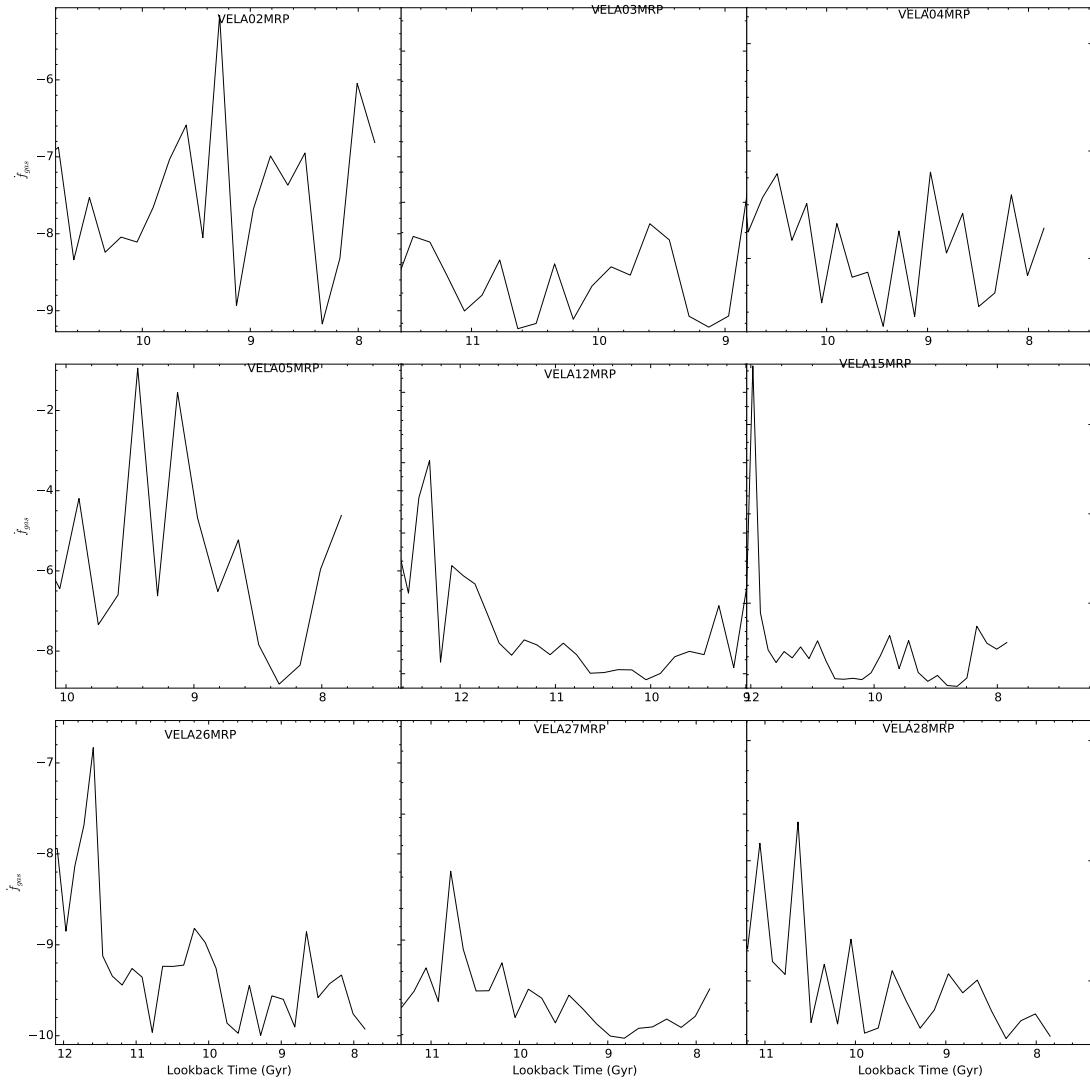


Figure 4.13 \dot{f}_{gas} for all VELA galaxies a function of time.

CHAPTER 4. VELA SIMULATIONS

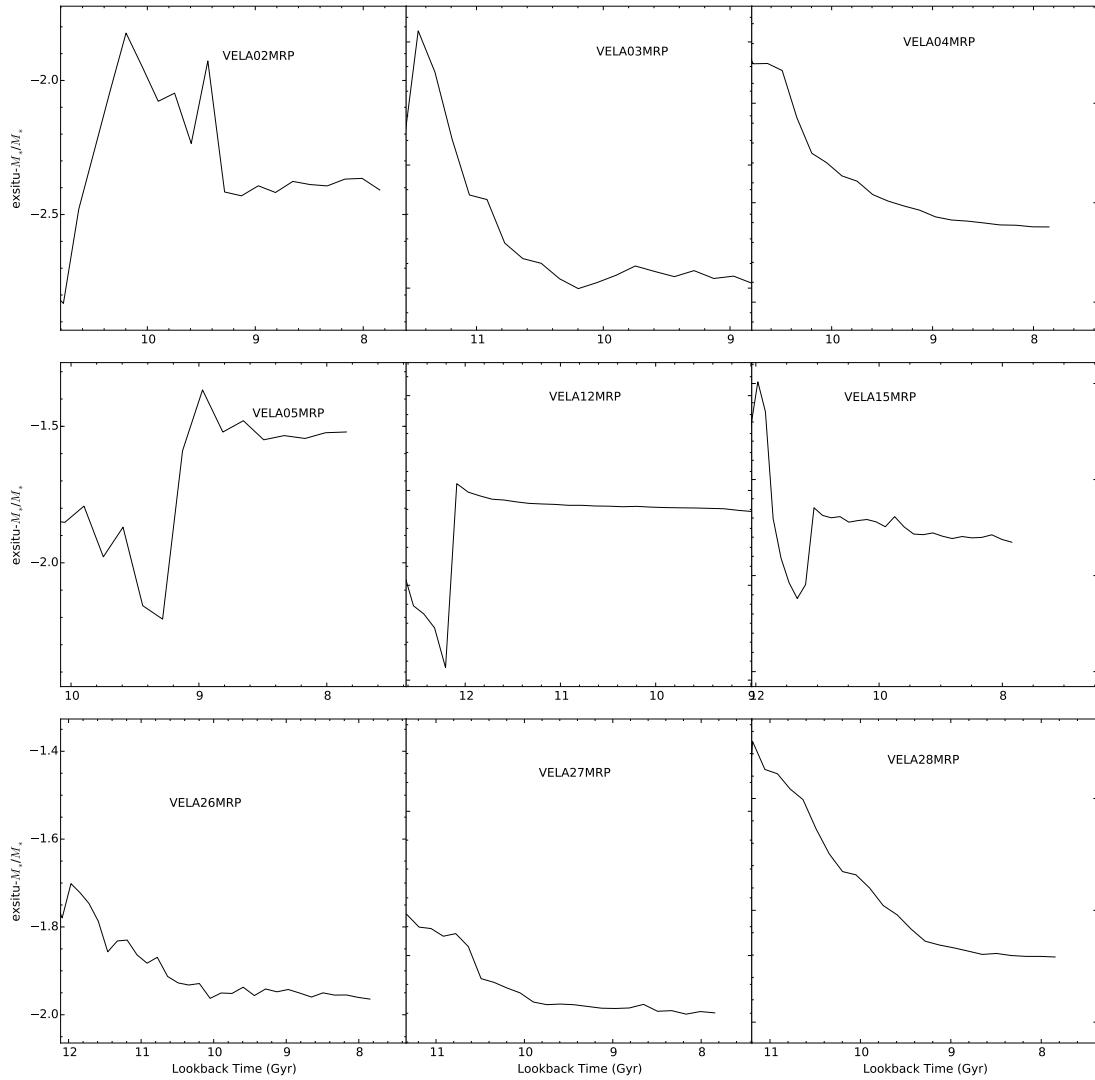


Figure 4.14 ex-situ M_*/M_* for all VELA galaxies a function of time.

CHAPTER 4. VELA SIMULATIONS

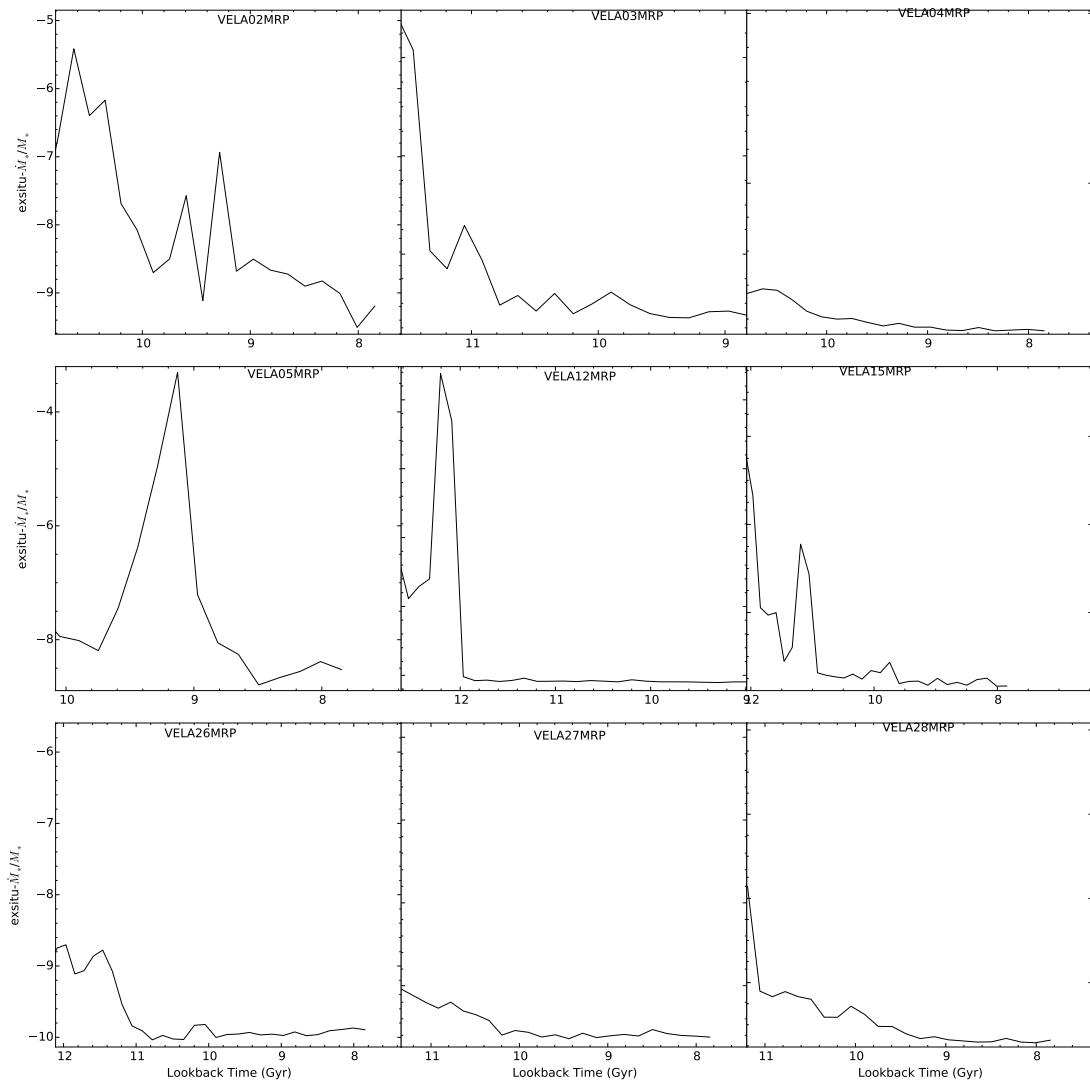


Figure 4.15 ex-situ \dot{M}_*/M_* for all VELA galaxies a function of time.

CHAPTER 4. VELA SIMULATIONS

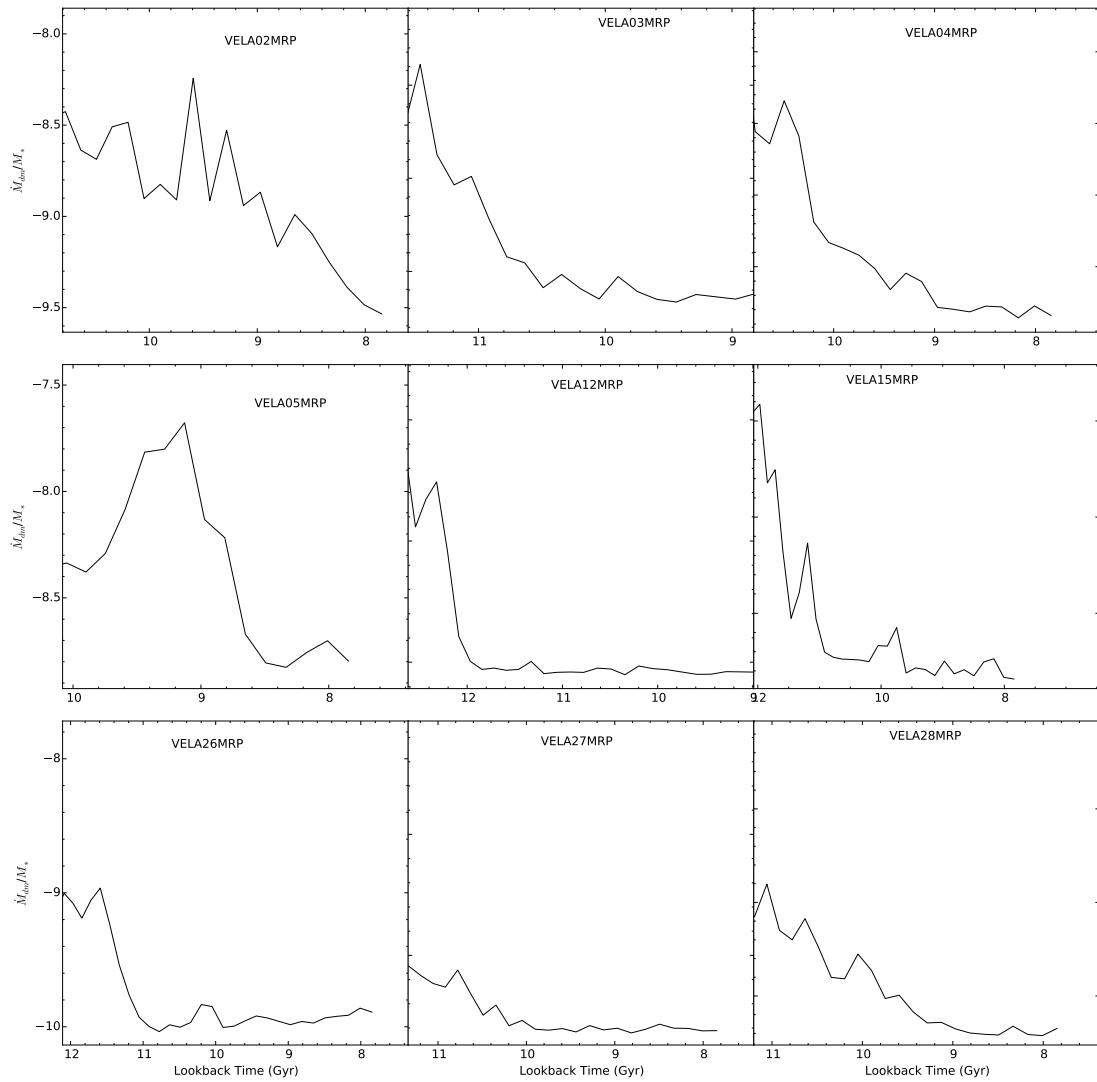


Figure 4.16 \dot{M}_{dm}/M_* for all VELA galaxies a function of time.

CHAPTER 4. VELA SIMULATIONS

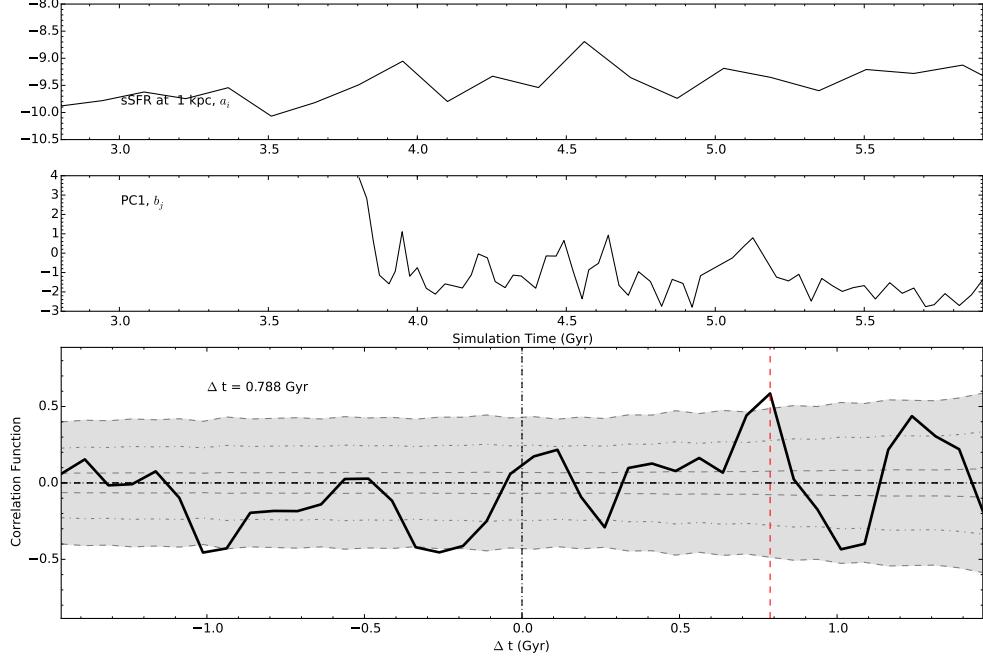


Figure 4.17 (*top panel*): Specific star-formation rate of VELA02 (as seen face-on) as a function of time. (*middle panel*): PC1 of VELA02 (as seen face-on) as a function of time. (*bottom panel*): The cross-correlation between SFR and PC1 (bulge strength). The gray regions represent the 3σ level of significance, correlations above and below this region are important. The vertical red dashed line represents the location of Δt . We find the cross correlation is multiply peaked with the strongest correlation occurring when sSFR lags PC1 by ~ 800 Myrs but anti-correlation is observed with sSFR leading by ~ 1 Gyrs or ~ 400 Myrs. A decrease in PC1 implies the bulge is strengthening, and a positive correlation implies sSFR declines as the central bulge becomes stronger. Conversely, an increase in PC1 implies sSFR increases as the disk is growing. A stronger bulge can lead to quenching, while a growing disk is associated with star-formation.

CHAPTER 4. VELA SIMULATIONS

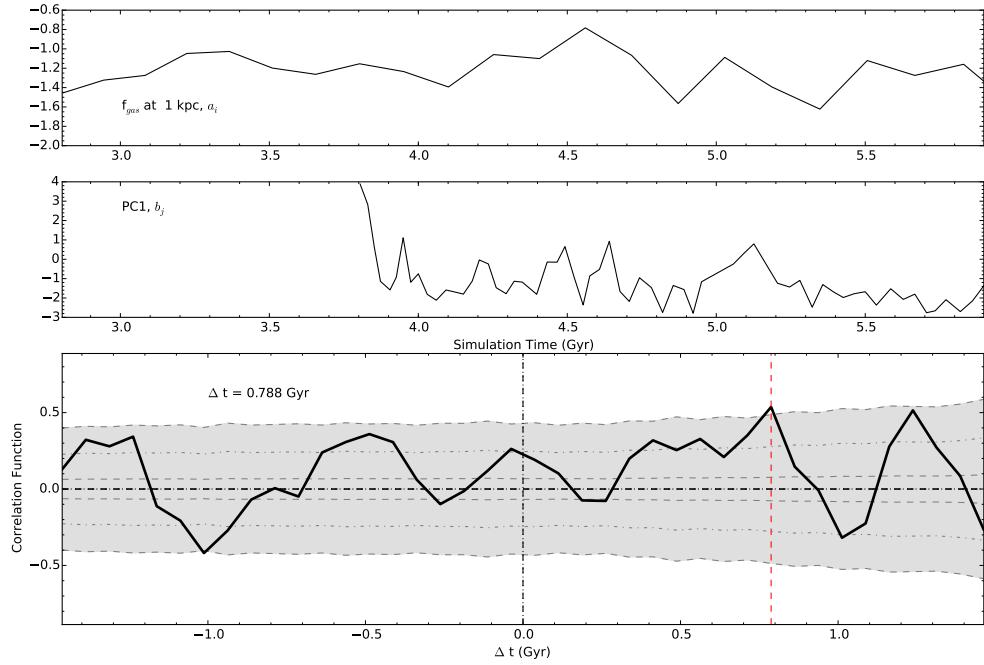


Figure 4.18 (*top panel*): Gas fraction in the central 1 kpc of VELA02 (as seen face-on) as a function of time. (*middle panel*): PC1 of VELA02 (as seen face-on) as a function of time. (*bottom panel*): The cross-correlation between f_{gas} and PC1 (bulge strength). The gray regions represent the 3σ level of significance, correlations above and below this region are important. The vertical red dashed line represents the location of Δt . We find a moderate positive correlation with f_{gas} lagging PC1 by ~ 800 Myrs. However, this correlation is not much stronger than 3σ .

CHAPTER 4. VELA SIMULATIONS

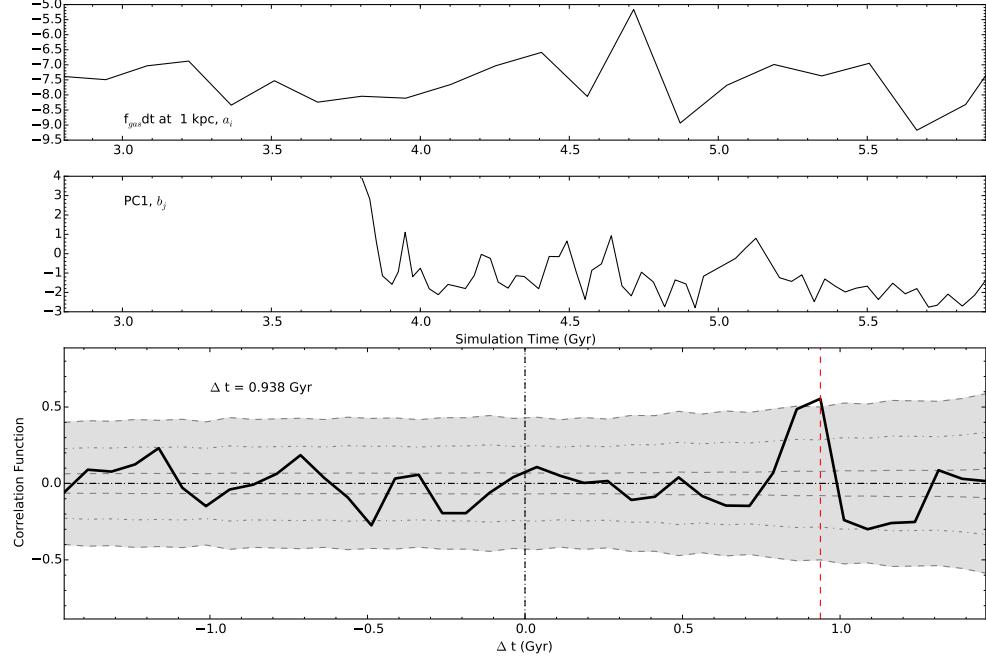


Figure 4.19 (*top panel*): Rate of gas fraction into the central 1 kpc of VELA02 (as seen face-on) as a function of time. (*middle panel*): PC1 of VELA02 (as seen face-on) as a function of time. (*bottom panel*): The cross-correlation between \dot{f}_{gas} and PC1 (bulge strength). The gray regions represent the 3σ level of significance, correlations above and below this region are important. The vertical red dashed line represents the location of Δt . The strongest correlation exists for \dot{f}_{gas} lagging PC1 by ~ 900 Myrs. This correlation is not stronger than 3σ but is the only correlation stronger than 2σ . A decrease in PC1 implies the bulge is strengthening, and a positively correlated \dot{f}_{gas} implies gas inflow slows down as the central bulge becomes stronger. Conversely, an increase in PC1 implies gas rate increases as the disk grows.

CHAPTER 4. VELA SIMULATIONS

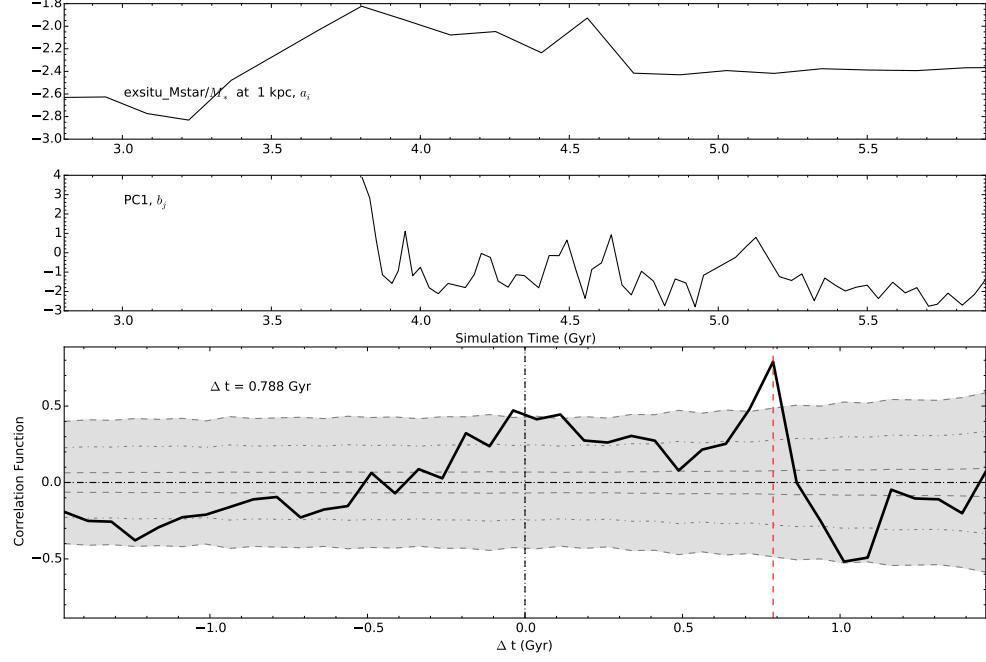


Figure 4.20 (*top panel*): ex-situ-stellar mass into the central 1 kpc of VELA02 (as seen face-on) as a function of time. (*middle panel*): PC1 of VELA02 (as seen face-on) as a function of time. (*bottom panel*): The cross-correlation between ex-situ- \dot{M}_* and PC1 (bulge strength). The gray regions represent the 3σ level of significance, correlations above and below this region are important. The vertical red dashed line represents the location of Δt . The strongest correlation occurs with ex-situ- M_* lagging PC1 by ~ 800 Myrs. This correlation suggests the bulge becomes more compact as less ex-situ stellar mass is being accreted by the galaxy.

CHAPTER 4. VELA SIMULATIONS

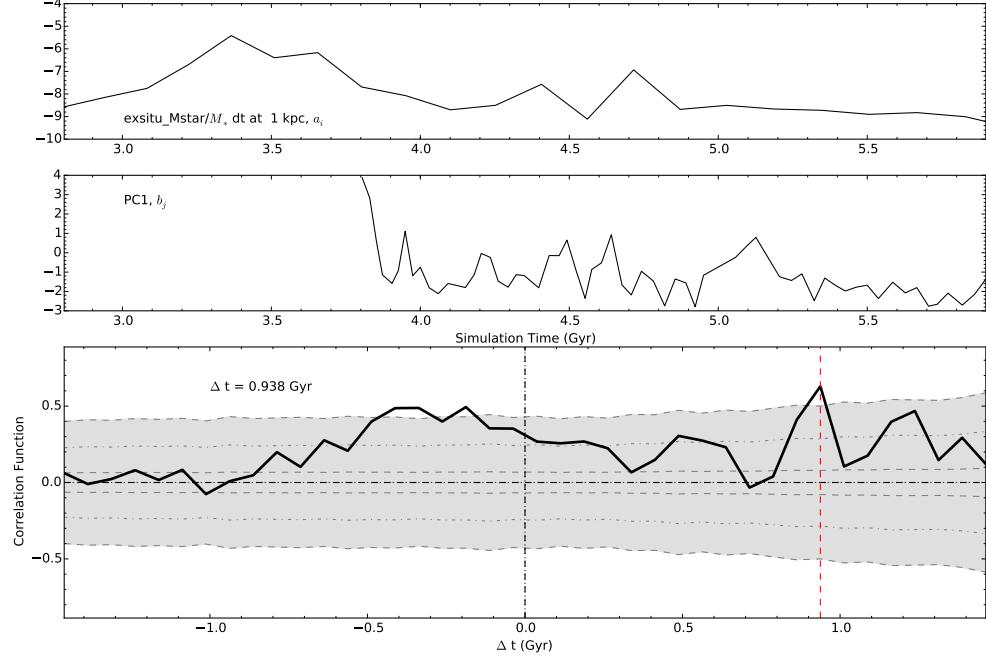


Figure 4.21 (*top panel*): Rate of ex-situ-stellar mass into the central 1 kpc of VELA02 (as seen face-on) as a function of time. (*middle panel*): PC1 of VELA02 (as seen face-on) as a function of time. (*bottom panel*): The cross-correlation between ex-situ- \dot{M}_* and PC1. The gray regions represent the 3σ level of significance, correlations above and below this region are important. The vertical red dashed line represents the location of Δt . There is a moderately strong correlation with ex-situ- \dot{M}_* lagging PC1 by ~ 940 Myrs. The correlation is slightly stronger than 3σ but only for a single time step, which casts doubts upon the relative importance.

CHAPTER 4. VELA SIMULATIONS

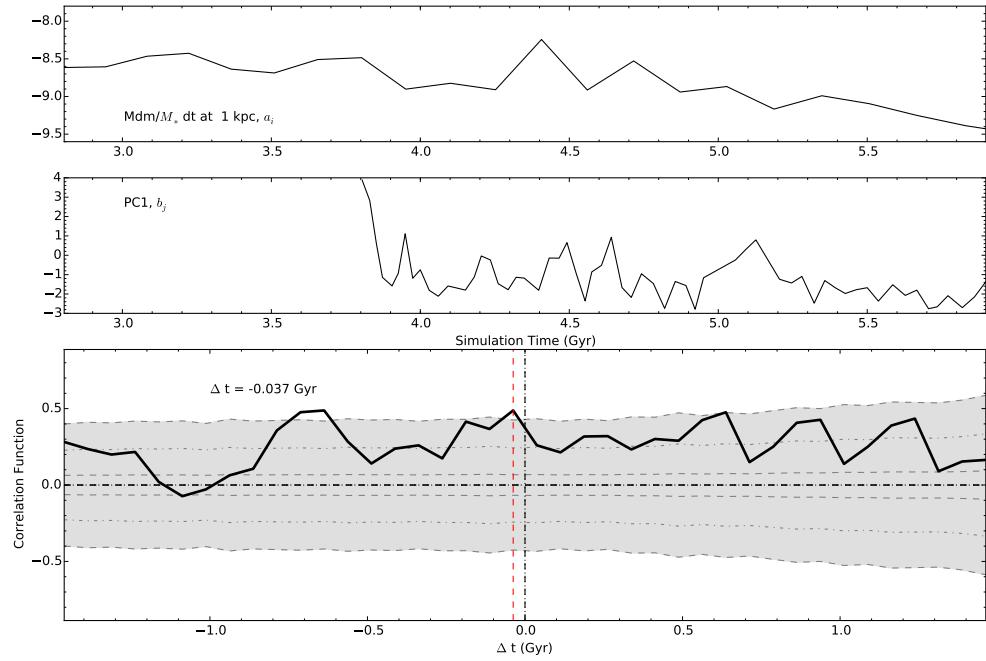


Figure 4.22 (*top panel*): Rate of dark matter mass into the central 1 kpc of VELA02 (as seen face-on) as a function of time. (*middle panel*): PC1 of VELA02 (as seen face-on) as a function of time. (*bottom panel*): The cross-correlation between \dot{M}_{dm} and PC1 (bulge strength). The gray regions represent the 3σ level of significance, correlations above and below this region are important. The vertical red dashed line represents the location of Δt . The correlation between dark matter mass rate and PC1 does not appear to be very significant for VELA02.

CHAPTER 4. VELA SIMULATIONS

questions we first calculate the cross-correlations present in each VELA galaxy. By themselves these cross-correlations for each VELA galaxy can be difficult to interpret. To understand the average cross-correlation for VELA galaxies we stack the cross-correlations between PC1 and ex-situ- M_* , ex-situ- \dot{M}_* , SFR, f_{gas} , \dot{M}_{dm} , and \dot{f}_{gas} for 9/10 VELA galaxies. We did not include the results of VELA14 because this galaxy was not measured at as many time steps as the other galaxies. We were not able to sample the full range of time lags/leads as the remaining galaxies. We stack the cross-correlation results for these galaxies to get an idea of the general correlations between morphology and physical parameters in the simulation. This analysis can easily be scaled up when more VELA galaxies have been CANDELized and a more representative sample of galaxies has been created.

Figure 4.23 show the cross-correlations between PC1 and physical parameters at the inner kpc. There is strong correlation between many physical parameters and PC1. At 1 kpc there is a strong correlation for PC1 with sSFR, f_{gas} and ex-situ- M_* . Each of these correlations are stronger than 3σ for between 100 Myrs and 1Gyr. PC1 has strong but wide correlations with \dot{M}_{dm} and ex-situ- \dot{M}_* . Figure 4.24 shows that at the 10 kpc scale, PC1 is most strongly correlated with \dot{f}_{gas} and ex-situ- M_* . The remaining physical parameters have very wide distributions of time delay and thus are not important by our definition.

Figures 4.25 and 4.26 show that PC2 is anti correlated with all measured physical parameters. However, each correlation is stronger than 3σ for > 1 Gyr, which we

CHAPTER 4. VELA SIMULATIONS

have defined as not important. PC2 is not very well correlated with any physical parameter.

In Figure 4.27 we see that only the moderate correlations between PC3 and f_{gas} , ex-situ mass or sSFR for the inner kpc. The time lag between PC3 and sSFR or f_{gas} is only ~ 40 Myrs, which we can say all occur simultaneously. The galaxy experiences a morphological change slightly before f_{gas} migrates to the central kpc. The migration of M_{gas} must be a slow process. While the gas slowly cycles inward the shape and structure of the galaxy has changed before the gas settles in the central kpc. Ex-situ mass leads PC3 by ~ 500 Myrs, followed by a moderate anti-correlation with ex-situ mass lagging PC3 by ~ 100 Myrs. This could be an example of a merger event leading to a change in PC3 by disturbing the galaxy, followed by either a decrease of ex-situ stellar mass (through the first pass of a merger) or the creation of more stellar mass in the galaxy pushing the ex-situ stellar mass/ M_* ratio lower.

Figure 4.28 shows that at the 10 kpc scale, only PC3 and sSFR have a strong correlation. There is roughly no time delay between the two time series indicating that asymmetric morphologies and sSFR are deeply intertwined. The stacked correlations for \dot{M}_{dm} , ex-situ- M_* , ex-situ- \dot{M}_* , \dot{f}_{gas} ex-situ- M_* , \dot{M}_{dm} are not as statistically significant since none of their cross-correlation coefficients are larger than the 3σ threshold.

CHAPTER 4. VELA SIMULATIONS

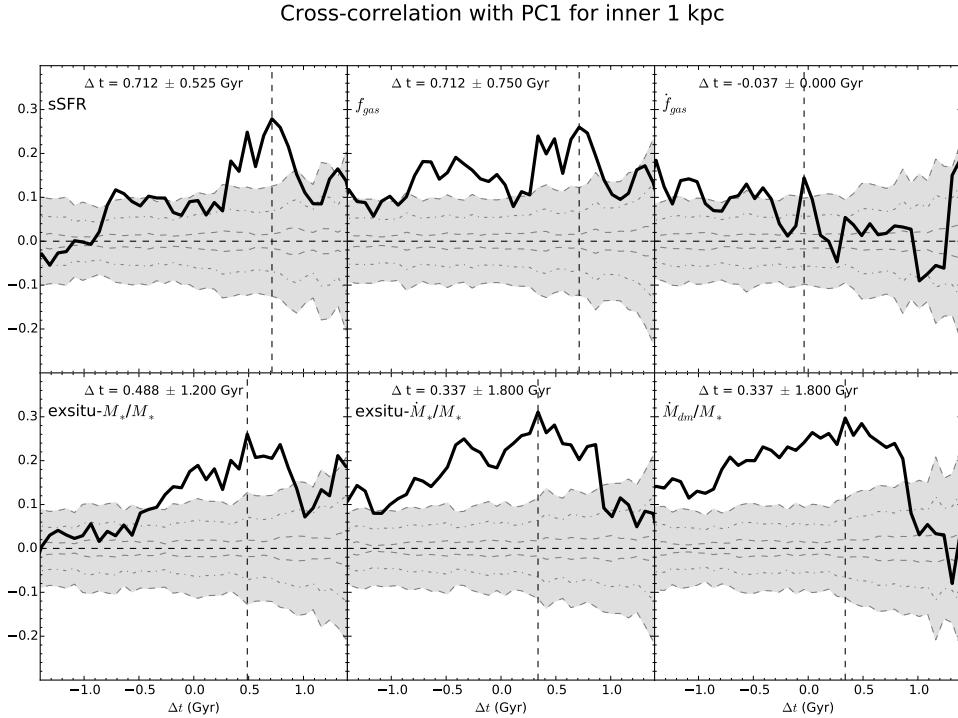


Figure 4.23 The stacked cross-correlation between PC1 and ex-situ M_*/M_* , f_{gas} , \dot{M}_{dm}/M_* , ex-situ \dot{M}_*/M_* , sSFR, and \dot{f}_{gas} and for 9/10 VELA galaxies as observed at 1 kpc. PC1 is most strongly correlated with sSFR, f_{gas} and ex-situ M_*/M_* . Each physical property leads PC1 by 500-700 Myrs implying the the bulge strengthens in a compaction phase before the galaxy quenches or that disk growth is a lengthy process before star-formation and gas accretion are maximized.

CHAPTER 4. VELA SIMULATIONS

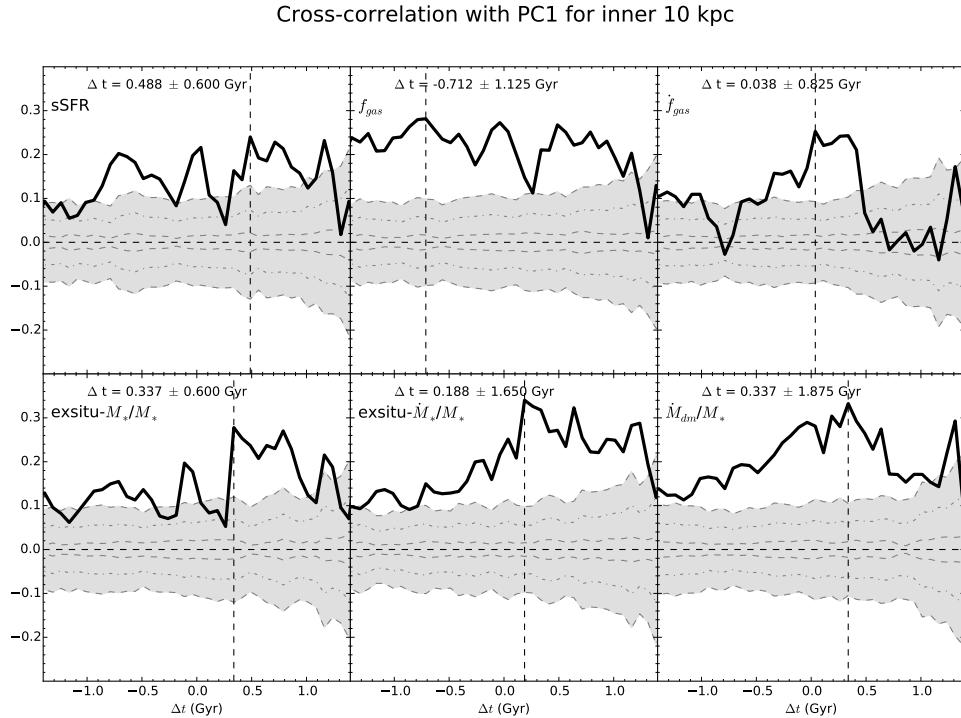


Figure 4.24 The stacked cross-correlation between PC1 and ex-situ- M_*/M_* , f_{gas} , \dot{M}_{dm}/M_* , ex-situ- \dot{M}_*/M_* , sSFR, and \dot{f}_{gas} and for 9/10 VELA galaxies as observed at 10 kpc. PC1 is most strongly correlated with sSFR, \dot{f}_{gas} and ex-situ M_*/M_* . PC1 and \dot{f}_{gas} very nearly instantly influenced by one another and SFR is influenced nearly 500 Myrs later which shows that star-formation responds strongly to a build up of fuel and a disk.

CHAPTER 4. VELA SIMULATIONS

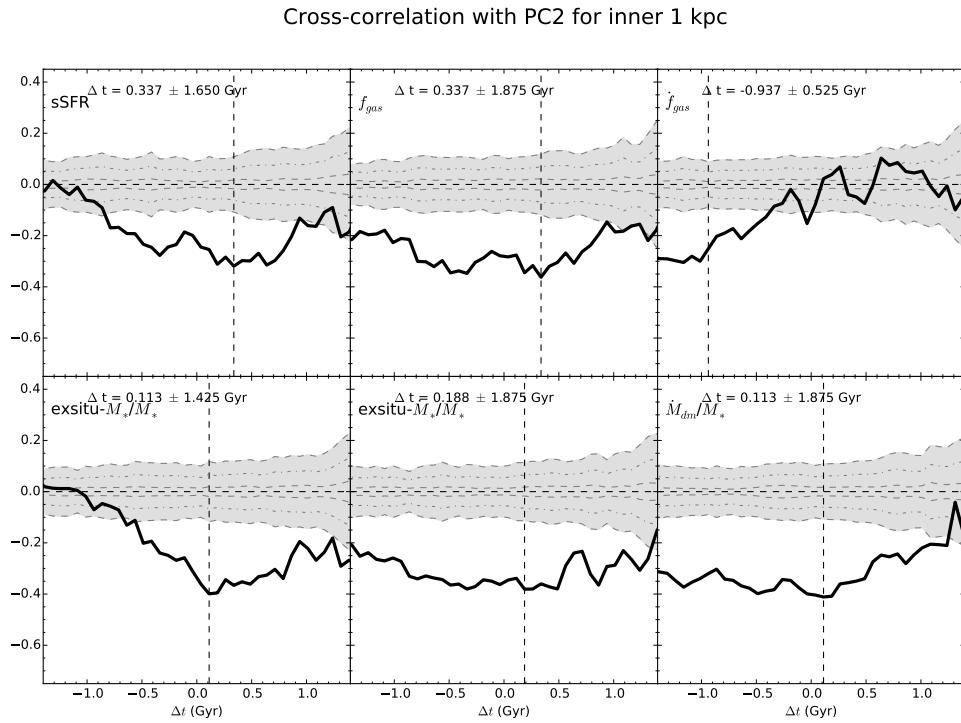


Figure 4.25 The stacked cross-correlation between PC2 and ex-situ- M_*/M_* , f_{gas} , \dot{M}_{dm}/M_* , ex-situ- \dot{M}_*/M_* , sSFR, and \dot{f}_{gas} and for 9/10 VELA galaxies as observed at 1 kpc. PC2 displays only weak and very wide anti-correlations with physical properties.

CHAPTER 4. VELA SIMULATIONS

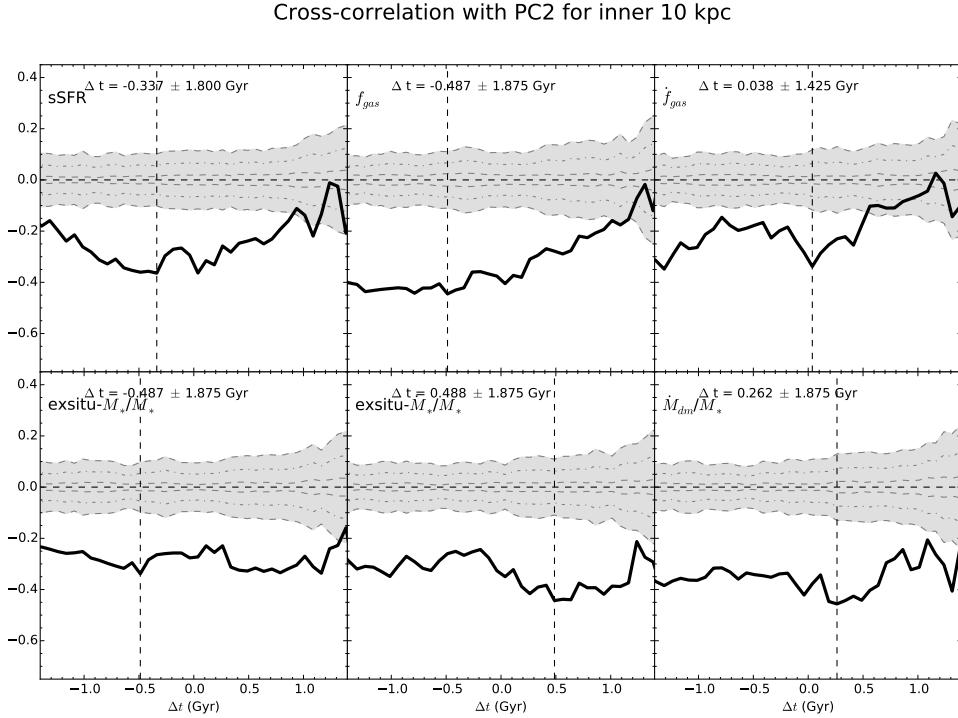


Figure 4.26 The stacked cross-correlation between PC2 and ex-situ- M_*/M_* , f_{gas} , \dot{M}_{dm}/M_* , ex-situ- \dot{M}_*/M_* , sSFR, and \dot{f}_{gas} and for 9/10 VELA galaxies as observed at 10 kpc. PC2 displays only weak and very wide anti-correlations with physical properties.

CHAPTER 4. VELA SIMULATIONS

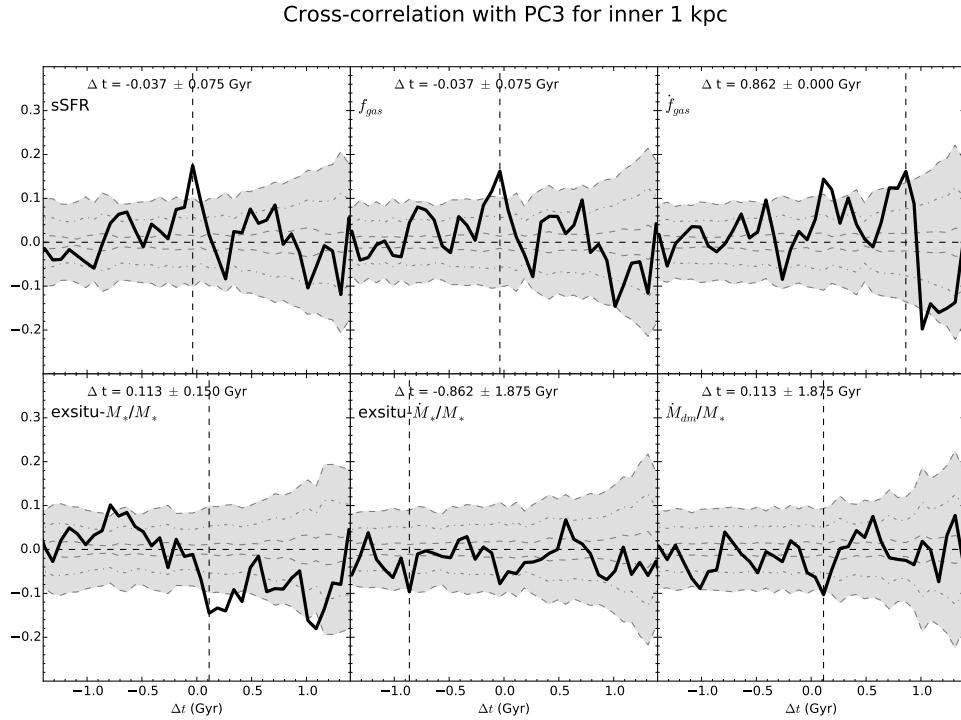


Figure 4.27 The stacked cross-correlation between PC3 and ex-situ- M_*/M_* , f_{gas} , \dot{M}_{dm}/M_* , ex-situ- \dot{M}_*/M_* , sSFR, and \dot{f}_{gas} and for 9/10 VELA galaxies as observed at 1 kpc. sSFR and f_{gas} have a very moderately positive cross-correlation with PC3 and no time lag.

CHAPTER 4. VELA SIMULATIONS

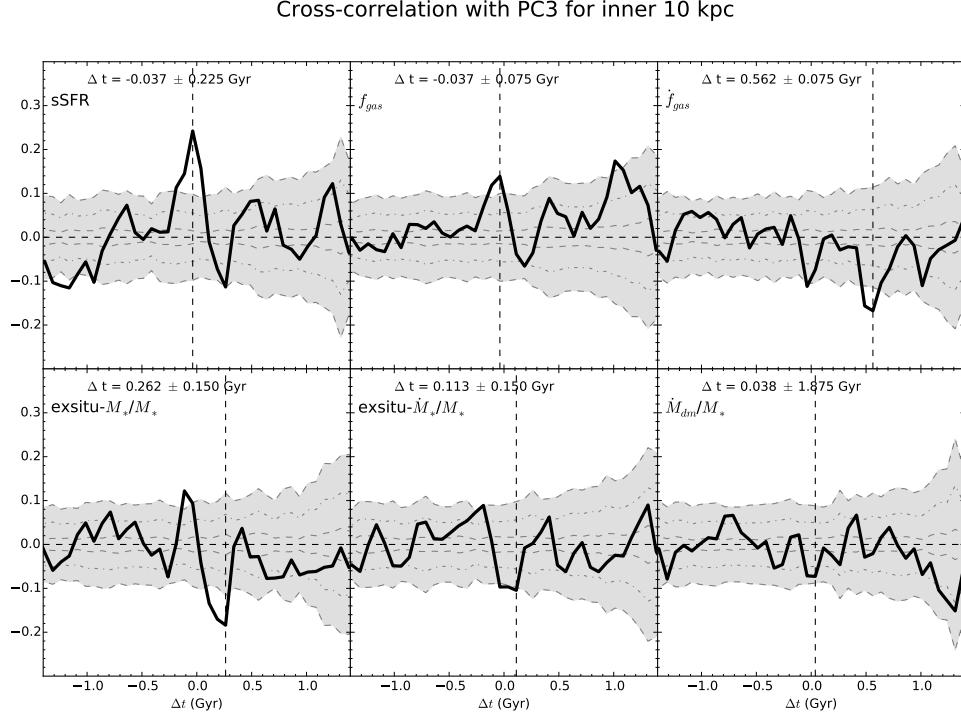


Figure 4.28 The stacked cross-correlation between PC3 and ex-situ- M_*/M_* , f_{gas} , \dot{M}_{dm}/M_* , ex-situ- \dot{M}_*/M_* , sSFR, and \dot{f}_{gas} and for 9/10 VELA galaxies as observed at 10 kpc. sSFR has a strong positive cross-correlation with PC3 and no time lag. Star-formation and asymmetric morphology influence each other nearly instantly.

4.5 Discussion

One of the most important attributes of the VELA simulations is the ability to track individual galaxies across cosmic time. This allows for direct observations of the mergers, disk instabilities and other physical mechanisms influencing galaxy morphology. The delay time between mechanisms leading to morphological changes or vice versa can be measured through a cross-correlation of the morphological (PC values) and physical properties (such as SFR and gas fraction) time series. The PC values for VELA galaxies are based on the PC weights defined by the $z \sim 1.5$ CANDELS sample. These PC weights are used so that direct comparison between CANDELS and VELA galaxies is possible.

However, there are some dissimilarities between the observed CANDELS sample and the simulated VELA sample. Each sample targets a slightly different mass range. VELA galaxies rarely grow above ($10^{10} M_{\odot}$) whereas CANDELS observed galaxies are explicitly selected to be more massive than $10^{10} M_{\odot}$. There are $\sim 1,200$ galaxies in the CANDELS sample compared to only 9 in the VELA suite. The 9 VELA galaxies are each measured for at least 40 time steps and 6 viewing angles, leading to a catalog of 2,400 galaxies. The VELA simulations are only probing a small subsample of galactic masses but they do offer greater resolution and insight into how individual evolve over time.

We cross-correlated the time series of PC1 (bulge strength), PC2 (concentration) and PC3 (asymmetry) with the time series of physical parameters (sSFR, f_{gas} ,

CHAPTER 4. VELA SIMULATIONS

\dot{M}_{dm}/M_* , ex-situ- M_*/M_* and ex-situ- \dot{M}_*/M_*) to understand the cause and effect between morphology, star formation and mergers as a function of time. This analysis allows us to study how sudden or delayed the galaxy morphology and physical processes react to one another.

4.5.1 Is PC1 an indicator of evolution?

PC1 correlates most strongly with sSFR, f_{gas} , and ex-situ- M_*/M_* when measured for the inner kpc. We observe that sSFR lags PC1 by ~ 700 Myrs, f_{gas} lags by ~ 700 and ex-situ- M_* lags PC1 by ~ 500 Myrs. At 10 kpc, PC1 correlates most strongly with ex-situ- M_*/M_* , sSFR and \dot{f}_{gas} . We observe that sSFR lags PC1 by ~ 500 Myrs, ex-situ- M_* lags PC1 by ~ 350 Myrs and f_{gas} has no time lag from PC1. The time lag between sSFR and ex-situ M_* with PC1 are shorter (by ~ 150 Myrs) at the 10 kpc scale than the inner kpc. Whereas, PC1 is correlated the gas fraction rate is correlated with at 10 kpc instead of the gas fraction.

As PC1 decreases (bulge strength increases) so does the sSFR and gas fraction. The sSFR and f_{gas} decrease ~ 700 Myrs after a galaxy has experienced bulge growth. Conversely, correlations also suggest that if a galaxy becomes more disk dominated (PC1 increases) then sSFR and f_{gas} increase.

As f_{gas} begins to decline, the galaxy becomes more bulge-dominated. Quickly after (or perhaps simultaneously) a galaxy has become bulge-dominated \dot{f}_{gas} declines steeply.

CHAPTER 4. VELA SIMULATIONS

A decline in the gas supply, either through decreased accretion of cold streams or the continual conversion of gas into stars, eventually leads to a rapid decline in star-formation; creating a compact, quenched red nugget galaxy (Dekel & Burkert, 2014). These compact, quenched galaxies have been observed in numerous studies of massive, $z \sim 2-3$ galaxies (van Dokkum et al., 2008; Whitaker et al., 2012) and gradually grow to become the massive ellipticals observed in the local universe (Dekel & Burkert, 2014; Tacchella et al., 2015). Morphological quenching is the result of a sufficiently massive central bulge stabilizing the disk from further fragmentation and will shut down star formation (Martig et al., 2009; Genzel et al., 2014; Tacchella et al., 2015).

In the VELA simulations, star-formation in the central kpc continues to increase until the galactic bulge reaches an asymptotic mass and quenching begins (Zolotov et al., 2015). Less massive galaxies compact later and to a lesser degree than higher mass galaxies (a process known as “down-sizing”). Although, in the PC1 stacked cross-correlation PC1 leads the sSFR by ~ 700 Myrs for both the inner kpc and outer regions of the galaxy. This suggests the star-formation in the central bulge and outer regions react concurrently with evolving bulge strength.

The rapid increase of ex-situ stellar mass is indicative of a merger event (Zolotov et al., 2015). At both the inner kpc and 10 kpc length scales, ex-situ- \dot{M}_*/M_* and PC1 correlate strongly (PC1 leads by ~ 300 Myrs, which is ~ 200 Myrs before ex-situ mass and PC1 correlate). PC1 decreases (strengthening bulge) and is followed by ex-situ-

CHAPTER 4. VELA SIMULATIONS

\dot{M}_*/M_* decreasing most strongly. Once ex-situ- \dot{M}_*/M_* has decreased for ~ 200 Myrs then the ex-situ stellar mass reaches a minimum. Conversely, cross-correlations also state increases of PC1 (weaker bulge and stronger disk) and a merger event increases the sSFR and f_{gas} nearly 100 Myrs later.

The relationship between mergers and compaction is not always uniform. For some galaxies, a merger triggers the compaction phase, but for other galaxies a merger is not required to begin compaction. This potentially explains why the cross-correlation between PC1 and ex-situ- M_* has such a wide distribution. Non-linear perturbations in the disk (also known as violent disk instabilities or simply secular processes) are consistently observed in simulations (Zolotov et al., 2015) and may be caused by inflowing gas or merger events. Both of these causes lead to galaxy compaction and quenched star-formation.

As the gas fraction for the entire galaxy increases the strength of the disk increases. An increase in for the entire galaxy \dot{f}_{gas} precedes sSFR for the inner kpc. Once the galaxy has built up a supply of gas then sSFR can be triggered. On the other hand, if f_{gas} declines, then the sharpest decline in gas fraction coincides with the growth of the bulge. Decreases in the amount and rate of ex-situ stellar mass and gas fraction precede star-formation quenching.

4.5.2 How PC2 interacts with galaxy evolution

PC2 (strongly influenced by concentration) anti-correlates weakly with all physical measurements. The correlation distributions are very wide, indicating a large uncertainty in the time lag/lead measurement. This is true both for the inner kpc and 10 kpc length scales.

In VELA simulations, the compaction peak is correlated with the peak of star-formation (Zolotov et al., 2015). An inexact correlation between the minimum effective radius and maximal core gas density exists. In some cases the minimum effective radius is reached before the maximum core gas density and other times afterwards.

An anti-correlation between PC2 and \dot{M}_{dm}/M_* or ex-situ- M_* implies that as more dark matter or ex-situ stellar mass is accreted by a galaxy the weaker the concentration becomes. Galaxies are not becoming more concentrated with the inclusion of more dark matter and ex-situ stellar mass. Both of these parameters are indicative of merger phenomena. Implying compaction (which would increase the concentration of a galaxy) can not be the result of merger activity alone.

Qualitatively, the anti-correlations between PC2 and each physical parameter agrees with the scenario established from the correlations between PC1 and the physical parameters. When PC2 increases it means the concentration increases, but when PC1 increases the bulge strength decreases. Bulge strength and concentration are understandably related and so we see the same parameters which cause a decrease/increase in bulge strength also leading to a decrease/increase in concentration.

CHAPTER 4. VELA SIMULATIONS

However, since the distribution of delay times are very wide we can not say much more about the causative relation between PC2 morphology (concentration) and physical features of a galaxy.

4.5.3 PC3 and Star-Formation

PC3, which is highly dependent upon the asymmetry of a galaxy, correlates with sSFR and f_{gas} with nearly no time lag (~ 40 Myr) between them. PC3 correlates even more strongly with sSFR when measured for the entire galaxy (10 kpc), again with nearly no time lag. The correlation between PC3 and f_{gas} shows that disturbed morphology is directly related to the amount of gas. These correlations are well understood (e.g. Wuyts et al., 2011), galaxies undergoing strong star formation are typically more disk dominated and asymmetrical than quiescent galaxies. Likewise, galaxies can not form stars without a reservoir of gas from which to draw upon. The nearly simultaneous correlations between sSFR, gas fraction and PC3 are a good indicator that the simulations are behaving like we think they should.

Previous studies (Zolotov et al., 2015; Tomassetti et al., 2016) show the central kpc of a VELA galaxy experiences gradual mass growth and star-formation followed by a phase of gas-rich compaction, leading to an increasing gas mass and star-formation rate. Once the star-formation rate peaks a strong outflow rate reduces the amount of available gas and quenches star-formation in the central kpc (Zolotov et al., 2015). However, these works did not specifically look at the connection between morphology

CHAPTER 4. VELA SIMULATIONS

and the star-formation evolution. Our results show that morphology (particularly asymmetry) is strongly related to the gas mass and sSFR.

For stars to have formed in the bulge there must be a large reservoir of available gas. Starbursts in the bulge would be evidence of a “blue-nugget” phase of galaxy evolution (Dekel & Burkert, 2014) and agree with observational evidence of compact, SF galaxies (Barro et al., 2013, 2014a; Bruce et al., 2014a,b; Williams et al., 2014).

High redshift galaxies experience mergers and gas accretion which can ignite violent disk instabilities (VDIs, Kereš et al., 2005). The high central stellar mass density of massive galaxies at $z \sim 2.2$ suggests mergers and VDIs are important mechanisms for forming central bulges (Tacchella et al., 2015).

In the inner kpc, the correlations between \dot{f}_{gas} , \dot{M}_{dm}/M_* , ex-situ- \dot{M}_*/M_* and PC3 are quite weak. These correlations do not surpass the 3σ confidence level. On the other hand, the correlation between PC3 and ex-situ- M_*/M_* appears to be double peaked, a positive correlation occurring with PC3 lagging ex-situ- M_*/M_* by ~ 700 Myrs and an anti-correlation occurring when PC3 leads ex-situ- M_*/M_* by ~ 110 Myrs. The anti-correlation is important beyond the 3σ confidence level, suggesting

We could be witnessing the first pass of a merger accumulating ex-situ stellar mass which leads to a disruption in the morphology of the central galaxy (PC3 increase lagging the ex-situ stellar mass increase). Either, 1) the merger creates new stars (increases galaxy mass) and thus decreases the ex-situ- M_*/M_* ratio or 2) the ex-situ- M_*/M_* decrease is a result of the merging galaxy passing through the central galaxy.

CHAPTER 4. VELA SIMULATIONS

Both of these scenarios are plausible, and likely occur in tandem. This correlation is also visible at the 10 kpc scale. Except at this scale, the initial merger event PC3 only lags the ex-situ- M_*/M_* by \sim 40 Myrs (nearly co-temporal), but the anti-correlation still exists with PC3 leading the ex-situ stellar mass by \sim 250 Myrs. Intriguingly, the correlation between ex-situ- \dot{M}_*/M_* resembles a time shifted correlation of ex-situ- M_*/M_* , just not as statistically significant. An increasing or decreasing ex-situ- \dot{M}_*/M_* can directly correspond to the amount of ex-situ mass in a galaxy (i.e. faster mergers bring in more stars).

4.6 Summary

We utilized a principal component analysis of the non-parametric morphological statistics measured in Snyder et al. (2015b) of the VELA hydrodynamical simulation galaxies. These galaxies have corresponding physical data, such as gas mass, stellar mass, star-formation rate and dark matter mass measured at (nearly) equal intervals that we then cross-correlate with the PC results. We stack the results of a discrete cross correlation between the times series of PCs and physical parameters from 9 VELA galaxies. We discover each of the first three PCs correlates differently with these physical parameters: PC1 (bulge strength) is correlated strongly with ex-situ stellar mass, the gas fraction and sSFR; PC2 (concentration) is very weakly anti-correlated with all physical features; PC3 (asymmetry) is strongly correlated with

CHAPTER 4. VELA SIMULATIONS

sSFR at all length scales and with gas fraction in the central kpc.

We use the PC results to classify every time step and every viewing angle into the PC groups defined for the $1.4 < z < 2$ CANDELS galaxies sample (Peth et al., 2016). A vast majority of galaxies are classified into groups 0 and 6 (the bulge-dominated, spheroidal groups); next most into groups 4 and 8 (the bulge+disk systems), and only a small handful of galaxies are classified into groups 1, 2 or 5 (the disk-dominated groups). Visual inspection of the VELA galaxies agrees with the lack of disk-dominated galaxies.

Morphological evolution in the VELA simulations is complicated. A major merger can cause a previously disk-dominated galaxy to settle into a bulge-dominated spheroidal galaxy, by first traversing through a disturbed state. Conversely, a major merger can cause a bulge-dominated galaxy to become more disturbed, while simultaneously accreting a faint disk. However, these disks are not always long lived and typically these galaxies will settle into a bulge-dominated spheroid. Minor mergers have a comparably minor effect upon the morphology of the galaxy. Any change appears to be short lived and moderate. Typically, a minor merger will only cause bulge-dominated galaxies to fluctuate between no disk and a very weak disk (groups 6 and 0). In cases where a major merger has not occurred for $\gtrsim 1$ Gyr secular processes appear able to grow a disk around a previously bulge-dominated system. These cases are rare and only observable in VELA27.

Increasing PC1 (disk growth/bulge weakening) is correlated with sSFR, gas inflow

CHAPTER 4. VELA SIMULATIONS

and ex-situ stellar mass on very long timescales ($\Delta t \sim 700$ Myrs). Decreasing PC1 is correlated with decreasing gas fractions. Star-formation is possible when the disk is growing, but when the bulge grows and the gas fraction decreases, so does the star-formation rate.

PC3 (asymmetry) correlates well with sSFR, ex-situ mass accretion and gas fraction on very short time scales ($\Delta t \sim 0$ Myrs). As galaxies become more disk-dominated star-formation is triggered, while a build-up of the bulge will shut down star-formation. Since mergers are rare in VELA simulations, more galaxies with better time-sampling will be required to further understand the role of mergers on morphology and star-formation.

Chapter 5

Summary and Future Directions

Exactly how galaxies evolve from compact and possibly irregular systems at high redshift to the giant elliptical or spiral galaxies in the local Universe remains a very active area of research. The relationships between morphology and star-formation (Wuyts et al., 2011) or mergers (Toomre, 1977) are very well established. We can leverage this information at a variety of cosmic epochs to determine how galaxy evolution is affected by a host of physical mechanisms. Major mergers, minor mergers, gas accretion and secular processes have all been invoked to explain the morphological transformations, cosmic SFR decline, size evolution and bulge growth observed since $z \sim 2$. Each mechanism affects the morphology, size evolution and bulge growth in different ways. This thesis used machine learning to quantify morphology to further understand aspects of galaxy evolution.

5.1 Galaxy Morphological Classifications Using PCA

To investigate the role of morphology in communicating galaxy evolution we need to classify galaxies and identify the possible galaxy quenching mechanisms. However, important (albeit rare and subtle) processes driving galaxy morphology and star-formation may be missed by traditional spiral, elliptical, irregular or Sérsic bulge/disk classifications. To overcome this limitation, we use a principal component analysis of non-parametric morphological indicators (concentration, asymmetry, Gini coefficient, M_{20} , multi-mode, intensity and deviation) measured at rest-frame B -band (corresponding to HST/WFC3 F125W at $1.4 < z < 2$) to trace the natural distribution of massive ($> 10^{10} M_\odot$) galaxy morphologies.

Principal component analysis (PCA) quantified the correlations between these morphological indicators and determines the relative importance of each. The first three principal components (PCs) capture ~ 75 per cent of the variance inherent to our sample. We interpret the first principal component (PC) as bulge strength, the second PC as dominated by concentration and the third PC as dominated by asymmetry. Both PC1 and PC2 correlate with the visual appearance of a central bulge and predict galaxy quiescence. PC1 is a better predictor of quenching than stellar mass, and as good as other structural indicators (Sérsic-n or compactness).

5.2 Random Forest Classifications of Pan-STARRS Galaxies

Galaxies have long been visually classified and in recent years large scale collaborations such as GalaxyZoo and CANDELS (Lintott et al., 2008a; Kartaltepe et al., 2015) have undertaken the problem of cataloging and annotating galaxies. Researchers are tasked with deciding between disk, elliptical, irregular, merger remnant and many other classes of galaxies.

Merger identification plays a very important in understanding their role in the formation and evolution of galaxies, in terms of structural assembly, star-formation and nuclear activity. The frequency and characteristics of merging galaxies lead to an understanding of the physical processes that influence galaxies. Merging galaxies display unique morphological characteristics (such as tidal tails or double nuclei) from disks or ellipticals, but the exact nature of these characteristics is unique to each merger. There are many different types and stages of merger events which make classification challenging.

We develop a classification scheme utilizing a supervised machine learning technique, random forest, to classify galaxies into mergers and non-mergers. This method is trained upon the quantitative morphological measurements using photometry from Pan-STARRS coinciding with visually classified Galaxy Zoo galaxies.

With 93% completeness and 93% accuracy random forest is able to distinguish

CHAPTER 5. SUMMARY

mergers from non-merger galaxies using a variety of input features (PCs, non-parametric morphologies, sSFR, M_* , rest-frame color). The training sample includes galaxies that have already merged, and those in interacting pairs with visible tidal disruptions that will one day merge. Asymmetry is by far the most important indicator of whether a galaxy is experiencing a merger and the next most important features include: PC7, PC5, PC3, deviation and $d(G, M_{20})$.

We are able to extend the random forest classifications to a sample of local ($z < 0.05$) MaNGA galaxies that has been trained using the Galaxy Zoo/Pan-STARRS sample. These classifications are 100% complete and 78% specificity but a large number of false positives in relation to true positives (142 vs. 5). An investigation into the false positive galaxies with the largest random forest probability of being a merger determines that 14 have bright foreground stars contaminating the morphological measurements, 13 have either irregular or disturbed morphology and 3 have strong tidal features. The classes of disturbed galaxies with or without tidal features could provide an additional source of merging galaxies only visible with the deeper imaging of Pan-STARRS.

5.3 Studying Galaxy Morphology Using VELA Simulation Suite

The morphology of a galaxy indicates a myriad of process, including major/minor mergers, tidal stripping and violent disk instabilities form and assemble the system. Understanding exactly how the morphology is influenced by these mechanisms is difficult in observational studies that are only able to capture a single snapshot of a galaxy. The causality inherent in the relationships between structural evolution and properties such as star-formation rate, colors or mass can not be determined directly.

Galaxy simulations provide an avenue to study how individual galaxies evolve over time and which mechanisms exactly are responsible for morphological changes. The VELA simulation suite (Ceverino et al., 2010a; Ceverino & Klypin, 2009; Ceverino et al., 2012; Dekel et al., 2013; Ceverino et al., 2014) contains 9 galaxies with physical measurements (gas fraction, ex-situ stellar mass, star-formation rate, dark matter mass) and “CANDELized” morphological measurements (Snyder et al., 2015b). The relationships and causality between morphology and physical mechanisms can be shown through an discrete correlation function between the morphology and physical property time series.

We investigated the temporal relationship between morphology and these physical parameters to discover that each of the first three PCs correlates differently with these physical parameters: PC1 is strongly correlated with ex-situ stellar mass, the

CHAPTER 5. SUMMARY

gas fraction, sSFR for the inner kpc and \dot{f}_{gas} at the 10 kpc length scale; PC2 is weakly anti-correlated with all physical properties; PC3 is strongly correlated with sSFR at all length scales and with gas fraction in the central kpc. The strong correlations between PC1 and physical parameters show the strength of the bulge is directly related to how much star formation and gas is present. Additionally, disk growth is a long, slow process eventually influencing star-formation. Meanwhile, correlations between PC3, sSFR and gas show that galaxies become the most disturbed at the same time that star-formation is peaking.

5.4 Mergers Can Grow Bulges and Regulate Star-formation

Massive structures grow hierarchically in a Λ CDM universe (White & Rees, 1978) and the most massive dark matter halos (and galaxies) experience more mergers than less massive halos. Mergers can cause large disturbances in galaxy morphology, restructure galaxies, assemble classical bulges and regulate star-formation. However, measurements of galaxy mergers have not led to a complete picture of how important they are at different times (Lin et al., 2004; Lotz et al., 2008; Conselice et al., 2009; Lotz et al., 2011) because of differences in how mergers are defined (either in the pre-merger stage of pairs or post-merger stage of a disturbed galaxy).

Elliptical galaxies and central bulges can be the result of disk galaxy mergers

CHAPTER 5. SUMMARY

(Lynden-Bell, 1967; Toomre, 1977), particularly dissipational mergers (Barnes, 1988; Hernquist, 1992). Even though gas-poor mergers are more common than gas-rich mergers (Lotz et al., 2011) they alone can not account for the size growth of ellipticals (Brooks & Christensen, 2016 and references therein). Semi-analytic models (SAMs; Porter et al., 2014) and cosmological simulations (Oser et al., 2012) have shown that a combination of all different flavors of mergers (gas-rich, gas-poor, major, minor) is required to match the scaling relations of observed galaxies (Faber & Jackson, 1976; Kormendy, 1977a).

This thesis investigated different (but compatible) aspects of galaxy evolution and assembly across cosmic time through the use of the (primarily) compact high redshift ($1.36 < z < 1.97$) galaxies, local ($z < 0.1$) galaxy mergers, and a suite of simulated galaxies bridging part of the difference in time. By studying galaxy morphology across cosmic time we can probe the physical mechanisms (primarily mergers) which lead to bulge growth and regulate star-formation.

There is no single diagnostic defining mergers across all cosmic time but a consensus on merger identification can be reached through multiple diagnostic statistics. Quantitative morphological statistics (Gini, M_{20} , concentration, asymmetry, PCs, etc.) have shown an ability to identify major and minor mergers with varying success. As shown in Chapter 2, Gini– M_{20} and C–A show a distinct ability to create pure but incomplete samples of merging galaxies. For local galaxies, PC7, and to a lesser extent PC5, present an interesting consequence of the random forest classification;

CHAPTER 5. SUMMARY

principal components that may not capture much of the variance are found to be quite important for distinguishing mergers from non-mergers in local galaxies. With enhanced resolution photometry the analysis of identifying mergers in local galaxies can be extended to higher redshift.

Clearly galaxies evolved significantly from the compact high-redshift galaxies seen in CANDELS to the large spiral and elliptical galaxies seen in Pan-STARRS at low redshift. High resolution galaxy simulations (such as VELA) allow for careful study of the relationship between mergers, morphology and galactic assembly.

The relationship between mergers and morphology can be quite complicated. For bulge-dominated galaxies a merger can help regrow a disk, but for disk-dominated galaxies a merger can destroy the disk leading to violent relaxation and a spheroidal merger remnant. The build up of central bulges in the VELA simulations \sim 500 Myrs before subsequent quenching of star-formation shows that the presence of a bulge plays a large factor into the ability for a galaxy to form stars. The build up of a bulge is not a straightforward process, there are many competing factors such as mergers and gas outflows that seek to both build and not build a bulge. While disk growth is equally related to increasing star-formation, on equivalently long time scales. Star-formation does not reach a maximum until \sim 500 Myrs after a disk is fully grown.

Quantitative morphology has shown the capacity to identify mergers in an effort to characterize the evolution of bulge growth, star-formation regulation, and galaxy

CHAPTER 5. SUMMARY

assembly. Future studies with larger surveys of observed galaxies, greater number of simulated galaxies and deeper/higher resolution imaging will be able to further the understanding of how morphology and galaxy assembly are deeply intertwined.

5.5 Future Work

Star-formation can be quenched in many ways and with a reliable morphology classification for different epochs we can begin to answer some intriguing open questions: whether star-formation quenching is occurring at the same time as the bulge is forming? A temporal connection between these two could have important consequences on how galaxies have been quenching star-formation.

The PCA and random forest classifications defined here can be used to study the evolution of star-formation for a variety of morphological types as a function of time, star-formation rates and mass. The prevalence of certain morphological types can be indicators for the importance of mergers and secular processes for assembling galaxy structure.

We can extend the random forest classifications to higher redshift either through collecting more high redshift user classified galaxies, or by artificially redshifting our sample of low redshift mergers/non-mergers. The CANDELS and Galaxy Zoo teams are continuing the arduous process of collecting user determined morphological classifications, extending to higher redshifts. However, the primary drawback with this

CHAPTER 5. SUMMARY

method is the true hallmarks of merger activity (tidal tails, faint disturbances in the disk) are very hard to impossible to observe given the resolution of high redshift galaxies. Eventually, new space telescopes such as WFIRST and very high resolution telescopes such as HDST will provide a tool to observe high redshift galaxies for larger samples or much greater resolution, but these programs are still years to decades away from completion.

In the upcoming years and decades large telescope surveys such as LSST, Pan-STARRS and E-ELT will come online providing the astronomical community with petabytes of raw images. With this much data, problems such as the visual classifications of galaxies will quickly become intractable without the use automated machine-learning methods.

We can also use simulations, such as VELA, as a training set for random forest classifications. The VELA simulations are powerful because not only are they available at the resolution of high redshift HST but also the images could be adapted to match the resolution of lower redshift observations or the higher resolution of future missions.

The number of simulations available in VELA will begin to grow as newer, faster algorithms and computers will make simulating more galaxies possible. Already there are nearly 20 galaxies that have been simulated in VELA but have not been “CANDELized” which could nearly triple the total amount of simulated data. The process of CANDELizing these galaxies is currently underway (Snyder et al., inprep) and

CHAPTER 5. SUMMARY

should be available in the near term. An increase in the number of galaxies will increase the resolution and interpretability of our cross-correlation importance and lag analysis.

Additional galaxy simulation suites such as Illustris (Vogelsberger et al., 2014) and EAGLE (McAlpine et al., 2015) may be a productive avenue to study the connection between morphology and physical features. These simulations have a very large number of galaxies which will greatly enhance the statistics of any analysis.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Abraham, R. G., van den Bergh, S., Glazebrook, K., et al. 1996, ApJS, 107, 1

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

Agarwal, A., Gupta, A. C., Bachev, R., et al. 2015, MNRAS, 451, 3882

Agertz, O., Teyssier, R., & Moore, B. 2011, MNRAS, 410, 1391

Andrae, R., Jahnke, K., & Melchior, P. 2011, MNRAS, 411, 385

Baldry, I. K., & Glazebrook, K. 2003, ApJ, 593, 258

Ball, N. M., & Brunner, R. J. 2010, International Journal of Modern Physics D, 19, 1049

Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2008a, ApJ, 683, 12

—. 2007, ApJ, 663, 774

BIBLIOGRAPHY

- Ball, N. M., Brunner, R. J., Myers, A. D., & Tcheng, D. 2006, ApJ, 650, 497
- Ball, N. M., Loveday, J., & Brunner, R. J. 2008b, MNRAS, 383, 907
- Ball, N. M., Loveday, J., Fukugita, M., et al. 2004, MNRAS, 348, 1038
- Barnes, J. E. 1988, ApJ, 331, 699
- Barnes, J. E., & Hernquist, L. 1996, ApJ, 471, 115
- Barro, G., Faber, S. M., Pérez-González, P. G., et al. 2013, ApJ, 765, 104
- . 2014a, ApJ, 791, 52
- Barro, G., Trump, J. R., Koo, D. C., et al. 2014b, ArXiv e-prints, arXiv:1405.7042
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
- Bell, E. F., van der Wel, A., Papovich, C., et al. 2012, ApJ, 753, 167
- Bell, E. F., Wolf, C., Meisenheimer, K., et al. 2004, ApJ, 608, 752
- Bell, E. F., Naab, T., McIntosh, D. H., et al. 2006, ApJ, 640, 241
- Ben-Hur, A., & Guyon, I. 2003, in Methods in Molecular Biology, Vol. 224, Functional Genomics, ed. M. Brownstein & A. Khodursky (Humana Press), 159–182
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
- Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645

BIBLIOGRAPHY

- Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
- Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349
- Bournaud, F., & Elmegreen, B. G. 2009, ApJ, 694, L158
- Bournaud, F., Elmegreen, B. G., & Elmegreen, D. M. 2007, ApJ, 670, 237
- Bournaud, F., Elmegreen, B. G., & Martig, M. 2009, ApJ, 707, L1
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
- Breiman, L. 2001, Machine Learning, 45, 5
- Brennan, R., Pandya, V., Somerville, R. S., et al. 2015, MNRAS, 451, 2933
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
- Brooks, A., & Christensen, C. 2016, Galactic Bulges, 418, 317
- Brooks, A. M., Governato, F., Quinn, T., Brook, C. B., & Wadsley, J. 2009, ApJ, 694, 396
- Brown, M. J. I., Moustakas, J., Smith, J.-D. T., et al. 2014, ApJS, 212, 18
- Bruce, V. A., Dunlop, J. S., Cirasuolo, M., et al. 2012, MNRAS, 427, 1666
- Bruce, V. A., Dunlop, J. S., McLure, R. J., et al. 2014a, MNRAS, 444, 1001
- . 2014b, MNRAS, 444, 1660

BIBLIOGRAPHY

- Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
- Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
- Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
- Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, ApJ, 712, 511
- Cattaneo, A., Faber, S. M., Binney, J., et al. 2009, Nature, 460, 213
- Ceverino, D., Dekel, A., & Bournaud, F. 2010a, MNRAS, 404, 2151
- . 2010b, MNRAS, 404, 2151
- Ceverino, D., Dekel, A., Mandelker, N., et al. 2012, MNRAS, 420, 3490
- Ceverino, D., & Klypin, A. 2009, ApJ, 695, 292
- Ceverino, D., Klypin, A., Klimek, E. S., et al. 2014, MNRAS, 442, 1545
- Ceverino, D., Primack, J., & Dekel, A. 2015, MNRAS, 453, 408
- Ceverino, D., Sánchez Almeida, J., Muñoz Tuñón, C., et al. 2016, MNRAS, 457, 2605
- Chabrier, G. 2003, PASP, 115, 763
- Cisternas, M., Jahnke, K., Inskip, K. J., et al. 2011, ApJ, 726, 57

BIBLIOGRAPHY

- Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
- Conselice, C. J. 2014, ARA&A, 52, 291
- Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
- Conselice, C. J., Chapman, S. C., & Windhorst, R. A. 2003, ApJ, 596, L5
- Conselice, C. J., Yang, C., & Bluck, A. F. L. 2009, MNRAS, 394, 1956
- Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839
- Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
- Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93
- Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010a, MNRAS, 401, 1043
- . 2010b, MNRAS, 401, 1552
- de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247
- Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
- Dekel, A., & Burkert, A. 2014, MNRAS, 438, 1870
- Dekel, A., Sari, R., & Ceverino, D. 2009a, ApJ, 703, 785
- Dekel, A., Zolotov, A., Tweed, D., et al. 2013, MNRAS, 435, 999
- Dekel, A., Birnboim, Y., Engel, G., et al. 2009b, Nature, 457, 451

BIBLIOGRAPHY

- Edelson, R. A., & Krolik, J. H. 1988, ApJ, 333, 646
- Elbaz, D., & Cesarsky, C. J. 2003, Science, 300, 270
- Ellison, S. L., Patton, D. R., Mendel, J. T., & Scudder, J. M. 2011, MNRAS, 418, 2043
- Elmegreen, B. G., Bournaud, F., & Elmegreen, D. M. 2008, ApJ, 688, 67
- Engel, H., Tacconi, L. J., Davies, R. I., et al. 2010, ApJ, 724, 233
- Everitt, B., & Hothorn, T. 2006, A Handbook of Statistical Analyses Using R (Chapman & Hall/CRC)
- Faber, S. M., & Jackson, R. E. 1976, ApJ, 204, 668
- Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, ApJ, 665, 265
- Fabian, A. C. 2012, ARA&A, 50, 455
- Fang, J. J., Faber, S. M., Koo, D. C., & Dekel, A. 2013, ApJ, 776, 63
- Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761
- Ferrarese, L., & Merritt, D. 2000, ApJ, 539, L9
- Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, ApJ, 706, 1364
- Freeman, P. E., Izbicki, R., Lee, A. B., et al. 2013, MNRAS, 434, 282
- Gadotti, D. A. 2009, MNRAS, 393, 1531

BIBLIOGRAPHY

- Galmetz, A., Grazian, A., Fontana, A., et al. 2013, ApJS, 206, 10
- Gaur, H., Gupta, A. C., Bachev, R., et al. 2015, MNRAS, 452, 4263
- Genzel, R., Burkert, A., Bouché, N., et al. 2008, ApJ, 687, 59
- Genzel, R., Förster Schreiber, N. M., Lang, P., et al. 2014, ApJ, 785, 75
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279
- Governato, F., Willman, B., Mayer, L., et al. 2007, MNRAS, 374, 1479
- Governato, F., Zolotov, A., Pontzen, A., et al. 2012, MNRAS, 422, 1231
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
- Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
- Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
- Haardt, F., & Madau, P. 1996, ApJ, 461, 20
- Haas, M., Chini, R., Ramolla, M., et al. 2011, A&A, 535, A73
- Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, ApJ, 613, 109
- Hernquist, L. 1992, ApJ, 400, 460

BIBLIOGRAPHY

- Hogg, D. W., Blanton, M. R., Brinchmann, J., et al. 2004, ApJ, 601, L29
- Hopkins, P. F., & Hernquist, L. 2009, ApJ, 694, 599
- Hopkins, P. F., Somerville, R. S., Cox, T. J., et al. 2009, MNRAS, 397, 802
- Hopkins, P. F., Bundy, K., Croton, D., et al. 2010, ApJ, 715, 202
- Hoyle, B., Masters, K. L., Nichol, R. C., et al. 2011, MNRAS, 415, 3627
- Hubble, E. P. 1926, ApJ, 64, doi:10.1086/143018
- Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, ApJS, 221, 8
- Hung, C.-L., Hayward, C. C., Smith, H. A., et al. 2016, ApJ, 816, 99
- Immeli, A., Samland, M., Gerhard, O., & Westera, P. 2004a, A&A, 413, 547
- Immeli, A., Samland, M., Westera, P., & Gerhard, O. 2004b, ApJ, 611, 20
- Ivezić, v., Tyson, J. A., Acosta, E., et al. 2008, arXiv:0805.2366v4
- Ivezić, Ž., Connolly, A., Vanderplas, J., & Gray, A. 2013, Statistics, Data Mining and Machine Learning in Astronomy (Princeton University Press)
- James, G., Witten, D., Hastie, T., & Tibshirani, R. 2014, An Introduction to Statistical Learning: With Applications in R (Springer Publishing Company, Incorporated)
- Jolliffe, I. T. 1986, Principal component analysis
- Jonsson, P. 2006, MNRAS, 372, 2

BIBLIOGRAPHY

- Jonsson, P., Groves, B. A., & Cox, T. J. 2010, MNRAS, 403, 17
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in Proc. SPIE, Vol. 7733, Ground-based and Airborne Telescopes III, 77330E
- Kamdar, H. M., Turk, M. J., & Brunner, R. J. 2016a, MNRAS, 455, 642
- . 2016b, MNRAS, 457, 1162
- Kampczyk, P., Lilly, S. J., Carollo, C. M., et al. 2007, ApJS, 172, 329
- Kartaltepe, J. S., Mozena, M., Kocevski, D., et al. 2015, ApJS, 221, 11
- Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, ApJ, 758, 106
- Kauffmann, G., & Haehnelt, M. 2000, MNRAS, 311, 576
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 54
- Keel, W. C., Manning, A. M., Holwerda, B. W., et al. 2013, PASP, 125, 2
- Kennicutt, Jr., R. C. 1998, ApJ, 498, 541
- Kennicutt, Jr., R. C., Roettiger, K. A., Keel, W. C., van der Hulst, J. M., & Hummel, E. 1987, AJ, 93, 1011
- Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
- Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

BIBLIOGRAPHY

- Kormendy, J. 1977a, ApJ, 218, 333
- . 1977b, ApJ, 217, 406
- Kormendy, J., & Kennicutt, Jr., R. C. 2004, ARA&A, 42, 603
- Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581
- Kravtsov, A. V. 2003, ApJ, 590, L1
- Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, ApJS, 111, 73
- Kriek, M., van Dokkum, P. G., Franx, M., Illingworth, G. D., & Magee, D. K. 2009, ApJ, 705, L71
- Kriek, M., van Dokkum, P. G., Franx, M., et al. 2006, ApJ, 649, L71
- Kroupa, P. 2001, MNRAS, 322, 231
- Labbé, I., Huang, J., Franx, M., et al. 2005, ApJ, 624, L81
- Lahav, O., Naim, A., Buta, R. J., et al. 1995, Science, 267, 859
- Lang, P., Wuyts, S., Somerville, R. S., et al. 2014, ApJ, 788, 11
- Lee, B., Giavalisco, M., Williams, C. C., et al. 2013, ApJ, 774, 47
- Lin, H., Yee, H. K. C., Carlberg, R. G., et al. 1999, ApJ, 518, 533
- Lin, L., Koo, D. C., Willmer, C. N. A., et al. 2004, ApJ, 617, L9

BIBLIOGRAPHY

- Lin, L., Jian, H.-Y., Foucaud, S., et al. 2014, ApJ, 782, 33
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
- Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008a, MNRAS, 389, 1179
- . 2008b, MNRAS, 389, 1179
- Lorenz, M. O. 1905, Publications of the American Statistical Association, Volume 9, Number 70, p. 209-219, 9, 209
- Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, ApJ, 742, 103
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, MNRAS, 404, 590
- . 2010b, MNRAS, 404, 575
- Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163
- Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177
- Lynden-Bell, D. 1967, MNRAS, 136, 101
- Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285
- Martig, M., Bournaud, F., Teyssier, R., & Dekel, A. 2009, ApJ, 707, 250
- Mather, J. C., Cheng, E. S., Eplee, Jr., R. E., et al. 1990, ApJ, 354, L37

BIBLIOGRAPHY

- McAlpine, S., Helly, J. C., Schaller, M., et al. 2015, ArXiv e-prints, arXiv:1510.01320
- Meurer, G. R., Heckman, T. M., Leitherer, C., et al. 1995, AJ, 110, 2665
- Mihos, J. C., & Hernquist, L. 1994, ApJ, 437, L47
- Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, ApJ, 808, 101
- Moody, C. E., Guo, Y., Mandelker, N., et al. 2014, MNRAS, 444, 1389
- Mozena, M. W. 2013, PhD thesis, University of California, Santa Cruz
- Naab, T., Jesseit, R., & Burkert, A. 2006a, MNRAS, 372, 839
- Naab, T., Johansson, P. H., & Ostriker, J. P. 2009, ApJ, 699, L178
- Naab, T., Khochfar, S., & Burkert, A. 2006b, ApJ, 636, L81
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43
- Noguchi, M. 1999, ApJ, 514, 77
- Oser, L., Naab, T., Ostriker, J. P., & Johansson, P. H. 2012, ApJ, 744, 63
- Papovich, C., Bassett, R., Lotz, J. M., et al. 2012, ApJ, 750, 93
- Patel, S. G., Kelson, D. D., Holden, B. P., Franx, M., & Illingworth, G. D. 2011, ApJ, 735, 53
- Pearson, K. 1901, Philosophical Magazine Series 6, 2, 559

BIBLIOGRAPHY

- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
—. 2010, AJ, 139, 2097
- Peth, M. A., Lotz, J. M., Freeman, P. E., et al. 2016, MNRAS, 458, 963
- Petrosian, V. 1976, ApJ, 209, L1
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A1
- Porter, L. A., Somerville, R. S., Primack, J. R., & Johansson, P. H. 2014, MNRAS, 444, 942
- Ripley, B. D. 1981, Spatial statistics
—. 1988, Statistical inference for spatial processes
- Robertson, B., Bullock, J. S., Font, A. S., Johnston, K. V., & Hernquist, L. 2005, ApJ, 632, 872
- Robertson, D. R. S., Gallo, L. C., Zoghbi, A., & Fabian, A. C. 2015, MNRAS, 453, 3455
- Sandage, A., Binggeli, B., & Tammann, G. A. 1985, AJ, 90, 1759
- Sanders, D. B., & Mirabel, I. F. 1996, ARA&A, 34, 749

BIBLIOGRAPHY

- Sargent, M. T., Carollo, C. M., Lilly, S. J., et al. 2007, ApJS, 172, 434
- Scannapieco, C., Gadotti, D. A., Jonsson, P., & White, S. D. M. 2010, MNRAS, 407, L41
- Scarlata, C., Carollo, C. M., Lilly, S., et al. 2007a, ApJS, 172, 406
- Scarlata, C., Carollo, C. M., Lilly, S. J., et al. 2007b, ApJS, 172, 494
- Schawinski, K., Treister, E., Urry, C. M., et al. 2011, ApJ, 727, L31
- Schawinski, K., Khochfar, S., Kaviraj, S., et al. 2006, Nature, 442, 888
- Schawinski, K., Urry, C. M., Virani, S., et al. 2010, ApJ, 711, 284
- Sersic, J. L. 1968, Atlas de galaxias australes
- Shankar, F., Marulli, F., Mathur, S., Bernardi, M., & Bournaud, F. 2012, A&A, 540, A23
- Silk, J., & Rees, M. J. 1998, A&A, 331, L1
- Simard, L., & Pritchett, C. J. 1998, ApJ, 505, 96
- Snyder, G. F., Hayward, C. C., Sajina, A., et al. 2013, ApJ, 768, 168
- Snyder, G. F., Lotz, J., Moody, C., et al. 2015a, MNRAS, 451, 4290
- Snyder, G. F., Torrey, P., Lotz, J. M., et al. 2015b, MNRAS, 454, 1886

BIBLIOGRAPHY

- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008a, MNRAS, 391, 481
- . 2008b, MNRAS, 391, 481
- Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
- Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, MNRAS, 458, 242
- Tacchella, S., Carollo, C. M., Renzini, A., et al. 2015, Science, 348, 314
- Taghizadeh-Popp, M., Heinis, S., & Szalay, A. S. 2012, ApJ, 755, 143
- Taniguchi, Y. 1999, ApJ, 524, 65
- Terlevich, R., Davies, R. L., Faber, S. M., & Burstein, D. 1981, MNRAS, 196, 381
- Thilker, D. A., Vinsen, K., & Galaxy Properties Key Project, P. 2014, in American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts #223, 116.11
- Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, MNRAS, 458, 4477
- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, ApJ, 783, 85
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, ApJ, 750, 99

BIBLIOGRAPHY

- Toomre, A. 1977, in Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley & R. B. G. Larson, D. Campbell, 401
- Toomre, A., & Toomre, J. 1972, ApJ, 178, 623
- Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, MNRAS, 438, 1985
- Trujillo-Gomez, S., Klypin, A., Colín, P., et al. 2015, MNRAS, 446, 1140
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014a, ApJ, 788, 28
- van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014b, ApJ, 788, 28
- van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJ, 677, L5
- van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 709, 1018
- Vapnik, V. N., & Vapnik, V. 1998, Statistical learning theory, Vol. 1 (Wiley New York)
- Villforth, C., Hamann, F., Koekemoer, A., et al. 2013, ArXiv e-prints, arXiv:1303.1874
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
- Wellons, S., Torrey, P., Ma, C.-P., et al. 2015, MNRAS, 449, 361
- Whitaker, K. E., Kriek, M., van Dokkum, P. G., et al. 2012, ApJ, 745, 179

BIBLIOGRAPHY

- White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
- Williams, C. C., Giavalisco, M., Cassata, P., et al. 2014, ApJ, 780, 1
- Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labb  , I. 2009, ApJ, 691, 1879
- Wuyts, S., Labb  , I., Franx, M., et al. 2007, ApJ, 655, 51
- Wuyts, S., F  rster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000a, AJ, 120, 1579
- . 2000b, AJ, 120, 1579
- Zhao, D. H., Mo, H. J., Jing, Y. P., & B  rner, G. 2003, MNRAS, 339, 12
- Zolotov, A., Dekel, A., Mandelker, N., et al. 2015, MNRAS, 450, 2327

Vita

Michael Andrew Peth was born November, 21st 1988 to Nancy and Stephen Peth. He received a B. S. degree in Astronomy & Astrophysics and a B. S. degree in Physics from the Pennsylvania State University. He was admitted into the Schreyer's Honors College in 2008. He began his astronomy research during the spring of his sophomore year working with Dr. Mercedes Richards on developing a novel method for visualizing 3D tomography data for a binary star system. He then continued his research as a summer intern working with Dr. Donald Schneider and Dr. Nic Ross on creating catalogs of quasars from the SDSS and UKIDSS surveys. Eventually writing his senior thesis entitled: *Near Infrared Photometric Properties of 130,000 Quasars: A SDSS-ULAS Catalog and the K-Band Quasar Luminosity Functions*. In the fall of 2010 he enrolled in the Physics & Astronomy P.h.D. program at the Johns Hopkins University. Under the guidance of Dr. Jennifer Lotz his research has focused on studying the evolution of galaxies through morphology with the use of machine learning techniques.