

# Empirical Star Formation Estimates in Galaxy Evolution

by

Nityasri Mandyam Doddamane

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

NEW YORK UNIVERSITY

JANUARY, 2019

---

Michael R. Blanton

© NITYASRI MANDYAM DODDAMANE  
ALL RIGHTS RESERVED, 2019

# Abstract

One of the crucial unsolved puzzles in astronomy has been the question of how galaxies end the process of star formation. A key piece in this puzzle is figuring out how to robustly infer star formation rates and the cosmic star formation histories of galaxies from observed photometric and spectroscopic data. The star formation history of a galaxy is encoded in its spectrum, which can be thought of as a fingerprint of a galaxy. However, spectroscopic information is often unavailable as it is more expensive to obtain. And where available, it can be limited in terms of interpretability by the dependence on the fiber diameter used and the physical properties of the galaxy in question.

The class of techniques known as SED (Spectral Energy Distribution) fitting tackles this problem by inferring the spectrum of a galaxy from photometric and spectroscopic data. Over the last decade, SED-fitting methods have become increasingly more sophisticated both in terms of model predictions as well as accounting for effects of more complex physical mechanisms such as dust absorption/emission. Another significant development, has been the rapidly growing field of IFU (Integral Field Unit) spectroscopy with a view to obtaining spatially resolved spectra with high S/N ratio of galaxies. With the advent of ongoing surveys such as MaNGA (Mapping Nearby Galaxies at Apache Point Observatory), the largest IFU-based survey thus far, we are better poised to answer this question than ever

before.

In this dissertation, I examine these two techniques for estimating star formation histories. First, using data from NASA Sloan Atlas (NSA) catalog along with the Wide-field Infrared Survey (WISE), I compare star formation rates obtained from two different methods: one, a UV-to-IR SED fitting method that accounts for dust and the other, a purely UV photometry-based approach. Using galaxy environments as a third independent parameter, I find a population of dust obscured star formers that masquerade as much lower star formation galaxies when only UV-optical information is available, but which live in the same large scale environments as other star forming galaxies.

In the second part, I examine the robustness of star formation history measurements from one of the largest available and most influential catalogs available, the Sloan Digital Sky Survey (SDSS) Legacy Survey. This catalog is limited by the fact that it uses small aperture fibers (3 arcsec diameter) to measure galaxies that can sometimes have much larger angular extent. Spatially resolved spectroscopy from MaNGA measures galaxies more completely, but for a much smaller sample. I use aperture measurements of key spectral indicators of age, stellar mass and star formation history such as the  $H\delta_A$  absorption line index and the  $D_{n4000}$  break to quantify the effects of the small fiber aperture on the Legacy Survey sample results. MRB suggests to add: From these results I show that biases exist in the measurements of stellar mass from the SDSS Legacy Survey, but that for most galaxies these biases are small.

# Contents

ABSTRACT	3
DEDICATION	8
ACKNOWLEDGMENTS	5
o INTRODUCTION	6
o.1 A Brief History of Galaxy Evolution . . . . .	6
o.2 Questions in Galaxy Evolution . . . . .	11
o.3 Observational Indicators of Star Formation . . . . .	12
o.4 Star Formation in the Local Universe . . . . .	13
1 UV TO IR STAR FORMATION INDICATORS AND ENVIRONMENTS	14
1.1 Introduction . . . . .	15
1.2 Constructing a local sample spanning Ultraviolet to Infrared Imaging . . .	17
1.3 Estimating the Specific Star Formation Rates . . . . .	19

1.4	Environments . . . . .	22
1.5	The Environments of the Outliers . . . . .	26
1.6	Summary and Conclusion . . . . .	28
2	APERTURE EFFECTS IN STELLAR MASS ESTIMATES	30
2.1	Introduction . . . . .	31
2.2	The $H\delta_A - D_{n4000}$ plane . . . . .	34
2.3	Manga Overview . . . . .	37
2.4	Data . . . . .	37
2.5	Methods . . . . .	38
2.6	Comparison to Full Aperture Measurements . . . . .	42
2.7	Offsets in the $H\delta_A$ - $D_{n4000}$ plane . . . . .	44
2.8	Discussion . . . . .	44
3	CONCLUSION	49
	APPENDIX A SOME EXTRA STUFF	50

# Listing of figures

1.1	Short figure name. . . . .	18
1.2	Short figure name. . . . .	23
1.3	Short figure name. . . . .	25
1.4	Short figure name. . . . .	27
1.5	Short figure name. . . . .	28
2.1	Short figure name. . . . .	34
2.2	Short figure name. . . . .	36
2.3	Short figure name. . . . .	39
2.4	Short figure name. . . . .	41
2.5	Short figure name. . . . .	43
2.6	Short figure name. . . . .	45
2.7	Short figure name. . . . .	46
2.8	Short figure name. . . . .	47

2.9	Short figure name. . . . .	48
-----	----------------------------	----



THIS IS THE DEDICATION.

# Acknowledgments

HERE'S WHERE I ACKNOWLEDGE a lot of people.

# 0

## Introduction

### 0.1 A BRIEF HISTORY OF GALAXY EVOLUTION

MRB says: I would think of these first three sections as close to just a single section. I might also suggest the opening section be "Questions in galaxy evolution." You could have a paragraph or two on the thumbnail history — Herschel to Slipher and Hubble to modern day. Then give the theoretical context. Fluctuations grow, halos collapse, baryons radiate their

energy are collapse in the center, form disks, and those disks form stars. Stellar and AGN processes may feedback. Then the questions: what ends star formation at high mass? what regulates it at low mass? how are stellar mass and DM mass related? Then finally mention simulations that exist that compare the two, if reliable observations can be made to compare to.

#### 0.1.1 THE DISCOVERY OF GALAXIES

One of the momentous shifts in scientific thinking was the realization that our Sun is one of many stars in the Universe and further that the gravitationally bound system of stars, stellar remnants, gas and dust that the Sun is a part of is one of many such structures known as galaxies. The bright band of stars and dust that we observe in the night sky, which we now know as our own galaxy, Milky Way, has been a puzzling topic since ancient times. The Greek philosopher Democritus is known to have speculated that this might be a band of distant stars but this was later eclipsed by Aristotelian physics. The astronomers of medieval Islam, such as Alhazen, al-Biruni and Ibn-Bajjah, centuries later, also hypothesized that the Milky Way was made of many stars such as our own Sun. However, observational proof for this came only in the 16th century when the Italian astronomer, Galileo Galelei pointed his telescope at this band and confirmed that the Milky Way was indeed made up of a huge number of faint stars.

Since the debate on a geocentric versus heliocentric Universe was still ongoing, Galileo's observations weren't given credence until more than century later, during which time, the development of Keplerian orbital mechanics and Newtonian Theory of Gravitation had resulted in a successful explanation of the Solar System and its dynamics. The idea that the

galaxy might be a rotating configuration of a huge number of stars held together by gravitational forces that contained smaller gravitationally bound configurations such as our Solar System was first theorized by the english astronomer, Thomas Wright, in 1750. On the heels of his observation that the faint (so-called) “nebulae” observed might be distant galaxies that came out of “external creations”, philosopher Immanuel Kant hypothesized the existence of many such “island Universes” that could potentially form and evolve independently from our own, thus laying the underpinnings of the study of galaxy formation and evolution.

#### 0.1.2 EARLY STUDIES OF GALAXIES

William Herschel, in the 1780’s, surveyed the stars in the Milky Way in multiple directions and discovered that the density of stars was greater on one side than the other. One the earliest catalogues of galaxy-like objects was developed by Messier, also towards the end of the 18th century, following which William Herschel assembled a catalog of 5000 nebulae. In the 19th century, with advances in optics and instrumentation, the first telescope to distinguish between elliptical and spiral galaxies was built by Lord Rosse. Additionally, this telescope could resolve point sources within nebulae, thus confirming that galaxies were indeed “island universes” as Wright and Kant had surmised.

In early 20th century, astronomers were beginning to get interested in the chemical composition of these nebulae and started recording their spectra. With the advent of Einstein’s Special Theory of Relativity, one could calculate the radial velocity of a “nebula” based on how its spectrum is Doppler-shifted. And so it happened that the American astronomer, Vesto Slipher, became the first person to measure galactic redshifts in 1912. While study-

ing the chemical composition of bright spiral nebulae, he noticed that they were all highly Doppler-shifted, with estimated radial velocities that were much higher than the velocities of stars in the Milky Way and that there were more “red-shifted” (i.e. moving away from us) nebulae than blue-shifted ones, an observation that would later propel Edwin Hubble to propose that the Universe was expanding.

In the 1920’s a series of observations made by Edwin Hubble, Ernst Opik, etc. confirmed that Andromeda was not a part of the Milky Way and was a galaxy in its own right, thus effectively settling the “Great Debate” of the times and confirming that our galaxy was just one of many galaxies in the Universe. Georges Lemaître, a Belgian physicist, predicted on theoretical grounds, that the redshifts of galaxies should increase with distance using Einstein’s General Theory of Relativity. In 1929, Edwin Hubble looked at the distances and velocities of 46 galaxies and observed the same (Hubble cite <http://adsabs.harvard.edu/abs/1929PNAS...15..168H>): that their radial velocities increased with distance from us thus concluding that a law that we know today as Hubble’s Law (or alternately, the Hubble-Lemaitre Law.)

Edwin Hubble further analyzed the morphologies of these galaxies and came up with the Hubble Sequence, a classification of galaxy morphology, a system that endures till date. Swiss astrophysicist, Fritz Zwicky, in 1933, while studying galaxy clusters, noticed that the orbits of the galaxies were not accounted for by the mass of the seen objects leading him to believe that there must be some “missing” mass (Zwicky cite: ) that does not interact electromagnetically, thus remaining unseen (He called it *dunkle Materie*, i.e., “dark matter”).

Along with the new theories of spacetime, an expanding Universe that possibly began with a big bang, the existence of matter that doesn’t interact electromagnetically, the study of galaxies in the 1930’s thus set a paradigm shift in astronomy and cosmology in motion.

The new questions were: How were the earliest galaxies formed? What would explain the diversity in morphologies seen in galaxies today? Is the Universe set to expand indefinitely? What drives the expansion of the Universe? If the Universe's origins were homogeneous, how do we explain the heterogeneous nature of the populations of galaxies seen?

### 0.1.3 GALAXY EVOLUTION AND COSMOLOGY

The explorations in the latter half of the 20th century confirmed a few hypotheses discussed above, along with answering a few of those questions. In the 1960's and 70's Vera Rubin, Kent Ford and Ken Freeman analyzed rotation curves of spiral galaxies and provided further strong evidence for the existence of dark matter (). Rubin inferred that most galaxies contain around six times as much dark matter as visible mass () Rubin found that most galaxies must contain about six times as much dark as visible mass, thus placing new constraints on galaxy formation and evolution in the Universe.

The two predominant theories of the origin of the Universe was the "Big Bang Theory", proposed by Georges Lemaitre and the "Steady State Theory"(essentially an unchanging Universe whose density remains the same inspite of expansion due to a continuous creation of matter) proposed by Fred Hoyle. However with theories of Big Bang Nucleosynthesis (the  $\alpha\beta\gamma$  paper () that successfully explained how the elements of the of the Universe came to be formed after the Big Bang and radio source counts which were correctly accounted for by the Big Bang Theory swung cosmologists to favor this over the Steady State Theory. The final nail in the coffin was the successful detection of the Cosmic Microwave Background (CMB) (explained below), predicted by the Big Bang Theory(cite Penzias Wilson 1964). According to the Big Bang Theory, the early Universe remained hot (above 10,000 K) for

several hundred thousand years, a state that should be detectable even today in the form of microwave radiation (redshifted according to Hubble's Law) and the CMB is indeed detectable as a very low energy radiation with a blackbody temperature of  $2.7\text{K}$ .

The discovery of dark matter coupled with the confirmation of the Big Bang Theory necessitated a theory of dark matter that would successfully explain the observable Universe. In the 1980's, competing theories of hot and cold dark matter (cite ) were proposed. Eventually the cold dark matter theories won out as their prediction of the anisotropies in the CMB (cite Peebles paper: ) were successfully verified by the COBE (Cosmic Background Explorer) probe (cite ).

Before the turn of the century, the remarkable discovery of the accelerating universe resulted in the resurrection of Einstein's cosmological constant and effectively confirmed  $\Lambda$ CDM as the clear winner among the cold dark matter models. The Big Bang Theory, together with the  $\Lambda$ CDM model, this forms much of the basis of modern cosmology and the foundations of theories of galaxy formation and evolution.

## 0.2 QUESTIONS IN GALAXY EVOLUTION

### 0.2.1 STRUCTURE FORMATION IN THE UNIVERSE

Then give the theoretical context. Fluctuations grow, halos collapse, baryons radiate their energy are collapse in the center, form disks, and those disks form stars. Stellar and AGN processes may feedback.

The current model of cosmology suggests that there was a single event, the so-called "Big Bang" which resulted in the appearance of expanding space-time containing radiation, followed by an exponential expansion of space (cosmic inflation) which established the initial



properties of the Universe: homogeneous, isotropic and flat. Tiny perturbations in this early Universe are responsible for structure formation.

Navarro Frankel White: .

Weinberg: .

### 0.2.2 FUNDAMENTAL QUESTIONS IN GALAXY EVOLUTION

Then the questions: what ends star formation at high mass? what regulates it at low mass? how are stellar mass and DM mass related? Then finally mention simulations that exist that compare the two, if reliable observations can be made to compare to.

### 0.3 OBSERVATIONAL INDICATORS OF STAR FORMATION

MRB says: I assume this section is about observations of SF in galaxies. That would make sense. You need to introduce importance of dust here.

MRB says: You probably want to also have something about observation of stellar mass distribution.

#### 0.3.1 WHAT CAN WE INFER FROM PHOTOMETRY?

MRB: Maybe combine this with the next.

#### 0.3.2 WHAT CAN WE INFER FROM A GALAXY SPECTRUM?

### 0.4 STAR FORMATION IN THE LOCAL UNIVERSE

MRB: Maybe Legacy, plus MaNGA. NSA can be mentioned as a reanalysis of Legacy photometry

o.4.1 THE AGE OF DIGITAL SURVEY ASTRONOMY

o.4.2 SDSS LEGACY SURVEY

# 1

## UV to IR Star Formation Indicators and Environments

## 1.1 INTRODUCTION

Determining the formation history of galaxies is one of the most important tasks in astronomy. The average star formation rates of galaxies have reduced with cosmic time (as reviewed by ?) and star-forming galaxies “transition” to red and quiescent ones over that time (*e.g.* ?????, among others). Many mechanisms have been proposed to explain the observed patterns of galaxy evolution but there are many unresolved questions regarding them. These mechanisms are known to have a some correlation with the environment, due to the fact that galaxy type depends on environment in the present day (??).

Star formation rate estimates are an important tool in the observational study of this problem (?). The light emitted from galaxies can reveal recent star formation through the ultraviolet light of massive stars, recombination line emission in the ionized gas around those stars, in infrared light from dust in the galaxies absorbing the ultraviolet starlight, and from radio synchrotron emission from the ionized gas. Ultraviolet, optical, and infrared images and spectra are available for numerous low redshift galaxies, from the Sloan Digital Sky Survey (SDSS; ?), the Wide-field Infrared Survey Explorer (WISE; ?), and the Galaxy Evolution Explorer (GALEX; ?).

In this chapter we compare two specific ways to estimate star formation rates from imaging photometry. We concentrate on imaging because it usually provides the most complete census of the galaxy; spectroscopy is often of only part of the galaxy, usually the center. We are interested specifically in comparing ultraviolet-based star formation indicators with indicators that account for the infrared light. The main difference between these types of indicators is how they account for dust absorption. The former ostensibly are corrected for

dust absorption, but we generally expect that the latter should be more reliable and complete. However, detectable infrared light is not always available.

The presence of dust is an important complication in any analysis of galaxy light. Dust absorbs preferentially in the UV/optical region of the galaxy spectrum and both extincts and reddens the photometric data. The extinction in the visual bandpasses in star forming galaxies is highly variable but is commonly of order 50%. This energy is re-emitted in the mid- to far-IR by poly-cyclic aromatic hydrocarbon molecules as well as warm and cold dust grains. Recent studies (?) find that in the nearby universe, almost 70% of the FUV luminosity is obscured by dust on an average. Although UV estimates of star formation do attempt to account for dust attenuation (?), the extensive reliance on UV SFR in the literature (*e.g.* ?????, among others) makes it important to understand how well the dust corrections for UV SFR estimates work relative to other methods of estimating SFR.

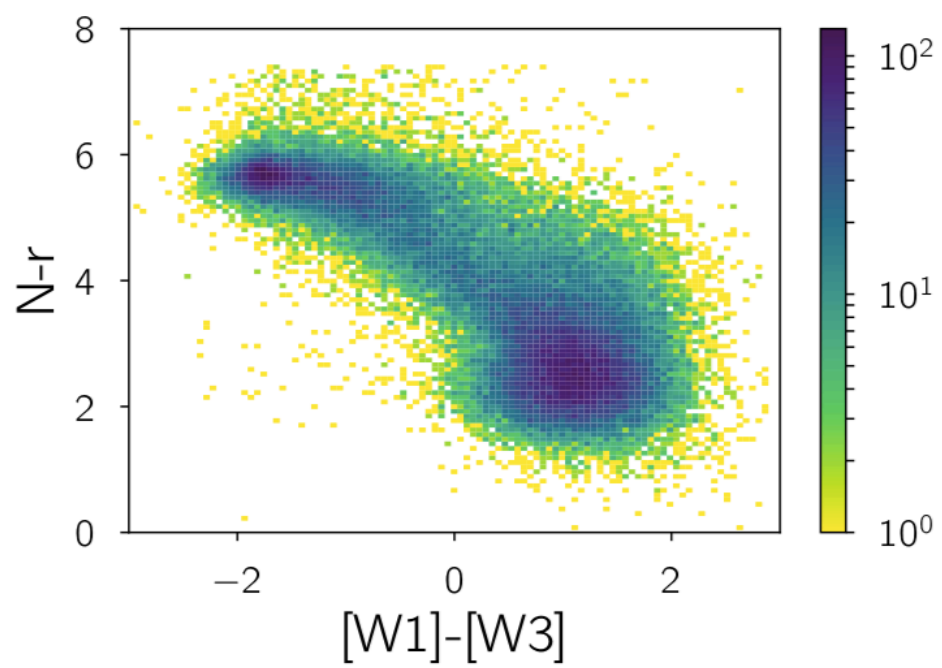
Here, we examine a sample of galaxies whose UV-IR photometry is available and estimate the star formation rate in two independent ways. First we exploit the fact that we have UV to IR photometry and perform SED fitting using MAGPHYS (?), which accounts for dust by using a simple method of energy balance to obtain the specific star formation rates (SSFRs). The other method involves using purely UV photometry to estimate both star formation rate and dust attenuation using the prescription given by Salim et al (?). We also estimate the environments of our population. We ask the following questions of this sample. How do these different star formation estimators disagree with each other? How does each estimate correlate with environment? Does one trace environment more closely? Under the assumption that the environment primarily correlates with the SSFR rather than dust obscuration's effect on the observables, studying this correlation will yield insights into

the relative accuracy of the methods.

## 1.2 CONSTRUCTING A LOCAL SAMPLE SPANNING ULTRAVIOLET TO INFRARED IMAGING

The sample on which we perform our measurements is based on the NASA-Sloan Atlas, a nearby galaxy sample which includes optical and ultraviolet imaging from SDSS and GALEX. We use NSA version v0\_1\_2, whose redshift range extends up to  $z = 0.055$  and includes SDSS (Data Release 9) and GALEX imaging for galaxies. The total number of objects in this NSA catalog is  $\sim 150,000$ . We define two samples from the NSA, one that we use to define environment (Environment Determining Population or EDP, hereafter) and one for which we determine star formation rates (Star Formation Rate Population or SFRP hereafter). We note that the SFRP is a subsample of the EDP. To get our EDP, we impose a volume-limited cut on this sample by retaining only galaxies with  $-24.5 < M_r < -18.5$ , which leaves us with 95,638 galaxies.

The star formation rate determinations we use here require both UV data and IR data. About 80% of our galaxies have UV data from GALEX analyzed in the NSA. To obtain infrared imaging for the data we use the Wide-Field Infrared Survey Explorer (WISE) data and match the objects from the NSA with the ALLWISE Source Catalog to get the four-band infrared fluxes for each galaxy. About 99% of galaxies in this sample have WISE data. We intend to study the evolution of these galaxies through the optical-IR color space, and therefore remove all galaxies with missing/faulty photometry in GALEX, SDSS, or WISE. This process includes limiting the color range of objects in the optical ( $7.5 > N - r \geq 0$ ) and infrared ( $3.0 > [W_1] - [W_3] \geq -3.0$ ) colors (top left corner of Fig. 1). In addition, we remove some objects near the edges of the sample, which do not have reliable environment



**Figure 1.1:** The Local sample distribution across optical and IR colors

determinations (see Section 1.4.2). Our SFRP ultimately consists of 61, 046 objects with absolute magnitudes in the UV ( $F$  and  $N$  bands), optical ( $u$ ,  $g$ ,  $r$ ,  $i$  and  $z$  bands) and infrared ( $W_1$ ,  $W_2$ ,  $W_3$  and  $W_4$  bands).

### 1.3 ESTIMATING THE SPECIFIC STAR FORMATION RATES

We compare two approaches to calculating specific star formation rates (SSFRs). The first uses UV, optical, and IR data. The second uses only the UV data.

#### 1.3.1 SED FITTING - MAGPHYS

We develop here a method to quickly estimate SSFRs based on UV-optical and infrared colors. We begin by sorting galaxies into bins of the  $(N-r) - ([W_1] - [W_3])$  color space ( $25 \times 25$  bins, where the optical and IR bin sizes are 0.29 and 0.24 in magnitudes respectively). Within each bin, we normalize the fluxes of each galaxy relative to a constant flux in the  $r$ -band, and then take the mean normalized flux in each band over all galaxies in the bin. This procedure yields a “template” SED for each bin in the color-color space.

We then use Multi-wavelength Analysis of Galaxy Physical Properties (?) software to infer the SSFR for each template SED. MAGPHYS is a physical model-based method to interpret the mid- and far-infrared spectral energy distributions of galaxies, self-consistently modeling the emission at ultraviolet, optical, and near-infrared wavelengths. For every input galaxy with a set of observed fluxes in different bands, MAGPHYS generates an optical and infrared library at that redshift and then samples all template spectra whose fluxes obey a simple principle of energy balance: that the amount of energy absorbed by dust in the UV/Optical matches the amount of infrared emission that is accounted for purely by dust.



Once the templates have been sub-sampled thus, MAGPHYS uses chi-squared fitting to see which combination best reproduces the observed fluxes along with the likelihood for the distributions. The results for the SSFRs thus obtained for each template are shown in the lower left panel of Fig 1.2 along with the number distribution of the galaxies across the chosen bins. Note that bins with  $< 5$  galaxies were omitted as those regions of the color-color space are obviously under-sampled.

The resulting distribution of the MAGPHYS-based SSFRs across the color-color space looks as we would expect it to for the most part. In UV-optical colors, blue galaxies have high SSFRs, and red galaxies have low SSFRs. However, there is also a dependence of SSFR on IR color; most notably, in the UV-optical green valley, the redder galaxies in the IR have higher SSFRs. These galaxies are likely to be dust-extincted in the UV, with the reemission by dust reddening the IR colors. In what follows, we will use the UV-optical and IR colors of individual galaxies, and the dependence of the SSFR on these colors shown in the lower left panel of Fig 1.2, to assign an SSFR to each individual galaxy.

### 1.3.2 UV STAR FORMATION RATES

We also explore a simple method of determining star formation rates developed by ? that depends only on the UV fluxes of each galaxy. It assigns a star formation rate that is proportional to the UV luminosity (specifically the FUV band if we are looking at GALEX). Dust attenuation is also accounted for in this method by looking at the ratio of luminosities in the FUV and NUV bands.

According to this prescription (?), the star formation rate is given by:

$$SFR = 1.08 \times 10^{-28} L_{FUV}^{\circ},$$

where  $L_{FUV}^{\circ}$  is the rest-frame FUV luminosity. This method accounts for dust attenuation of the FUV light as well by estimating an attenuation factor  $A_V$  as follows.

If  $N - r \geq 4.0$ , i.e. for the red sequence galaxies,

$$A_V = \begin{cases} 3.32(F - N) + 0.22, & \text{if}(F - N) < 0.95 \\ 3.37, & \text{if}(F - N) \geq 0.95. \end{cases}$$

If  $N - r < 4.0$ , i.e. for the blue sequence galaxies,

$$A_V = \begin{cases} 2.99(F - N) + 0.27, & \text{if}(F - N) < 0.90 \\ 2.96, & \text{if}(F - N) \geq 0.90. \end{cases}$$

This method uses the UV slope to estimate the dust attenuation. The physical idea is that when the UV light is dominated by massive young stars, these are hot enough that we are observing their Rayleigh-Jeans tail spectra, and this slope is basically constant with age of the population. Thus the observed UV color reveals the dust reddening.

The lower right panel of Fig. 1.2 shows the mean  $SSFR_{UV}$  as a function of color. Compared to the lower left panel, the  $SSFR_{UV}$ 's have a nearly monotonic relationship with UV-optical color whereas, as discussed above, the  $SSFR_{MAGPHYS}$ 's do not. Assuming the  $SSFR_{MAGPHYS}$ 's are closer to correct, the UV-optical green valley contains a population of galaxies that are not truly transitioning but instead are reddened by the presence of dust. Furthermore, the

nominally dust-corrected UV star formation rates do not successfully identify these galaxies.

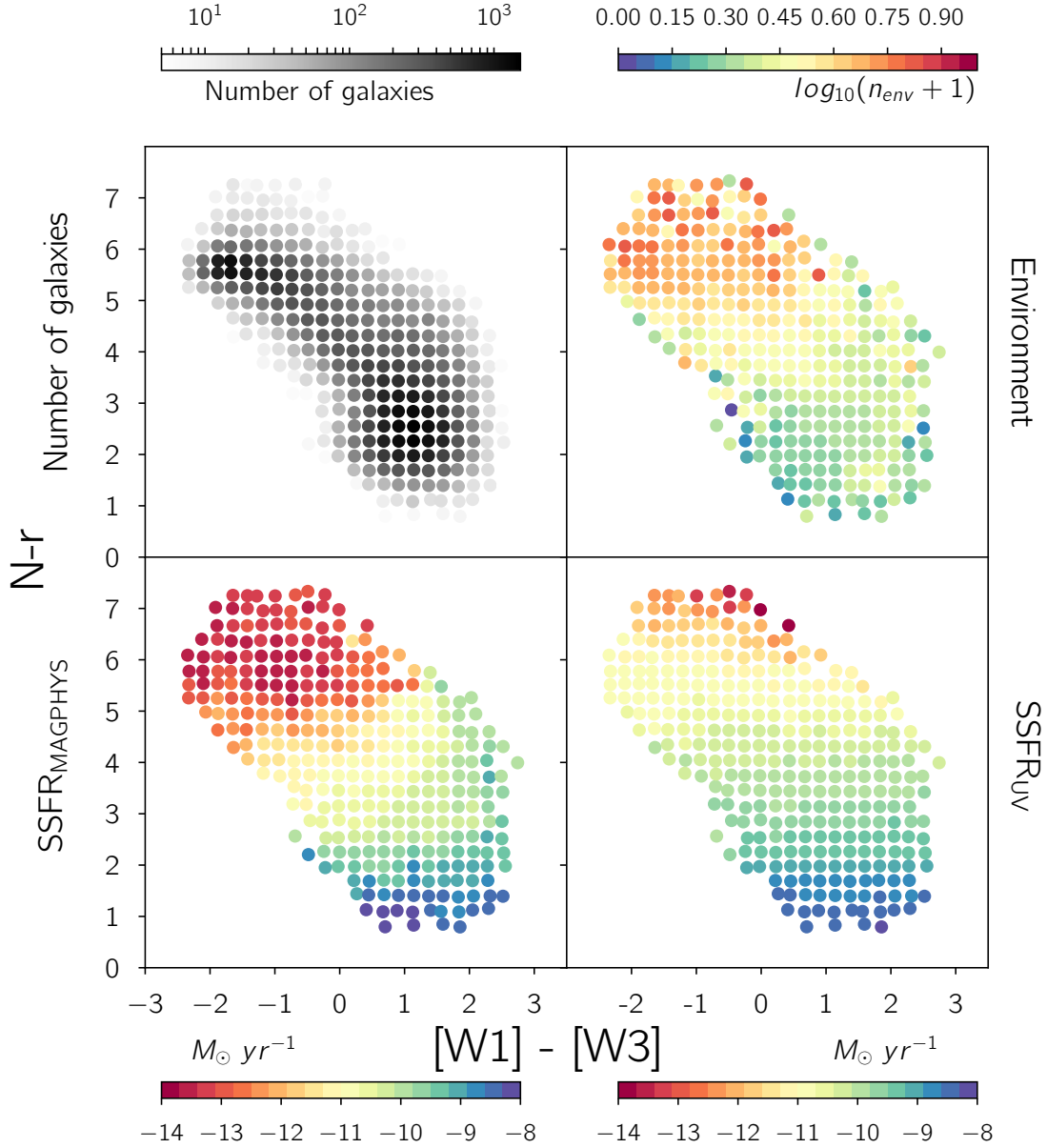
MRB: Quantify the fraction of galaxies like this; say within  $2.5 < N - r < 4.5$ , what fraction have  $[W_1] - [W_3] > 1.5$ , answering same question in 3 bins of absolute magnitude too.

#### 1.4 ENVIRONMENTS

We identified above a population of galaxies isolated in color space, which appeared to have dust-extincted star formation, such that the  $\text{SSFR}_{\text{UV}}$  estimate was much less than the  $\text{SSFR}_{\text{MAGPHYS}}$ . In order to confirm whether MAGPHYS is capturing an inherent property in this population of galaxies, we examine an independent physical property, namely the environments of our sample.

##### 1.4.1 MEASURES OF ENVIRONMENT

The environment of a galaxy can be defined in many ways, such as fixed aperture counts, distance to the  $n^{\text{th}}$  nearest neighbor, Voronoi volumes, etc (?). Here we use counts in a projected fixed aperture of radius 0.5 Mpc as our environment measure. Around every galaxy we construct a projected cylinder with a radius (in the transverse direction) of  $r_t = 0.5$  Mpc and a line of sight velocity window of  $v_{\text{los}} = \pm 1000 \text{ km/s}$ . For each galaxy for which we have star formation measurements, we count its number of neighbors ( $n_{\text{env}}$ ) in this cylinder (Fig. 1.2) in the Environment Defining Population.



**Figure 1.2:** The SSFR measurements and environment shown the color-color space. Each point in the plot is shown at the mean colors in each of the bins we use. The grey value or color of the points in each panel show the mean value in each bin for the quantity described by the corresponding color bar. *Top left:* Logarithmic number density in each of the bins; all bins with less than 5 galaxies were discarded in this and the other panels. *Top right:* Environment; in each bin, the average number of nearest neighbors is calculated in a projected cylinder ( $r_t = 0.5 \text{ Mpc}$  and  $v_{los} = \pm 1000 \text{ km/s}$ ). *Bottom left:* the Specific Star Formation Rates obtained from MAGPHYS. *Bottom right:* UV Specific Star Formation Rates estimated by using the method described in ?.

#### 1.4.2 EDGE EFFECTS

We must also account for the survey edges. For galaxies at the edge, part of the fixed aperture used to estimate the environments might lie outside the survey coverage. It is important to identify these galaxies and either discard them or assign an appropriate weight to  $n_{\text{env}}$  in order to account for the missing area.

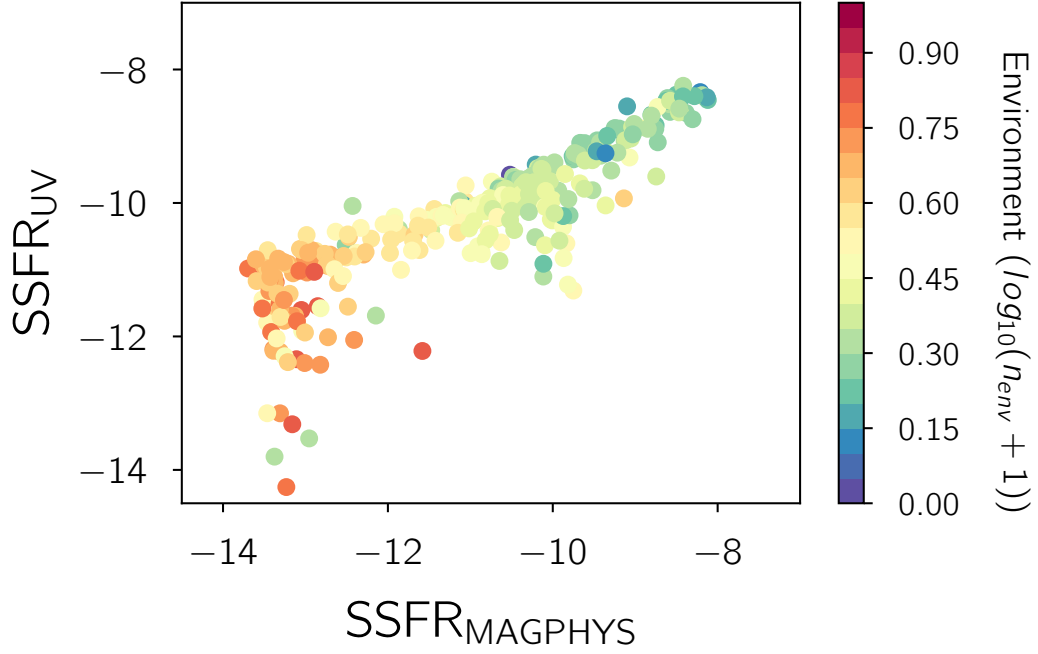
To identify the edges, we use (Swanson et. al.'s) *Mangle*, a suite of free open-source software designed to deal with complex angular masks in an efficient and accurate manner. First, the NYU-VAGC *Mangle*-format mask was used to obtain the angular mask for the NASA Sloan Atlas by using the *polyid* routine from *Mangle*. Then, the *ransack* routine was used to populate the mask with a random sample of  $N = 10,000,000$  galaxies. For each galaxy, we compute the angular separation  $\vartheta_i$  that corresponds to our  $0.5\mathcal{Mpc}$  aperture at the redshift of that galaxy. We then count the number of galaxies  $n_i$  that lie within this angular separation and compare the value obtained to the expected value:

$$\langle n \rangle_i = \frac{N}{\mathcal{A}_{\text{EDP}}} \times \pi \vartheta_i^2 \times f_{\text{thresh}}$$

$\mathcal{A}_{\text{EDP}}$  is the total area of the mask and  $f_{\text{thresh}}$  is the fractional threshold for the edge effect cut-off. In our case, we chose  $f_{\text{thresh}} = 0.8$ . Wherever  $n_i < \langle n \rangle_i$ , we consider the galaxy to be near an edge and discard it from our sample.

#### 1.4.3 ENVIRONMENTS IN COLOR-COLOR SPACE

The mean environment, quantified by  $(\langle n_{\text{env}} \rangle + 1)$  in each bin are shown in the upper right panel of Fig. 1.2 as a function of optical and infrared colors. As we expect from many



**Figure 1.3:** The two star formation rate estimates (from Fig.1.2) plotted against each other as a function of the environment; We notice two distinct set of outliers that seem to have lower UV SSFR's but similar environments to the galaxy bins with the same MAGPHYS SSFR's.

previous studies (?) the star-forming bluer galaxies tend to exist in less dense environments on average and the red-and-dead population tends to exist in more dense environments.

In the UV-optical green valley region ( $3 < N - r < 5$ ) there is some indication that the environment declines with  $[W_1] - [W_3]$  color at fixed UV-optical color. However, the mean environments in this plot cannot easily be interpreted, because the mean stellar mass in each bin is different, so some of the variation is driven by the dependence of environment on stellar mass.

## 1.5 THE ENVIRONMENTS OF THE OUTLIERS

Fig. 1.3 shows the relationship between the two SSFR measurements, with points colored according to the mean environment, for each bin in color-color space from Fig. 1.2. There is a set of bins in the range  $-10.5 < \text{SSFR}_{\text{MAGPHYS}} < -9.5$  that have lower  $\text{SSFR}_{\text{UV}}$  values than the general trend, and have environments similar to the other galaxy bins in the same MAGPHYS range. Hereafter we shall refer to this population as the “Outliers,” to distinguish them from the general trend line in Fig. 1.3.

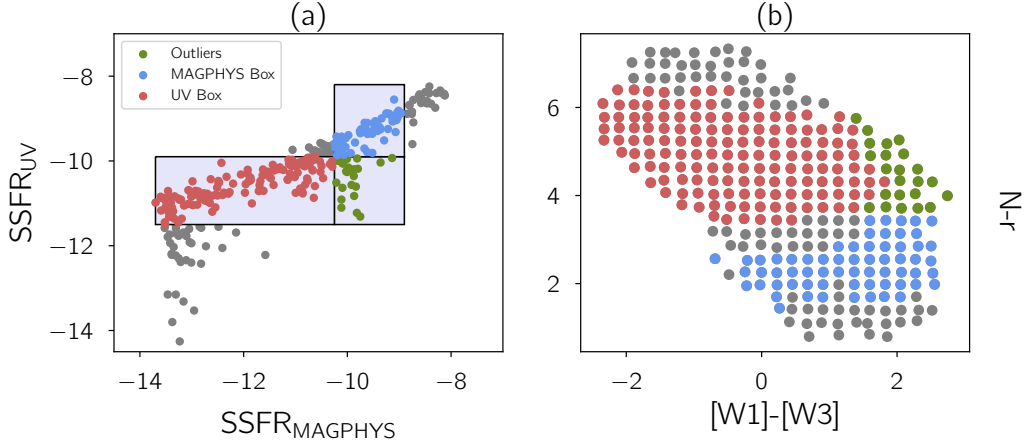
We identify these outliers more specifically in Fig. 1.4(a), as being in the lower right square. The galaxy bins with the same  $\text{SSFR}_{\text{MAGPHYS}}$  values (above the square) are identified as the “MAGPHYS Box” and the bins with the same  $\text{SSFR}_{\text{UV}}$  values (left of the square) are identified as “UV BOX” in Fig. 1.4(a).

.. change line styles of different boxes in Fig.1.5 so that the result can be inferred on b-and-w print...

### 1.5.1 GREEN-VALLEY INTERLOPERS

When we return to the color-color space (Fig. 1.4(b)) and show where the bins in each box lie, we see the trends we would expect. The Outliers lie in a similar UV-optical color range as the UV box but have higher IR color. The MAGPHYS Box occupies the bluer side in the optical color range while spanning almost the entire IR color range. The UV Box occupies the redder side in the optical color range while having lower IR color values. The Outlier bins are the same bins we previously identified as the galaxies in the UV-optical “green valley” that are there due to dust reddening.

To verify this, we unwrap the bins and look at the Probability Density Function of the



**Figure 1.4:** The outliers shown as a function of the Star Formation Rates as well as Optical and IR colors

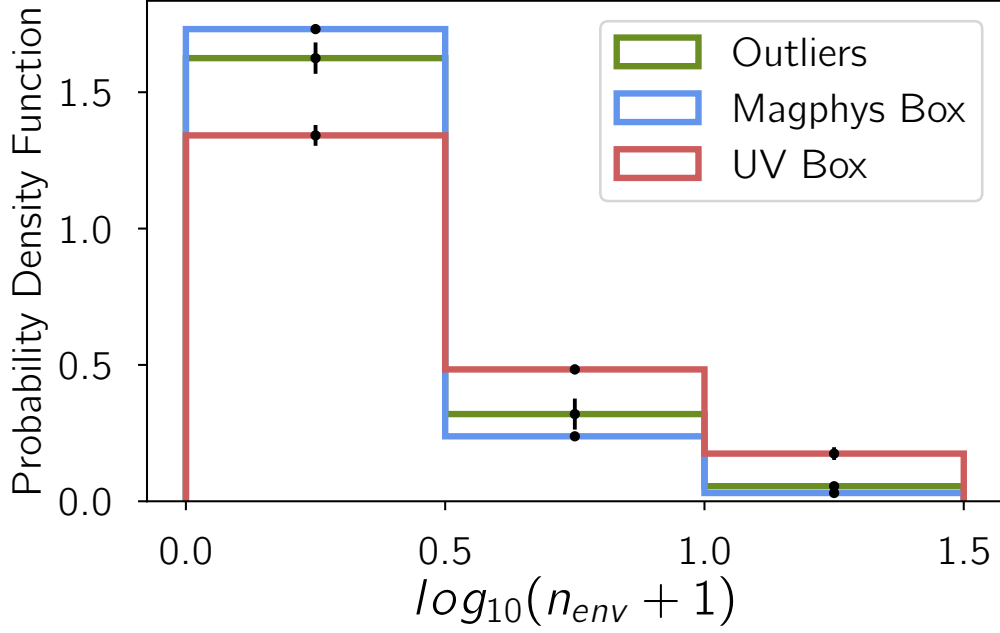
Environments of the galaxies in these three regions in the mass range  $9.5 < \mathcal{M}_* < 10.7$  and examine the distribution of environments in these three regions, the result of which is plotted in Fig. 1.5.

### 1.5.2 JACKKNIFE ERRORS

To calculate uncertainty in the estimated probability density functions of the environments,  $P_{i,\text{bin}}$ 's :  $i = 1, 2, 3$  for each of the populations, we use the standard jackknife technique. Jackknife re-sampling gives us an internal error estimate that tests how representative a measurement/trend is of the data it is estimated with. We divide our entire sample into 20 subsamples with nearly equal co-moving volumes and estimate the same probability density functions ( $P_i^j$ 's) for the whole sample while leaving out one subsample each time. We can then estimate our uncertainty for each bin in the PDF's thus:

$$\sigma_{i,\text{bin}} = \sqrt{\frac{N}{N-1} \sum_{j=1}^{j=N} (P_i^j - P_{i,\text{bin}})^2}$$





**Figure 1.5:** Probability density functions of the Environments of the three populations described in Fig. 2

The errors estimated in this manner account for Poisson shot noise and also the sample variance, the extra error associated with the fact that the density field varies across the survey. Although precision studies of large scale structure have found that latter effect is not perfectly accounted for with standard jackknife techniques, they are precise enough for our purposes here. The jackknife errors are shown in Fig. 4 and confirm our hypothesis: that the Outliers are a little different than the MAGPHYS Box but still, very different from the UV Box.

## 1.6 SUMMARY AND CONCLUSION

- From Fig 1, we see that  $SSFR_{MAGPHYS}$  identifies a region in the color-color space as

dust-obscured star forming galaxies and correlates better with the environments of the galaxies.

- At the higher star formation end, we find that the dust-obscured star-formers as identified by MAGPHYS have environments comparable to the blue star-forming galaxies, confirming that this is indeed a physical effect we're seeing.
- Comparing the environment distribution of the Outliers relative to the galaxies with (a) the same  $\text{SSFR}_{\text{MAGPHYS}}$ 's as the Outliers and (b) the same  $\text{SSFR}_{\text{UV}}$ 's as the Outliers (Fig. 1.5), we find that the Outliers indeed have a similar environment distribution to the galaxies that have the same  $\text{SSFR}_{\text{MAGPHYS}}$ 's, i.e., they seem to favor lower environment densities mimicking the behavior of star-formers.

# 2

## Aperture effects in Stellar Mass Estimates

## 2.1 INTRODUCTION

In this chapter we investigate methods of constraining the star formation histories and stellar masses of galaxies where galaxy spectra are available. In particular, we look at the largest galaxy catalog of estimated stellar masses, star formation rates and gas metallicities (the MPA-JHU catalog; NEED CITATION) obtained for the Sloan Digital Sky Survey (SDSS) Legacy Survey. The MPA-JHU catalog has, over the last decade and a half, been one of the most influential and widely-used catalogs in the fields of galaxy formation and evolution. Here we test a fundamental assumption of the catalog, which is that spectroscopic measurements of the central region of the galaxy yield sufficient information to constrain star formation histories and stellar masses for the galaxy as a whole. MRB says: You need to describe briefly why we want to measure these quantities. Call back to Chapter 0.

The SDSS spectra are obtained for multiple objects at a given time. This is done by connecting the spectrographs to fiber optic cables which are plugged into an aluminum plate that is placed in the telescope's focal plane. These optical fibers have a fiber diameter of  $3''$ . Thus for each galaxy in the SDSS, while photometry is available for the entire galaxy, spectra are available only for the region of the galaxy contained within this  $3''$  angular aperture. Thus, the fraction of the galaxy falling within the aperture may vary widely with the redshift of this galaxy. I describe below how the MPA-JHU catalog approached this problem.

The stellar masses and star formation histories in the catalog are obtained by looking at two key spectral indicators of starbursts and age of a galaxy: the Balmer  $H\delta_A$  absorption line index and the  $D_{n4000}$  break index, both of which are relatively insensitive to the metallicity-age degeneracy issue that spectral indicators in the redder side of the spectrum

are limited by. Because they are both measurements over a narrow range of wavelengths, that are similar to each other, and are independent of the absolute flux, they are designed to be insensitive to dust extinction within the galaxy. Using the distribution of the observed galaxies in the  $H\delta_A - D_{n4000}$  plane and by employing Stellar Population Synthesis (SPS) models by ? to model the evolution of galaxies in this plane, they are able to infer a mass-to-light ratio in the  $z$ -band (?) for every point in this phase space. From here, it is a short step to using the luminosity of the entire galaxy to infer the stellar mass and star formation history of the galaxy.

However, this methodology can be problematic when the mass-to-light ( $M/L$ ) ratio of the central part of a galaxy is not representative of the entire galaxy. This can be an issue where there are distinct bulge and disk components, such as for luminous spiral galaxies. However, with the availability of spatially resolved spectra for galaxies from IFU (Integral Field Unit) based surveys such as MaNGA (Mapping Nearby Galaxies at Apache Point (?)), SAMI (Sydney-Australian-Astronomical-Observatory Multi-object Integral-Field Spectrograph (?)), MUSE (Multi Unit Spectroscopic Explorer (?)) and CALIFA (Calar Alto Legacy Integral Field Area Survey) (?), we can actually put this question to the test and quantify the effect of aperture size on the SDSS stellar masses. We use MaNGA, the largest IFU-based survey thus far, which is based on SDSS, and ask the following questions: At any given redshift in the local Universe, how does the position of the galaxy in the  $H\delta_A - D_{n4000}$  plane as measured using a  $3''$  aperture compare to using a full aperture, i.e. spectra from all the spatially resolved regions in the galaxy.

### 2.1.1 THE SDSS SPECTRA

Describing the SDSS spectrograph and 3'' fiber diameter. References: (?) (York et al citation).

The galaxy spectra in SDSS are obtained through 3'' diameter fibers. The rest-frame wavelength range of the spectra at the median redshift is from 3500 Å to 8500 Å with a spectral resolution

$$R = \frac{\Delta\lambda}{\lambda} \approx 2000$$

. The spectra are calibrated using observations of F stars in each 3 degree field.

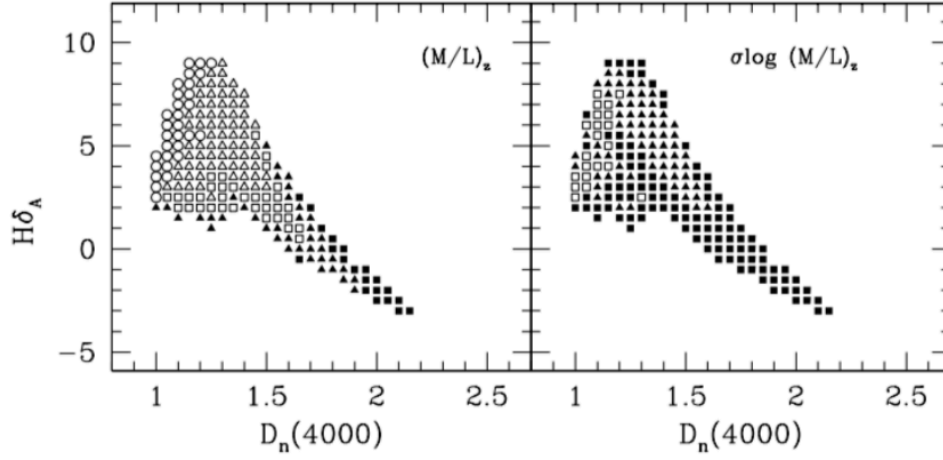
### 2.1.2 THE MPA JHU CATALOG

Introduction to and describing the importance of the largest catalogue of galaxy stellar masses (?), star formation rates (?) and gas metallicities (?) thus far. Including major results that rely on said SFR's.

MRB says: Then you need to mention some major results that depend on the MPA-JHU catalog measurements of this quantity, just a couple of highlights from the later section. TBD: Kewley et al 2006 (host galaxies of AGN's); Elbaz et al 2007 The reversal of the star formation-density relation in the distant universe; Kewley 2008 Metallicity Calibrations and the Mass-Metallicity Relation for Star-forming Galaxies; Peng et al 2010 (mass and environments as drivers etc); Kauffmann 2013 A re-examination of galactic conformity and a comparison with semi-analytic models of galaxy formation.

.....

The MPA-JHU catalog was developed as a result of a joint collaboration between MPA



**Figure 2.1:** The  $\delta$  grid to infer M/L ratios from the  $H\delta_A - D_{n4000}$  plane

and JHU right after the SDSS Data Release 1 (DR1) of spectroscopic observations. Using the previously mentioned spectral indicators ( $D_{n4000}$  and  $H\delta_A$ ) and broadband photometry from SDSS, they developed a method to derive maximum likelihood estimates of the stellar mass of a galaxy, the attenuation of optical light by dust and the fraction of stars in a galaxy formed in recent bursts. The dataset they used was 120,808 galaxies drawn from the SDSS DR1.

## 2.2 THE $H\delta_A - D_{n4000}$ PLANE

Two powerful spectral diagnostic tools that are valuable in constraining stellar masses and star formation histories in galaxies are the Balmer  $\delta$  absorption line index and the  $D_{n4000}$  break index in the optical spectrum of a galaxy (?). One of the key issues in spectral indicators for starbursts is the existence of the age-metallicity degeneracy in stellar populations (?). The heavier element lines are often prominent in both early type as well as metal rich galax-

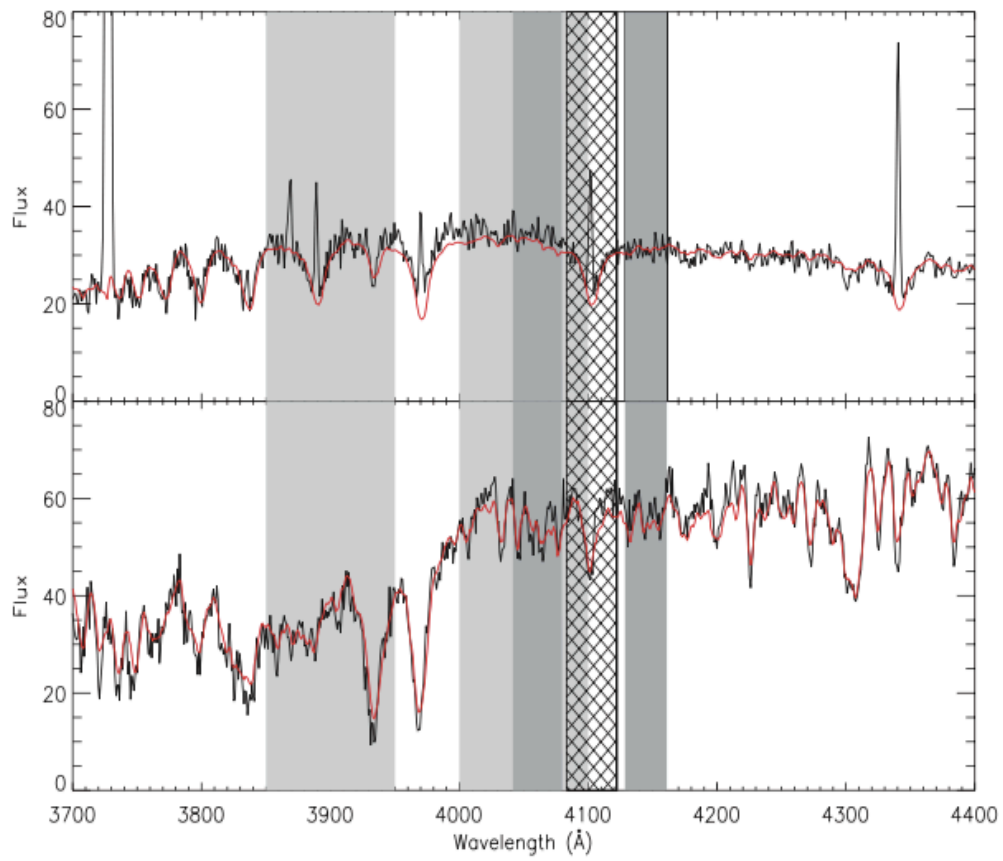
ies and the two are hard to distinguish in terms of interpretation as older stellar populations exhibit metal richness as well and without accurate modeling of the chemical evolution of galaxies, it is difficult to ascertain the reason. The Balmer  $H\alpha$  and  $H\beta$  are sensitive to the effects of metallicity as well. However both the Balmer  $H\delta$  absorption line as well as the  $D_{n4000}$  break occurring in the continuum are both in a bluer region of the spectrum relatively unaffected by metallicity constraints and together they form a powerful diagnostic tool to constrain the mass-to-light ratios of galaxies.

### 2.2.1 MEASURING THE $D_{n4000}$ INDEX

The most obvious discontinuity in the spectrum of a galaxy is often the  $D_{n4000}$  break (?), which is a consequence of the absorption into electronic excited states by neutral and partially ionized metals whose lines happen to overlap in the region just blueward of 4000 Å. Older stellar populations have spectra dominated by relatively cool giant stars, and under these conditions the partially ionized metal lines are strong. Younger stellar populations have spectra dominated by hot stars, for which these lines are weaker. In addition the break depends on stellar metallicity, exhibiting some of the same age-metallicity degeneracy that afflicts most spectroscopic indicators.

This break was defined in by ? as the ratio of the average flux density  $F_\nu$  in two bands that span  $\sim 200$  Å on either side of 4000 Å. ? defined a narrow band definition for the same where the bands that are now 100 Å wide with 3850 Å- 3950 Å as the “blue continuum” and 4000 Å- 4100 Å as the “red continuum”. The bandpass used in ? (2.2) is the same as the one introduced by ? and their motivation for doing so is that the narrow definition is more insensitive to reddening effects. I reproduce these measurements using the same narrow





**Figure 2.2:** PLACEHOLDER FIG; TBD: A figure showing D4000 and Hdelta breaks in spectra for early/late-type galaxies as well as the bimodality in D4000

band definition.

### 2.2.2 MEASURING THE $H\delta_A$ INDEX

## 2.3 MANGA OVERVIEW

Citations: (?)

### 2.3.1 INTRODUCTION TO INTEGRAL FIELD SPECTROSCOPY

What is Integral Field Spectroscopy? What is an Integral Field Unit?

### 2.3.2 THE MANGA IFU DESIGN

What is a spaxel? (?)

## 2.4 DATA

### 2.4.1 MANGA TARGET SELECTION AND DRP

Primary and Secondary Samples. NSA redshifts/Luminosity cut.

### 2.4.2 OUR SAMPLE

The most recent MaNGA product launch, MPL-8, which was announced in November 2018, containing products based on galaxy and stellar library observations from March 2014 - July 2018 serves as the source of our sample. It contains 950 plates - 6779 data cubes and 20649 stellar library stars. Out of these, 6468 are galaxies with measured NSA redshifts.

These are representative of all the IFU sizes (list them: 19,37... 127) and span a redshift range upto  $z = 0.15$ .

For each galaxy observation, depending on the IFU bundle size (say  $N_x, N_y$  each describing a position  $0.5''$  from the previous spaxel), we have  $(N_x, N_y)$  spectra which span 4563 wavelength points.

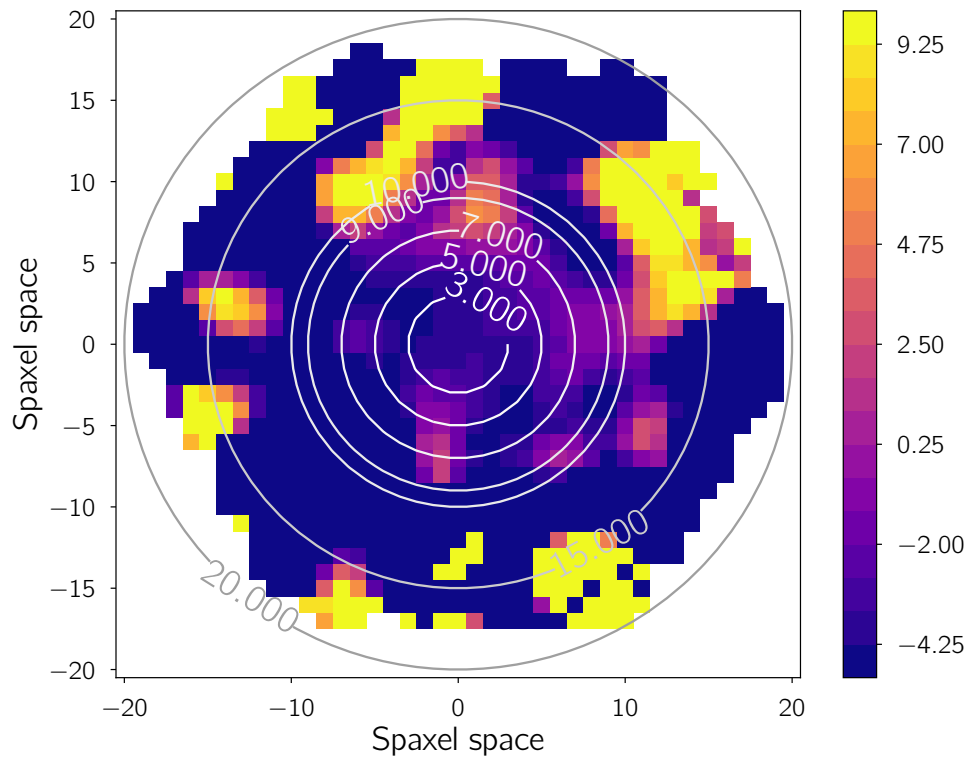
## 2.5 METHODS

### 2.5.1 VARIABLE APERTURE MEASUREMENTS

For any given datacube, we can determine which spaxels fall within an aperture radius of  $R$  arc-seconds as follows. As each spaxel spans a width of  $0.5''$  in along the “X” and “Y” direction, at any point  $(x, y)$  in the IFU image, the distance in arc-seconds of the centre of a spaxel from the central spaxel  $(x_c, y_c)$  in the IFU would be:

$$d = (r(x, y) - r(x_c, y_c)) = \sqrt{(0.5 * x)^2 + (0.5 * y)^2}$$

Thus, for every spaxel in the IFU, where  $d \leq R$ , that part of the galaxy would fall within the aperture and hence, we would include that spaxel in the measurement of whichever spectral index. Using this, for instance, we can simulate the SDSS fiber measurement, i.e., what the  $1.5''$  aperture radius used in SDSS would see versus the total galaxy or what we call a “full aperture” measurement.



**Figure 2.3:** Placeholder figure: TBD. Sample Manga galaxy and the Hdelta distributions with the 3" fiber at the three redshift bins

### 2.5.2 VARIABLE REDSHIFT MEASUREMENTS

One can alternately pose this question in terms of redshift. For all galaxies whose redshift  $z_{\text{obs}}$  is less than the redshift we are interested in,  $z_{\text{cutoff}}$ , we can ask the question: if we shift the said galaxy to the said cutoff point, where would its location on the  $H\delta_A - D_{n4000}$  plane be? How offset is this from the “full aperture” measurements?

Transverse angular distance varies with redshift as follows:

$$D_A(z) = \frac{D_M(z)}{1+z}$$

, where  $D_M(z)$  is the co-moving distance at redshift  $z$ .

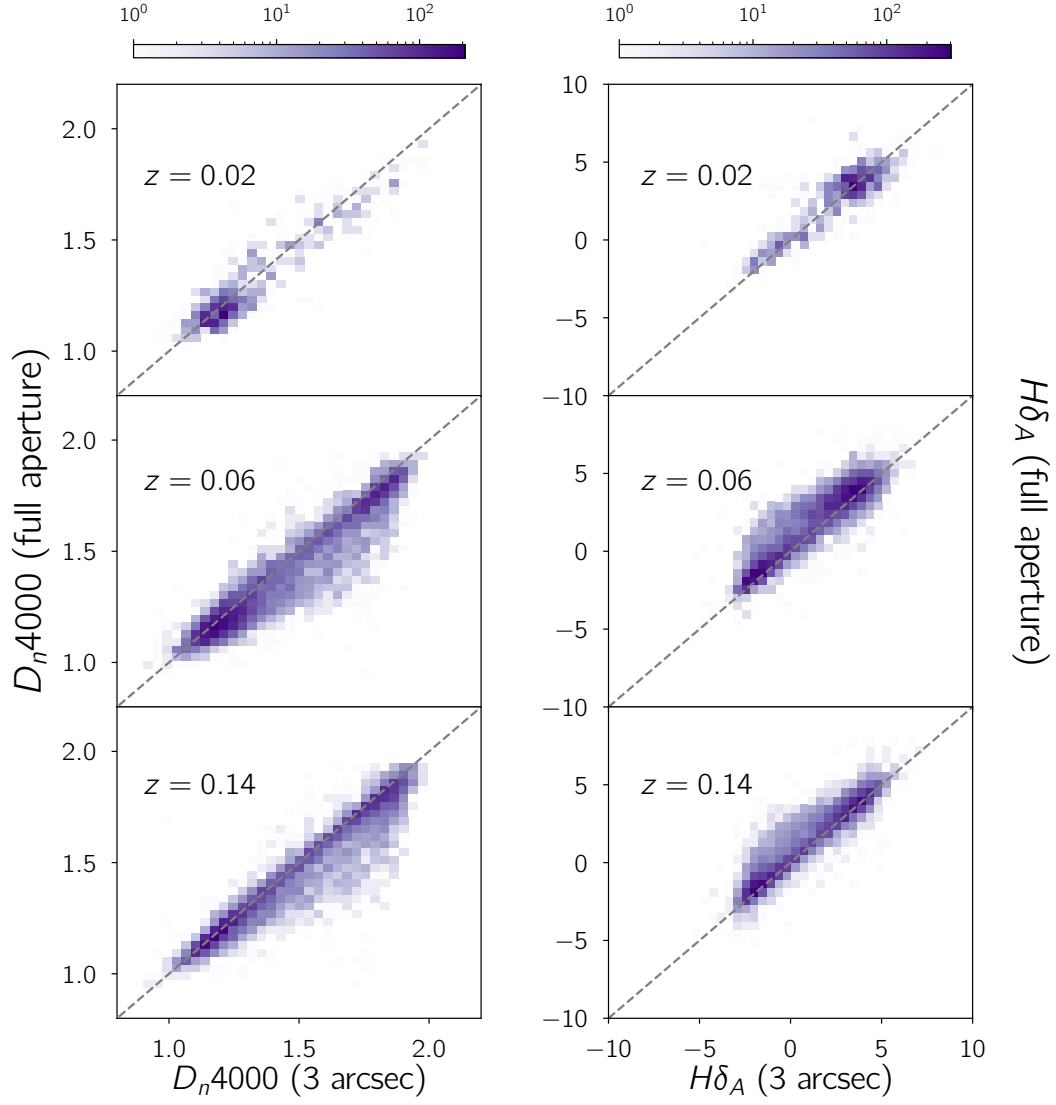
So when a galaxy at  $z_{\text{obs}}$  is shifted to  $z_{\text{cutoff}}$  the new distance  $d_{\text{new}}$  of spaxel  $(x, y)$  from the central spaxel  $(x_c, y_c)$  relates to the old distance thus:

$$d_{\text{new}} = \frac{(1+z_{\text{cutoff}}) \times D_M(z_{\text{obs}}) \times d}{(1+z_{\text{obs}}) \times D_M(z_{\text{cutoff}})}$$

Using the above, we can now figure out the spaxels that would fall within a 3” diameter aperture (say) at not only the observed redshift but also any redshift where we would like to collectively observe the behavior of the offset with a full aperture measurement in the  $H\delta_A - D_{n4000}$  plane.

### 2.5.3 $H\delta_A, D_{n4000}$ MEASUREMENTS WITHIN APERTURES

To get the  $H\delta_A$  and  $D_{n4000}$  index measurements for any aperture, we first redshift-correct the spectra obtained from all the spaxels within the aperture to rest-frame. As both the



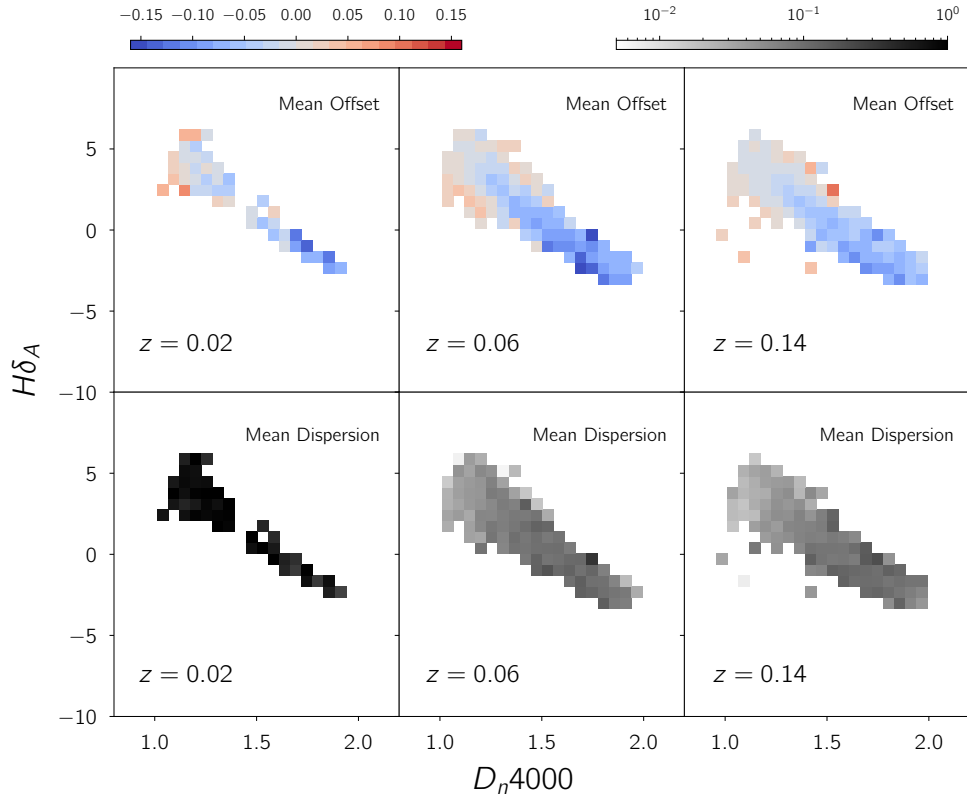
**Figure 2.4:** The  $D_{n4000}, H\delta_A$  indices measured at  $z = 0.02, 0.06, 0.14$  with a  $3''$  aperture compared to the full aperture measurement

equivalent width of the Balmer  $H\delta$  line as well as the  $D_{n4000}$  break rely on the continuum as well as a ratio of fluxes in the case of the latter, we add up the spectra of the spaxels that fall within any aperture before estimating either. We then follow the procedure described in Section 2.2 to calculate the indices. (In a footnote: the code for this is publicly available at [–provide github link–](#).)

## 2.6 COMPARISON TO FULL APERTURE MEASUREMENTS

To investigate how the full aperture measurements compare to the 3" measurements, we choose 3 different redshift bins  $z = 0.02, 0.06$  and  $0.14$  and observe how offset the  $H\delta_A$  and  $D_{n4000}$  measures are. For each redshift, I pick the galaxies whose redshift are below that redshift and “shift” them to the cutoff redshift as described in Section 2.5 and compare the full aperture measurements to the 3" measurements made at that redshift. The results of this are shown in Fig. 2.4, where the full aperture measurements are plotted against the 3" measurements at the respective redshifts with the colorbars indicating the number of galaxies in each bin. We note that the total number of galaxies in each redshift bin is 561 for  $z = 0.02$ , 5016 for  $z = 0.06$  and 6402 for  $z = 0.14$ .

From Fig. 2.4, we can infer the following. For most of the galaxies the offset is pretty minimal as it as they fall pretty close to the  $x = y$  line. We find that the scatter is the most at  $z = 0.06$  and tends towards higher  $D_{n4000}$  and lower  $H\delta_A$  values relative to the full aperture measurements. This can be interpreted as the effect of the apertures beginning to cover the galaxy disks which have lower star formation activity than the bulge for galaxies with strong bulges. The scatter lessens as we approach  $z = 0.14$ , close to the survey limit, where for most of the galaxies, all of the light gets accounted for by the 3" aperture at this



**Figure 2.5:** The mean offset and dispersion in the  $D_{n4000}$  index measured at  $z = 0.02, 0.06$  and  $0.14$  with a  $3''$  aperture from the full aperture measurement



point.

MRB Says: At all redshifts, we can see that there is a systematic tendency for galaxies to appear older through the fiber aperture than through the full aperture, and therefore to appear to have a higher M/L than they actually do. This would particularly affect the intermediate age galaxies (i.e. those with  $D_{n4000}$  in the green valley).

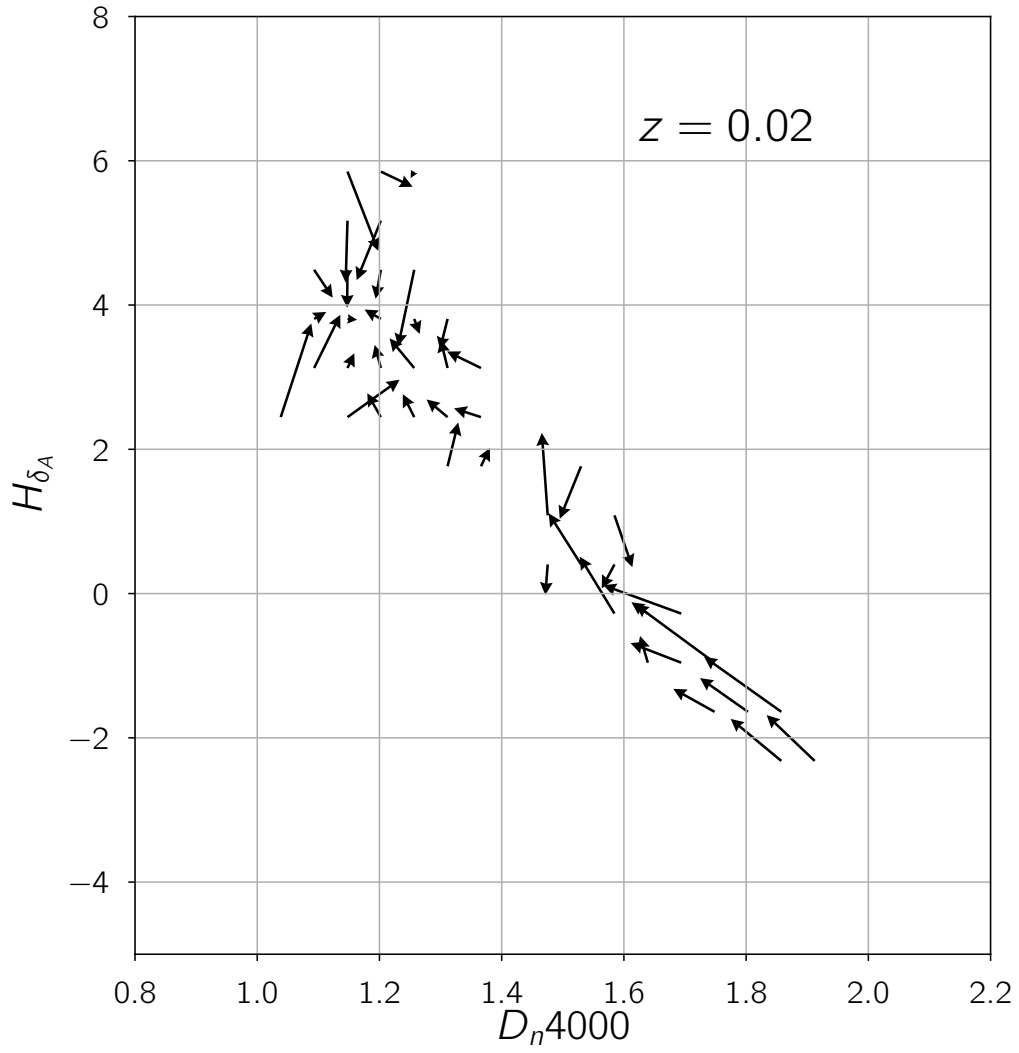
## 2.7 OFFSETS IN THE $H\delta_A$ - $D_{n4000}$ PLANE

## 2.8 DISCUSSION

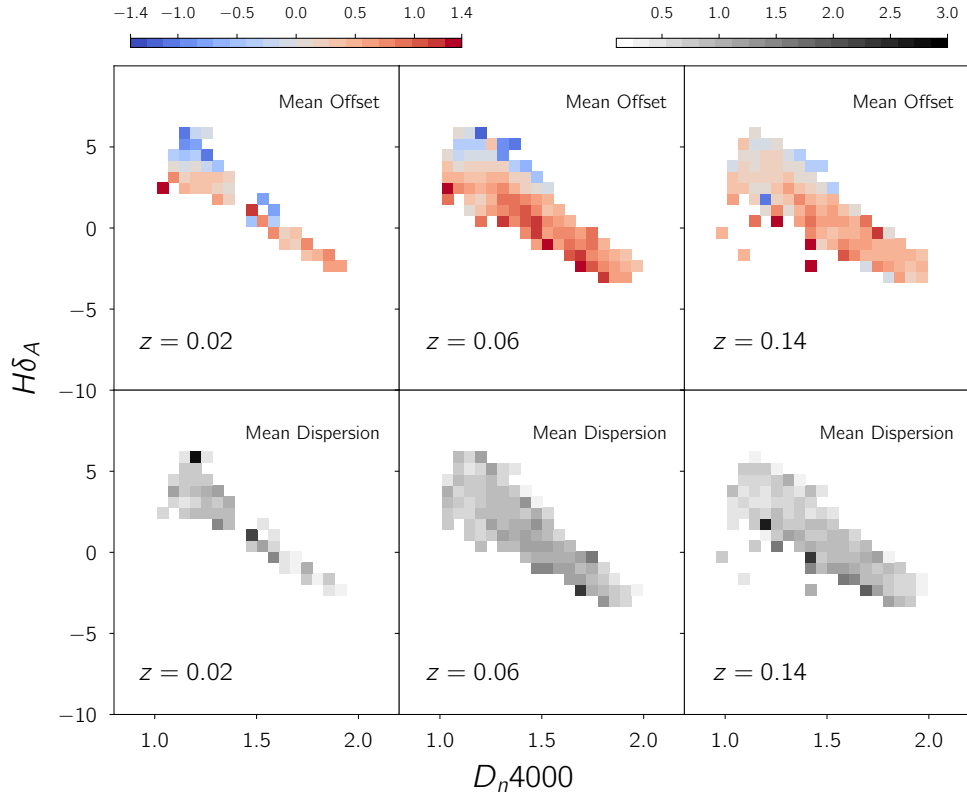
MRB Says: \* This mean tendency is \*relatively\* small. You should calculate from your grids the median offset for cells with  $D_{4000} > 1.3$  or something like that, and quote that number. You should then look carefully at the Kauffmann figure to translate this to a change in  $M/L_z$  — my guess is that it is around 0.1 dex for  $z=0.06$ . Do this for  $z=0.06$  and  $z=0.14$ .

\* The scatter about this mean is substantial relative to the mean. Again, calculate the median standard deviation over the cells with  $D_{4000} > 1.3$  to quantify this. My guess is that it is comparable to the mean itself. This means for individual galaxies the error can be a lot bigger. Again quote for both those redshifts.

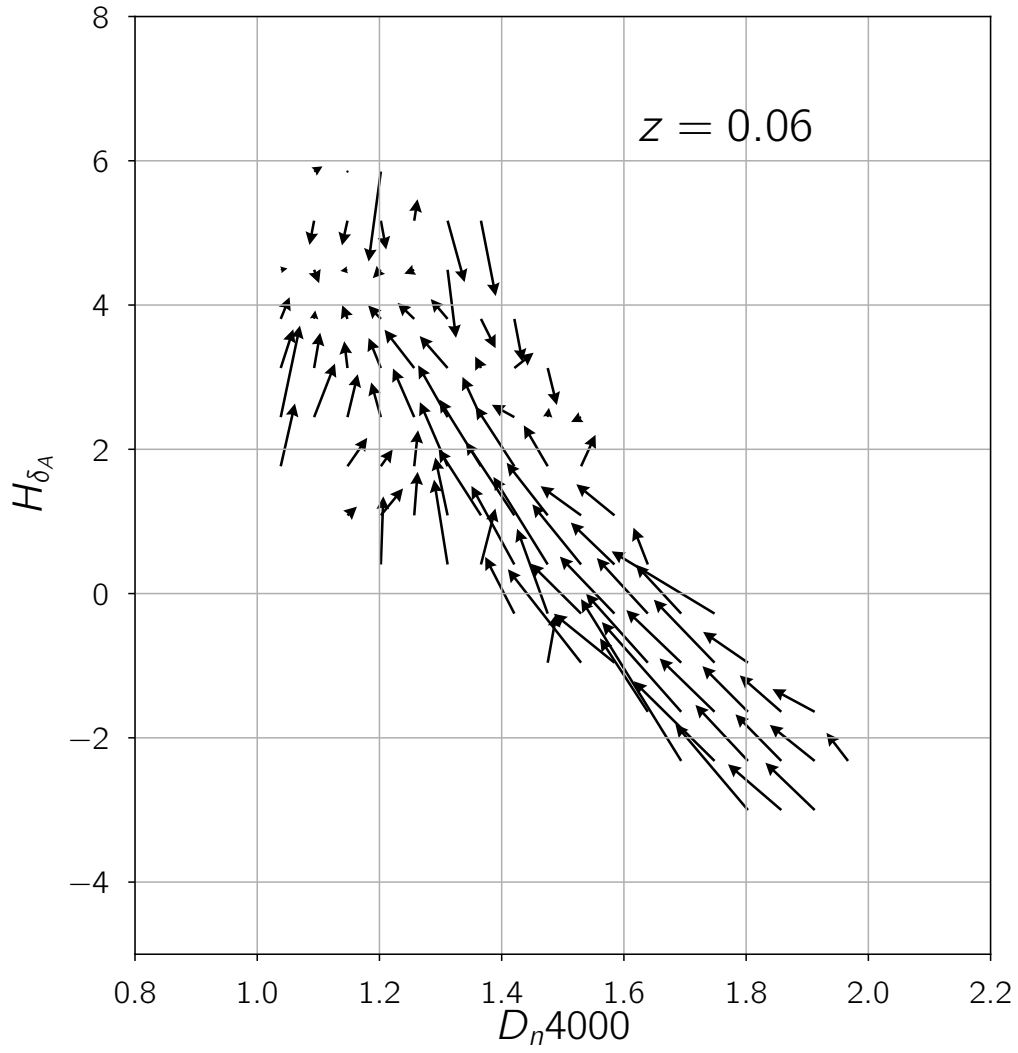
\* The main conclusion is that stellar masses more accurate than 0.1 dex (or whatever you see in the mean offset) are not possible, and that the scatter in the offset means the precision is degraded as well. This source of error is important to keep in mind but is probably not fatal for samples at  $z=0.1$  from MPA/JHU — most results done rely on stellar masses at that precision. Using this sample at lower redshift may be much more iffy.



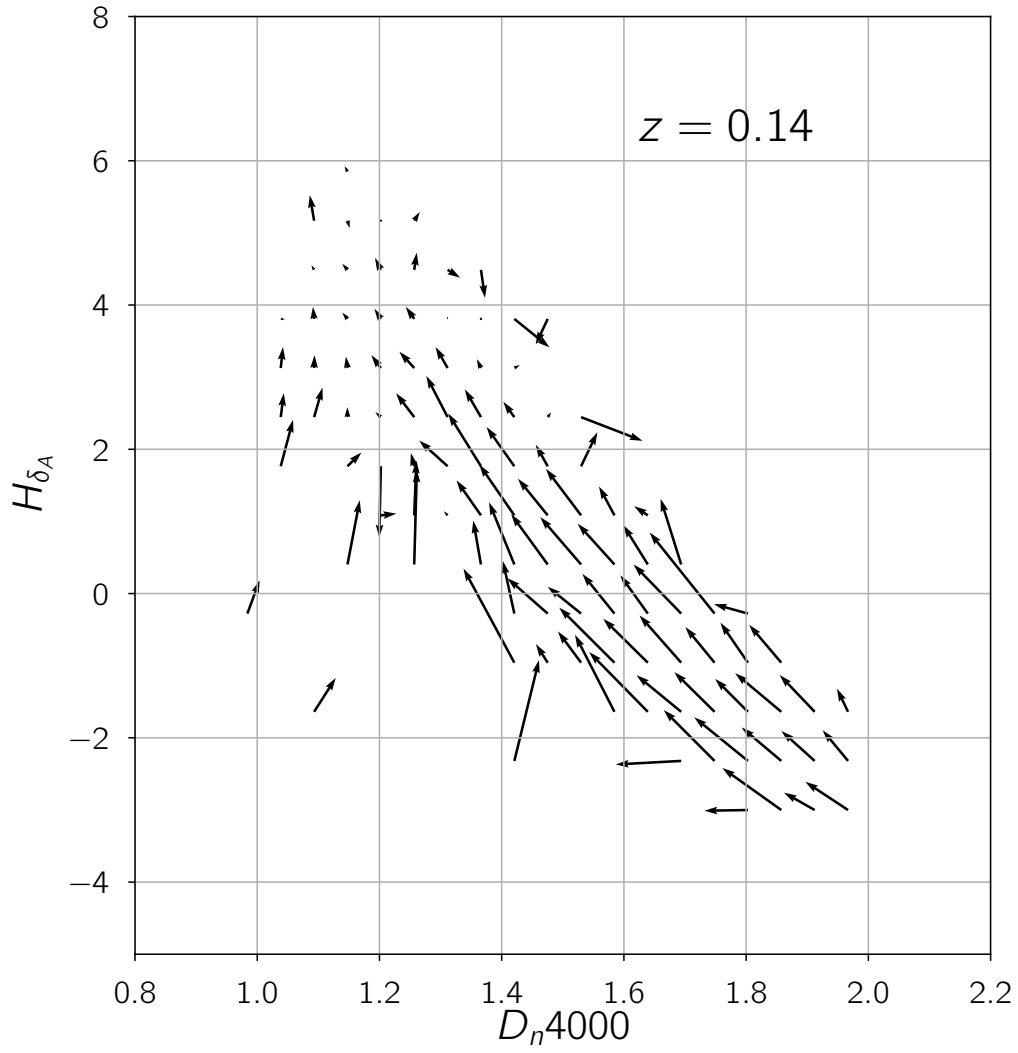
**Figure 2.6:** The mean offset and dispersion in the  $H_{\delta_A}$  index measured at  $z = 0.02, 0.06$  and  $0.14$  with a  $3''$  aperture from the full aperture measurement



**Figure 2.7:** The mean offset and dispersion in the  $h\delta_A$  index measured at  $z = 0.02, 0.06$  and  $0.14$  with a  $3''$  aperture from the full aperture measurement



**Figure 2.8:** The mean offset and dispersion in the  $h\delta_A$  index measured at  $z = 0.02, 0.06$  and  $0.14$  with a  $3''$  aperture from the full aperture measurement



**Figure 2.9:** The mean offset and dispersion in the  $H\delta_A$  index measured at  $z = 0.02, 0.06$  and  $0.14$  with a  $3''$  aperture from the full aperture measurement

# 3

## Conclusion



Some extra stuff

