Title
A Statistical Modeling Approach to Selection and Study of Galaxies at Different Phases of their Star-Formation Activity at High Redshift

Author
Shahidi, Abtin

Publication Date
2021

UNIVERSITY OF CALIFORNIA
RIVERSIDE

A Statistical Modeling Approach to Selection and Study of Galaxies at Different Phases of their Star-Formation Activity at High Redshift

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Abtin Shahidi

December 2021

Dissertation Committee:

Dr. Bahram Mobasher, Chairperson
Dr. Laura V. Sales
Dr. Brian Siana

The Dissertation of Abtin Shahidi is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First of all, I wanted to thank all the people who have helped me during my Ph.D. studies. I apologize if I forget to mention you all.

I am incredibly grateful to my parent, who supported me throughout my life and encouraged me to follow my dreams. Mom, thanks for teaching me how to be kind, forgiving, and persistent in reaching for my aspirations even if there are obstacles. Dad, I wish you were here with us so that I could share these moments with you. Losing you was the most challenging moment of my life, but I was lucky that you taught me how to be Stoic and withstand the adversities that life brings. Your memory is always with me, and your poetry calms my hectic mind like always.

I want to thank my advisor, Bahram Mobasher, for supporting me during these years and guiding me through some of the most challenging moments. I constantly learned from you and your insights in numerous areas. You always pointed me in the right direction with such positivity, thoughtfulness, and reality, gave me guidance and invaluable advice when I needed it, not just about academics. I will be forever indebted to you for all these. I always counted on your wise advice and will keep doing so in the future. Also, I wanted to thank Dr. Azin Mobasher for always treating me like a family and making studying away from my family much easier to handle.

I thank my best friend, Maedeh, who always listened to my complaints, helped me in difficulties, and kept me calm. I am fortunate to have you in my life, even though oceans separate us. I was very lucky to find Mostafa when I came to Riverside; your friendship is one that no matter how many days I don't talk to you or see you, we are just going to pick

up where we left things off. I deeply cherish the adventures we had and the memories we created.

I also want to thank some of my most amazing friends, Hooshang, Shouby, Irene, and Behnam. You always made me happier, offered me great friendship, and created incredible memories that I always remember. Thanks for sharing your knowledge and expertise, which I greatly benefited from in this study.

I also wanted to thank my groupmates in our research group, Nima, Zahra, Sina, Ali Ahmad, Marziye, Amanda, and Stephen. I especially thank two of my most amazing friends, Nima and Zahra. I learned a lot from you and your excellent suggestions. I could always count on you, and you guys always made me happy. I enjoyed every moment of studying and working next to you at UCR. I am also very grateful for finding Sina as a new friend and my housemate, as you provided me with a lot of support and help during the stressful months before graduation.

I am grateful for having the chance to find some fantastic friends during my stay in Riverside, Hamed, Amir, Zahra, Fariborz, Elahe, Ali, Shima, Guga, Farima, Mahdi, Aida, Niusha, Parnian, Justin, Amanda, Ethan, Christina, and Xinnan. I enjoyed the time I spent with you all.

I also wanted to thank the students and faculty in our department, which made the graduate school much more enjoyable and created a fantastic atmosphere. I especially thank my thesis committee members, Laura V. Sales and Brian Siana, for their insightful suggestions, helpful comments, and discussions. I was also fortunate to get many suggestions and insights from some of the best astronomers and astrophysicists during my Ph.D. in different projects, including Shoubaneh Hemmati, Hooshang Nayyeri, Henry C. Ferguson, Rachel Somerville, Charles L. Steinhardt, Istvan Szapud, Behnam Darvish, Tommy Wiklind, Mark Dickinson, Steven L Finkelstein, Camilla Pacifici, Casey Papovich, Adriano Fontana, Mauro Giavalisco, Anton Koekemoer, Jeffery Newman.

At last, I thank all my teachers from primary to graduate school, whom I am always indebted to.

I acknowledge the support by NASA MUREP Institutional Research Opportunity (MIRO) grant number NNX15AP99A for providing the funding that allowed me to undertake part of this research. Some part of this thesis is based on research that has been published and is a reprint of the material as it appears in journal article entitled: “Selection of Massive Evolved Galaxies at $3 \leq z \leq 4.5$ in the CANDELS Fields" published on July 1, 2020 in the Astrophysical Journal, Volume 897, Number 1 (Shahidi et al. 2020ApJ...897...44).

To mom and dad, Azar and Homayoun.

ABSTRACT OF THE DISSERTATION

A Statistical Modeling Approach to Selection and Study of Galaxies at Different Phases of their Star-Formation Activity at High Redshift

by

Abtin Shahidi

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, December 2021
Dr. Bahram Mobasher, Chairperson

This thesis focuses on the selection and study of galaxies based on their star formation activity at high redshift. I use multiple selection techniques from the traditional color selection to a Bayesian model averaging approach with Bayesian SED fitting to select the massive quiescent galaxies at $z \geq 3$ in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS). I compare predictions from empirical to the latest cosmological hydrodynamical simulations and find that number and stellar mass density are higher than predictions. I estimate their halo mass using abundance matching, which results in massive enough halos that shock heating for some fraction of the gas can explain part of their quenching process. But as cold streams are expected to be significant even for massive halos at these redshifts, other quenching mechanisms such as AGN feedback must have the dominant role. I then develop models for selecting these objects using a statistical learning approach to allow a robust and computationally efficient selection in the upcoming extensive surveys. I train and validate different methods using the mock catalog from the semi-analytic models for the CANDELS. Many of these techniques outperform the generic SED-fitting approach applied on the large catalogs and make more robust samples in terms of completeness and purity. Finally, I build a probabilistic model for jointly describing the galaxies' stellar mass, star-formation rate, and local density contrast, using a mixture model while accounting for different sources of uncertainties. I find that the effect of the environment on the prediction of a galaxy's star-formation activity is different when moving towards higher redshifts. The impact of the environment on the odds of being quiescent versus star-forming has small interaction with stellar mass at low redshift but shows strong interaction at high redshift (> 1) such that the effect of the environment is larger for more massive galaxies. The effect of the mass also depends on the environment and becomes larger in a denser environment. This is partly consistent with the picture that galaxies with halted cosmological gas accretion can become quiescent by depleting their gas reservoir through star-formation and outflows.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Within the past fifty years the standard model of cosmology have been developed with enormous success in describing the geometry and material content of the universe as a whole. This model is capable of reproducing the observed spatial structures in the universe that span wide range of scales from galactic scale to galaxy groups, clusters, and the cosmic web, starting from an initial condition set after the Hot Big Bang which is close to being homogeneous and carry the small quantum fluctuations imprints from the inflation (Frenk and White 2012). Perhaps the most astonishing prediction of this model, is the prediction of the angular power spectrum of anisotropies in cosmic microwave background which is one the greatest achievements of physical cosmology (Planck Collaboration et al. 2018).

Within the framework of the standard model, baryons fall into the dark matter potential through gravity and start to form the matter structure of the universe, and these structure grow through hierarchical clustering and merging of the smaller structure to form larger structures (White and Rees 1978). In the original picture of galaxy formation the primordial gas get shock heated to the virial temperature of the halo and cools down by losing its internal energy through radiative processes (e.g., Rees and Ostriker 1977; Silk 1977). The effectiveness of this cooling depends on the temperature, density, and composition of this gas. Gas in the most inner regions of the halo where the density is higher, cool down more efficiently, and by losing its pressure support form a disk, which grow through later accretion of the gas from outer region with longer cooling time. However, Birnboim and Dekel (2003) showed that these virial shocks do not form at low mass halos, and it does not develop for most halos at high redshift, in addition, Kereš et al. (2005) showed that most of the low mass galaxies get their gas through "cold mode" accretion where the gas does not heat up to the virial temperature of the halo. Only for the massive galaxies and at dense environment the effect of the from shock heating of the gas seems to become relevant. In this revised picture the amount of cold to hot mode accretion of the gas depends on the halo mass and which is related to local environment of the galaxies.

Even within the original picture, White and Rees (1978) showed that if only radiative cooling processes and gravity are involved the galaxy formation is too efficient and most of the gas must cool and form stars, which overpredict the number of low mass galaxies. This means that some additional baryonic physical process must be invoked to solve this "overcooling" problem. However, the problem seems to be much more complex as have been shown that solving the low mass end through star-formation related feedback and outflows, can produce very massive galaxies at the high mass end (Benson et al. 2003). Solving these problems requires better understanding of the complex baryonic physical processes involved in the galaxy formation. Study of their effectiveness and timescales over which they interact with galaxy to shape the the overall properties of the galaxy population, is one of the most complex problems in the field.

One of the main questions to answer is how early one could find population of quiescent galaxies. One piece of evidence comes from archaeological study of stellar population in the sample of local early type and quiescent galaxies located in massive clusters, which shows that galaxies form anti-hierarchically in the sense that the massive galaxies have formed earlier compared to the low mass galaxies and they seem to have formed rapidly during a short time scale though vigorous star-formation and must have quenched rapidly as well (e.g., Thomas et al. 2005). This result shows that we could expect to find these massive quiescent galaxies at high redshift particularly at high density regions. Finding these galaxies especially at higher redshift, when the timescales of the physical process are bounded by the age of the universe. Population of massive galaxies, which build up their mass sooner and at higher pace, have been found to higher and higher redshifts ($z > 3$) with little or no ongoing star-formations, in the universe that has plenty of gas and most galaxies are going through vigorous star-formation (e.g., Straatman et al. 2014). The physical model of galaxy formation must provide a reasonable avenue to form sufficient number of these massive objects with correct physical properties, and ways to make them quiescent when the universe was only $1 - 2$ Gyr old.

In the second chapter of this thesis, I provide a robust sample of these objects in the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS), using different methods from traditional color selection schemes either in the rest frame optical or in the observed Near IR, to Bayesian modeling of the Spectral Energy Distributing (SED) using different modeling assumptions and combining them using Bayesian model averaging. I will provide a comparison of their number and stellar mass densities with the latest prediction of some semi-empirical to latest cosmological hydro-dynamical simulations.

In the third chapter, I use the mock catalogs from semi-analytical models to show that many statistical learning methods can provide more robust sample of these objects compared to generic SED-fitting to the whole catalog. I provide an argument from a decision theory perspective that selection of any class of galaxies involves using a loss function and based on that I provide two pathways to find the optimal decision based on the application at hand, one that involves probabilistic modeling of the class labels given their observed quantities, and the other is an empirical approach. First, I provide a framework to handle missing values in the data through an iterative conditional modeling approach, and then by simulating the observational conditions of the CANDELS on the mock catalog from semi-analytic models, I build probabilistic or empirical models that can select the population of quiescent galaxies at high redshift from the observed photometeric data, in the presence of noisy, censored, and missing data. I will provide a analysis of the resulting models in terms of their ability to recover galaxies as a function of their stellar mass and redshift.

In the fourth chapter, I use the latest physical estimates in the CANDELS field including: photometric redshift measurements, star-formation rate using UV+IR calibration, and local density field measurements, in addition to our estimate of the stellar mass using SED-fitting. Then using a mass complete catalog across redshift of 0.5 to 3, I build a probabilistic model for the joint probability distribution over stellar mass, star-formation rate, local density field, and galaxy type, while taking into account the measurement uncertainties and biases such as Eddington-Jefferys bias. I assume that the true stellar masses are following a Schechter distribution, and the star-formation rate is a function of stellar mass and galaxy type using a mixture modeling approach, in which, I model the mixture proportions of galaxy type as a function of stellar mass and local density. Using this model, I do an inference for all the model parameters including the relation on the mixture proportions, star-formation rate versus mass, and stellar mass function, in addition to the latent variables of the model which are the "true" values of the physical measurements. Using this model one can allow a consistent flow of information from the noisy measurements to our model parameters. I will provide the results in terms of the evolution of the effects of the environment, and stellar mass in determining the location of the galaxies in the star-formation rate versus stellar mass plane across cosmic time. I then provide an possible extension of the model to limits below the completeness limit of the survey in order to put a better constraints on the parameters of the model.

Chapter 2

Selection of Massive Quiescent Galaxies at High Redshift in CANDELS

2.1 Introduction

In the standard ΛCDM paradigm, most of the mass in the universe resides in structures known as dark matter halos. These provide the gravitational well within which cold gas collapses, forms stars, and at a larger scale forms progenitors of the galaxies we observe today. (e.g., White and Rees 1978; Fall and Efstathiou 1980; Blumenthal et al. 1984; Frenk and White 2012; Wechsler and Tinker 2018). The dark matter halo itself forms from the gravitational collapse of initial perturbations in the density field at the very early universe (e.g., van Albada 1960; van Albada 1961; Peebles 1970; White 1976).

The emergence of the hierarchical model of structure formation (Press and Schechter 1974; Gott and Rees 1975; White and Rees 1978) supported by cosmological hydro-dynamical simulations and semi-analytical models suggest a bottom-up scenario, in which massive halos formed from a sequence of mergers (so-called merger trees) and mass accretion as opposed to initial rapid collapse models (e.g., White and Frenk 1991; Navarro and Benz 1991; Katz et al. 1992; Kauffmann et al. 1993; Lacey and Cole 1993; Somerville and Primack 1999). Discovery of very massive galaxies at high redshifts ($M_s \geq 10^{10} M_\odot$), which constitute most of the luminous baryonic component inside the dark matter halos, however, suggest a rapid build-up of the bulk of their stellar mass at $z > 2$, with intense star formation activity at early times. Observations at submillimeter further confirm the starburst populations with star formation rates exceeding hundreds of solar masses per year (e.g., Blain et al. 2002; Capak et al. 2008; Marchesini et al. 2010; Smolčić et al. 2015). There have been many recent spectroscopic confirmations of such sources at high redshift galaxies experiencing suppressed star formation activity (e.g., Whitaker et al. 2014b; Belli et al. 2014; Newman et al. 2015; Schreiber et al. 2018a; Newman et al. 2018a; Belli et al. 2017b; Belli et al. 2017a; Glazebrook et al. 2017; Belli et al. 2019; Forrest et al. 2020a). To use these systems to constrain galaxy formation and evolution scenarios and to study feedback and quenching mechanisms at early times requires a robust photometric selection of these objects followed by deep spectroscopic observations.

Searching for massive evolved galaxies at high redshifts is challenging due to the faint nature of these galaxies and their small number density, requiring multi-waveband deep imaging data over large areas. The Cosmic Assembly Near-infrared Deep Extragalactic

Legacy Survey (CANDELS) (Koekemoer et al. 2011; Grogin et al. 2011), is a treasury program on the Hubble Space Telescope (HST) providing deep multi-waveband imaging data, allowing detection of such massive systems when the universe was 1-2 Gyrs old. Studying these massive systems, with relatively small star formation activity, can help us understand the mass assembly of galaxies at very early times and estimate the baryonic content of the universe that turned into stars as well as studying primary physical processes responsible for the rapid star formation activity experienced by progenitors of these galaxies. Also, since the universe has a more cold gas reservoir at an early time, we would expect a high level of star-formation that persist longer. Therefore, the low star-formation activity of the massive evolved galaxies requires an explanation about physical mechanisms involved in quenching seen in these galaxies and perhaps other mechanisms for maintaining their low star formation activity. Measuring the evolution of the number/stellar mass density of these systems with relatively high stellar mass (i.e. $M_s \geq 10^{10} M_{\odot}$) and low specific star formation rate (i.e. sSFR $\leq 10^{-10}$ yr^{-1} for $z \sim 3$ targets) will shed light on some of the outstanding questions regarding early evolution of galaxies.

Over the last two decades, different techniques had been developed to identify different populations of high redshift galaxies (Cimatti et al. 2002; Roche et al. 2002; Daddi et al. 2004; Reddy et al. 2005; Mobasher et al. 2005; Lane et al. 2007; Daddi et al. 2007; Wiklind et al. 2007; Rodighiero et al. 2007; Grazian et al. 2007; Mancini et al. 2009; Fontana et al. 2009; Williams et al. 2009; Caputi et al. 2012; Arnouts et al. 2013, Whitaker et al. 2013; Barro et al. 2014; Nayyeri et al. 2014; Straatman et al. 2014; Fumagalli et al. 2016; Pacifici et al. 2016; Siudek et al. 2017; Fang et al. 2018; Merlin et al. 2018; Carnall et al.

2018; Carnall et al. 2019b; Carnall et al. 2020). However, different methods optimized to find the same population of high redshift galaxies, often result in different samples with varying levels of contamination when applied on the same data set.

This paper compares different techniques used for the selection of massive evolved galaxies at high redshifts. Here, we apply these methods to the same data set and compare the results. We then quantify the degree of confidence for each of the detected sources to be a member of the galaxy population in question.

In Section 4.2, we present the data. Different selection techniques are introduced in section 2.3 and applied on the CANDELS data to identify massive evolved galaxies. In Section 2.4, we compare these methods and analyze the effect of photometric errors on each of them. In Section 2.5, we compare the number and stellar mass densities of the quiescent galaxies with previous values reported from observations and cosmological simulations. We discuss our results in Section 2.6. We present our final catalog of massive evolved galaxies in the appendix 2.7.

We assume a standard cosmology with $H_0 = 70\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, $\Omega_b = 0.0486$, $\Omega_m = 0.3089$, and $\Omega_\Lambda = 0.7$ from Planck Collaboration et al. (2016) unless it is stated otherwise. All magnitudes are in the AB system where $\mathrm{m}_{AB} = 23.9 - 2.5\log(f_\nu/1\mu\mathrm{Jy})$ (Oke and Gunn 1983).

2.2 Data

We use the latest photometric catalogs from the CANDELS with consistent multi-waveband photometry, and physical parameters for all galaxies to the flux limit of the sample [1]. Details about the selection and photometry at different bands were carried out for all the CANDELS fields consisting of: GOODS-South (Guo et al. 2013; Santini et al. 2015), UDS (Galametz et al. 2013; Santini et al. 2015) , COSMOS (Nayyeri et al. 2017) , EGS (Stefanon et al. 2017), and GOODS-North (Barro et al. 2019). These catalogs contain observed photometry from the UV to near and mid-infrared wavelengths in many broadband and narrowband filters (Table 3.1), as well as inferred physical parameters.

The optical (*HST*/ACS) and near-IR (*HST*/WFC3) data were consistently combined with the mid-IR data (*Spitzer*/IRAC) and ground-based observations. For each of the CANDELS fields, the photometric catalogs were selected in *HST*/WFC3 F160W band using SExtractor (Bertin and Arnouts 1996). For low-resolution images, Template FITting (TFIT; Laidler et al. 2007) was performed to smooth the high-resolution image to low resolution and fit the best flux consistent with the *HST* data. TFIT creates a template using prior information on the position and light distribution profile of sources in the high-resolution to more robustly measure photometries from the lower resolution data. Table 2.2 lists the limiting magnitudes and survey areas covered for each of the CANDELS fields.

In this study, we use an improved version of the original Dahlen et al. (2013) photometric redshifts. The new catalogs are based on the probability density functions (PDF) measured by six groups using different template-based methods applied on the CANDELS photometric catalogs. The methods are different either in their choice of parameters or the code used. The PDFs from different groups were corrected and optimized for bias (optimal shift) and their variance (optimal scaling of the width of the PDF). After re-

---

[1] http://arcoiris.ucolick.org/candels/

Table 2.1: The observed bands from UV-to-NIR SEDs of galaxies across the five CANDELS fields

Field	**Filter set**
GOODS-S (Guo et al. (2013))	Blanco/CTIO U, VLT/VIMOS U, HST/ACS F435W, F606W, F775W, F814W, F850LP, HST/WFC3 F098M, F105W, F125W, F160W, VLT/ISAAC Ks, VLT/Hawk-I K_s, Spitzer/IRAC 3.6 μm, 4.5 μm, 5.8 μm, 8.0 μm
GOODS-N (Barro et al. (2019))	KPNO U, LBC U, HST/ACS F435W, F606W, F775W, F814W, F850LP, HST/WFC3 F105W, F125W, F140W, F160W, F275W, MOIRCS K, CFHT K_s, Spitzer/IRAC 3.6 μm, 4.5 μm, 5.8 μm, 8.0 μm
UDS (Galametz et al. (2013))	CFHT/MegaCam u, Subaru/Suprime-Cam B, V, R_c, i' , z' , HST/ACS F606W, F814W, HST/WFC3 F125W, F160W, VLT/Hawk-I Y, K_s, WFCAM/UKIRT J, H, K, Spitzer/IRAC 3.6 μm, 4.5 μm, 5.8 μm, 8.0 μm
EGS (Stefanon et al. (2017))	CFHT/MegaCam U^*, g' , r' , i' , z' , HST/ACS F606W, F814W, HST/WFC3 F125W, F140W, F160W, Mayall/NEWFIRM J1, J2, J3, H1, H2, K, CFHT/WIRCAM J, H, K_s, Spitzer/IRAC 3.6 μm, 4.5 μm, 5.8 μm, 8.0 μm
COSMOS (Nayyeri et al. (2017))	CFHT/MegaCam u^*, g^*, r^*, i^*, z^*, Subaru/Suprime-Cam B g^+, V, r^+, i^+, z^+, HST/ACS F606W, F814W, HST/WFC3 F125W, F160W, Subaru/Suprime-cam IA484, IA527, IA624, IA679, IA738, IA767, IB427, IB464, IB505, IB574, IB709, IB827, NB711, NB816, VLT/VISTA Y, J, H, K_s, Mayall/NEWFIRM J1, J2, J3, H1, H2, K, Spitzer/IRAC 3.6 μm, 4.5 μm, 5.8 μm, 8.0 μm

calibration, the PDFs were combined based on the minimum Frechet distance Alt and Godau (1995), which tracks the similarity between any pair of PDF curves, among the independent photo-z PDFs (analogous to the median of a set of numbers). The final catalog consist of the spectroscopic/3D-*HST* grism redshifts and the combined photometric redshifts (Kodra 2019). The point estimate redshifts used in this work show a normalized median absolute

Field	**Area** ($arcmin^2$)	**F160W** 5σ **Limiting Magnitude (AB)**
GOODS-North	170	27.80
GOODS-South	170	27.36
COSMOS	216	27.56
UDS	202	27.45
EGS	206	27.60

Table 2.2: The survey area and WFC3 F160W limiting magnitude for different CANDELS fields

deviation of $\sigma_{\rm NMAD} \sim 0.02$. The stellar masses were measured through SED fitting by keeping redshifts at their best values.

2.3 Selection Methods

Here we use a combination of several near-infrared selection techniques to identify massive quiescent galaxies at a redshift of $3 \leq z \leq 4.5$ in the CANDELS fields. This is the redshift interval that most of the galaxies exist; however, the full sample of galaxies is at a redshift of $2.8 \leq z \leq 5.4$. We define the massive quiescent population as those with $M_s \geq 10^{10} M_\odot$ and $sSFR(z) \leq 0.2/t_H(z)$ Gy^{-1} in which $t_H(z)$ is the age of the universe at redshift z in Gyrs. The results are cross-compared and used to compile a reliable catalog for the study of the number density and mass function of these systems in the redshift range mentioned. This comparison also allows a better understanding of the strengths and shortcomings of each method.

First, we use the color-selection technique based on rest-frame or observed colors for classification. Then, we explore selection methods based on the inferred physical properties of galaxies from their best-fitted model Spectral Energy Distributions (SEDs).

A serious source of uncertainty in these techniques is the choice of the somewhat subjective parameters that could significantly affect the outcome. These include the S/N in the flux in a particular band, color cuts in the color-space occupied by galaxies, flux limits, and parameters used to generate template SEDs. Galaxies that are close to the decision boundaries are most likely to be affected by the choice of these parameters. Therefore, to minimize the effect of these cuts for each technique, we define a likelihood function between zero and one that identifies the likelihood a given galaxy is a real member of the massive quiescent population within our specified redshift range. We then consistently combine all these measures to find a final value associated with individual galaxies to quantify the degree of confidence for each selection method and to compare different galaxies resulted from a particular selection. Then we estimate the uncertainty in number/stellar mass density measurements coming from the selection thresholds. We refer the reader to the Appendix for more details. Following this procedure, we select $5-10\%$ more candidates than using a step-function selection that misses galaxies close to the selection boundary. This difference is significant enough to affect the measurement of their number and mass densities. In the following section, we describe each selection method.

2.3.1 Rest-frame UVJ selection

This method uses rest-frame U, V, and J bands to select the quiescent population, as the rest frame $U-V$ probes the prominent Balmer break (3646) seen in post-starbursts. The $V-J$ color used to break the degeneracy between dusty star-forming and quiescent galaxies. We can classify galaxies using a color cut that separates the quiescent and star-forming regions in the UVJ plane. This method has been developed and extensively used for classifying galaxies in photometric surveys (e.g., Labbé et al. 2005; Wuyts et al. 2007; Williams et al. 2009; Arnouts et al. 2013, Whitaker et al. 2013; Barro et al. 2014; Straatman et al. 2014; Fumagalli et al. 2016, Siudek et al. 2017 Fang et al. 2018). However, measuring UVJ colors for galaxies at high redshifts is challenging, making the selection less reliable. A typical method to infer rest-frame UVJ colors is from the best-fitted model SED. The model SEDs are generally built assuming a τ model star formation history, which also becomes less reliable at high redshifts since the galaxy model needs more time to evolve into the quiescent regions (Merlin et al. 2018) and more generally suffer from other uncertainties associated with SED fits used to infer rest-frame colors. Therefore, the boundaries of the UVJ criteria are modified at different redshift bins (Whitaker et al. 2013). However, this can introduce contamination from the dusty star-forming galaxies, given the similarity of their rest-frame $U - V$ color. Nevertheless, given the overlap in the wavelength coverage, including medium-band data, can ameliorate some of these problems when selecting high-z galaxies (e.g., Spitler et al. 2014).

To measure the rest-frame UVJ colors, we employ the prescription described by Wolf et al. (2003) for the COMBO-17 survey, where the rest-frame colors are estimated from the best fit template SED. Rest-frame UVJ colors can also be measured by interpolating the observed bands which track the redshifted UVJ (e.g., Rudnick et al. 2003, Taylor et al. 2009, Williams et al. 2009). In the latter case, data from many bands outside of the UVJ region of the spectra is left unused while in constraining the best fit model SED, these data are beneficial (Brammer et al. 2011). We investigate the effect of the star-formation histories on the UVJ colors, when analyzing the effects of the photometric uncertainties in Section 2.4.

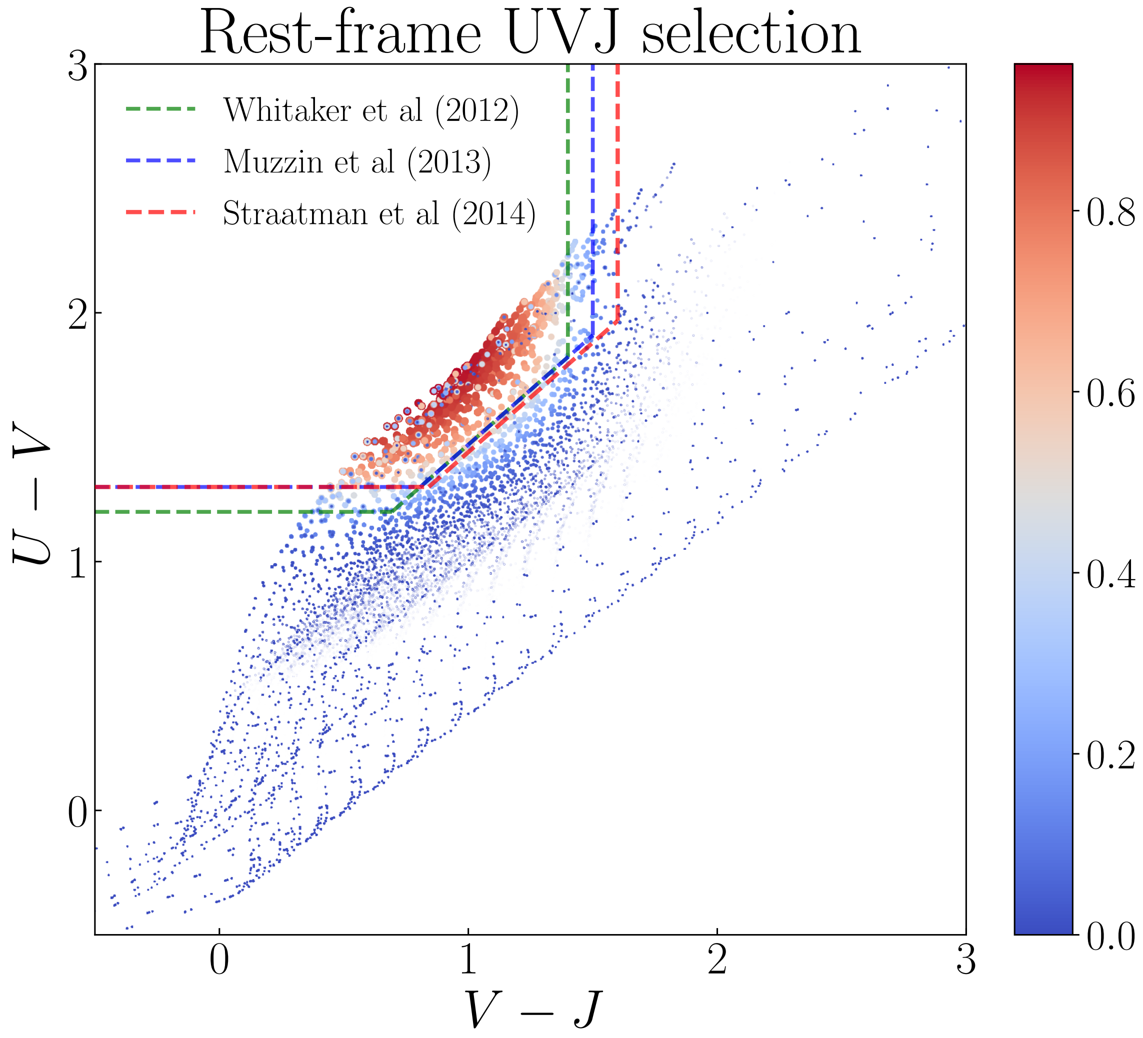


Figure 2.1: The rest-frame UVJ plane and colors from the best-fitted models for every galaxy are shown. The artificial patterns seen are from the limited number of models that do not span the total dynamic range of the UVJ colors of galaxies. Also, since we have imposed another criterion on the F125W-band, we do not select the low-z dusty solution counterparts. The color bar shows the likelihood grading for each selection. Redder data points are more likely to be quiescent.

SPS model	**τ Gyr**	**$E(B-V)$**
BC03 (Bruzual and Charlot (2003a))	(0.01, 30)	(0, 1.1)
IMF	**Dust attenuation law**	**Metallicity**
Chabrier (2003)	Calzetti et al. (2000)	{0.02, 0.008, 0.004}

Table 2.3: The range of the parameter used in the Le Phare code for exponentially declining star formation history, as well as the initial mass function and dust attenuation law used for fitting all the CANDELS.

We combine the selection criteria on the rest-frame UVJ plane employed in Muzzin et al. (2013), Whitaker et al. (2012a), and Straatman et al. (2014) (Figure 2.1). We define the grading scheme by giving a weight to individual galaxies based on their distance from the boundary defined in Table 2.6. (Figure 2.1). The combination of these selections is done by finding the likelihood that each galaxy is a member of the union of selected populations from different criteria. To find the rest-frame UVJ colors of CANDELS galaxies, we use the Le Phare SED fitting code (Arnouts et al. 1999, Ilbert et al. 2006) with standard libraries defined in Table 2.3 (Chartab et al. 2020).

In addition to the criteria on the UVJ plane, we impose a mild detection constraint on the observed J-band ($S/N > 2$), which probes blueward of the Balmer break for the highest redshift bins. This reassures that the break lies redward of the J-band, reducing the contamination from low-z interlopers. The results are presented in Figure 2.1.

2.3.2 Observed-color selection

The most commonly used methods for identifying the population of high redshift galaxies are variations of the drop-out technique, based on the observed colors of galaxies. This uses well-known features in the galaxy SEDs and follows them as they move to redder passbands when galaxy is redshifted. Examples of this are the Lyman break features used for the selection of UV bright Lyman break Galaxies (Steidel and Hamilton 1993; Steidel et al. 1995; Steidel et al. 2003; Reddy et al. 2005; more recently Stark et al. 2010; Bouwens et al. 2011; Bouwens et al. 2014; Roberts-Borsani et al. 2016; Oesch et al. 2016) and evolved systems using Balmer break features (Cimatti et al. 2002; Roche et al. 2002; Franx et al. 2003; van Dokkum et al. 2003; Daddi et al. 2004; Reddy et al. 2005; Mobasher et al. 2005; Lane et al. 2007; Daddi et al. 2007; Wiklind et al. 2007; Rodighiero et al. 2007; Caputi et al. 2012; Nayyeri et al. 2014; Girelli et al. 2019).

This technique uses the fact that magnitudes and colors are sensitive to redshift and the shape of their SEDs, which follows the physical properties of their stellar population and their interstellar medium. For example, for post-starburst galaxies, we can use Balmer break features from 3648 Balmer limit. Therefore, the observed colors can help us find the candidate galaxies directly from photometric measurements using a few photometric bands. For the redshift range of interest here, Balmer break features redshift towards near-IR wavelengths.

We use the observed colors to select candidates in the CANDELS fields with the color selection criteria from Nayyeri et al. (2014) (Figure 2.2). This selection was initially developed by finding the color criteria in the color-color space for a sample of old (low-z) quiescent, dusty starburst, and post starburst populations, using Bruzual and Charlot (2003a) stellar population synthesis evolutionary tracks and considering an IGM absorption following Madau (1995). In the redshift range of interest ($3 \leq z \leq 4.5$), $H - K$ color will constrain the presence of the Balmer break, accompanied by a $Y - J$ or $J - H$ constraint

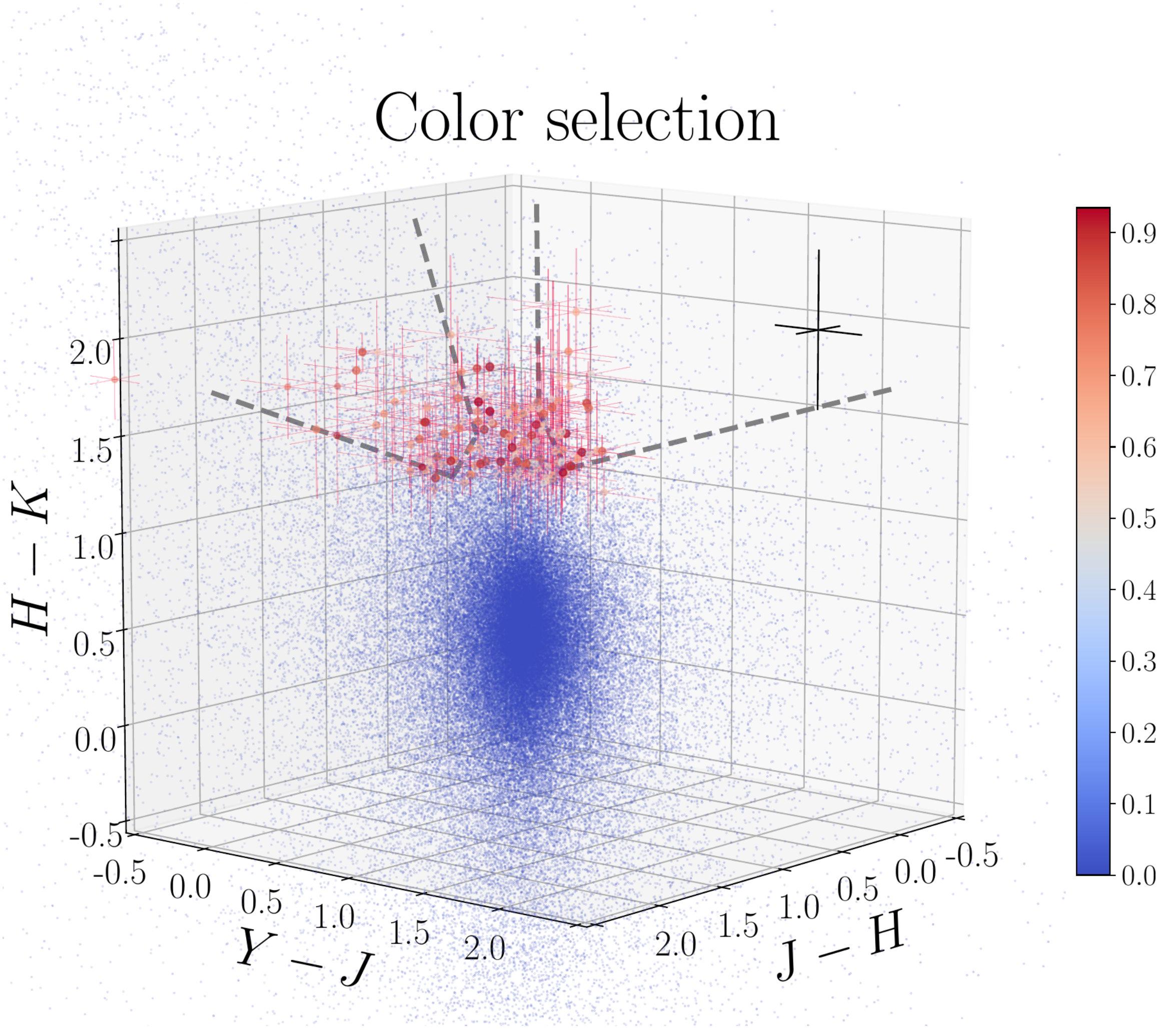


Figure 2.2: Observed color selection based on the criteria in Nayyeri et al. (2014) in 3 dimensions is shown (The gray dashed line shows these selection boundaries on the 2D planes of the colors). The redness of the points indicates the degree to which each galaxy is a member of the quiescent population. The error bars are plotted for sources with a likelihood higher than 0.5.

which can discriminate between dusty-starburst and quiescent galaxies. Also, we impose a non-detection requirement on the U and B bands since they fall blueward of the break. By doing this, we reduce the contamination from the low redshift interlopers.

Another likely source of contamination is due to the absence of nebular emission lines. These affect the selection based on broadband photometry since they can mimic a Balmer break feature, which can mislead the classification scheme (see Nayyeri et al. 2014; Straatman et al. 2014; Merlin et al. 2018). To minimize the effect of these lines, we perform the SED-fitting procedure described in Section 2.3.3 using libraries with and without the nebular line emissions.

We assign likelihoods to the selected sources, based on their proximity to the selection boundaries to reduce the dependence on the chosen selection boundaries (More details are presented in the Appendix 2.7). The final likelihood will be the average likelihood for all of the realizations of a galaxy within its error bars (Section 2.4.2).

2.3.3 Selection based on SED fitting

Here we discuss another method used for finding quiescent candidates based on the inferred physical properties measured from their SEDs. The advantage of this method is that it makes use of all the photometric data available and, therefore, imposes stronger constraints on the selection process.

Furthermore, it predicts the physical parameters for each galaxy. The disadvantage is that the method is model dependent and is based on the star formation history and extinction used to generate SED templates (Grazian et al. 2007; Fontana et al. 2009; Pacifici et al. 2016; Merlin et al. 2018; Carnall et al. 2018; Carnall et al. 2019b; Merlin et al. 2019).

We fit template SEDs for the candidates between $2.8 < z_{phot} < 5.4$ with masses larger than $10^{10} M_{\odot}$ using CANDELS photometric and spectroscopic redshifts (where available), and catalog of the physical properties. We rely only on the mass measurements since the stellar masses are less susceptible to the parameters chosen in the template libraries of galaxies, particularly the star formation history used for SED fitting (compare to the inferred SFRs)(Mobasher et al. 2015; Papovich et al. 2001; Shapley et al. 2001; Wuyts et al. 2009; Muzzin et al. 2009).

However, the presence of the nebular lines can affect stellar mass measurements. Hence, we treat the stellar mass as a free parameter (although we made a prior mass selection on the sub-sample) when fitting the sub-samples photometric measurements to keep the stellar mass uncertainty of this type limited to the initial selection.

We find the physical properties using Bayesian Analysis of Galaxies for Physical Inference and Parameter EStimation (BAGPIPES), a Bayesian SED fitting code (see Carnall et al. 2018). BAGPIPES uses 2016 version of the Bruzual and Charlot 2003a, and MULTINEST (Feroz and Hobson 2008; Feroz et al. 2009) for multimodal nested sampling algorithm (Skilling 2006).

The multimodal nested sampling algorithm employed is a huge improvement over a simple χ^2 fit, which is incapable of producing a reliable error estimate. Moreover, the Markov chain Monte Carlo (MCMC) algorithm employed in some SED-fitting procedures for finding the posterior of the physical properties can be problematic when sampling from multimodal posterior. These include models with large degeneracy, which is often the case when modeling a complex system such as a galaxy under large photometric uncertainties.

We fixed redshifts to their photo-z, and where available, the spec-z values. We built separate model libraries based on three prescriptions for SFHs: exponentially declining, top-hat, and double power-law forms (Behroozi et al. 2013b), assuming Chabrier (2003) IMF, dust attenuation based on Calzetti et al. (2000), and IGM absorption from Inoue et al. (2014) which is a revised version of the Madau (1995) model. Knowing that nebular lines emission can mimic a Balmer break-like feature and for controlling their effects on the physical properties and, consequently, the selection made based on them, we run the code with and without the nebular emission (since the target population is expected to have little to no nebular emissions). The priors used for physical parameters in the SED fitting are listed in Table 2.4. We then find the posterior distribution for each model parameter. Following Carnall et al. (2018), we define ψ_{SFR} as the ratio of the SFR at any given time ($SFR(t)$) to the average SFR over the age of a given galaxy ($< SFR(t) >$):

$$\psi_{SFR} = \frac{SFR(t)}{< SFR > (t)} = \frac{SFR(t)}{\frac{1}{t}\int_0^t SFR(t')dt'} \tag{2.1}$$

and we define the quiescent population as those with ψ_{SFR} less than 0.1 at the observed age of the galaxy. In other words, the quiescent galaxies are those with average SFR over the last 100 Myrs to be less than 10 percent of the average SFR over its lifetime.

The Bayesian nature of the SED fitting code allows us to have the posterior distribution for each galaxy and all the parameters associated with it. We apply the selection on the sample from the posterior and count the number of selected samples from the posterior versus the total count to find the expected value of the selection criteria under the posterior.

Table 2.4: The priors used for different free physical parameters and the fixed parameters used in the fit.

Star Formation History	**Free Parameter**	**Prior**	**Limits**	**Fixed Parameter**	**Value**
	A_V[1]	Uniform	(0, 2)	SPS models	BC03
Double Power Law:	$\log_{10}(M_{\rm formed}\ /\ \mathrm{M}_\odot)$[2]	Uniform	(1, 13)	IMF	Chabrier (2003)
	$\mathcal{Z}\ /\ \mathcal{Z}_\odot$[3]	Uniform	(0.2, 2.5)	$z_{\rm obs}$	$z_{\rm phot/spec}$
$SFR(t) = C[(t/\tau)^\alpha + (t/\tau)^{-\beta}]^{-1}$	τ / Gyr[4]	Uniform	(0, $t(z_{\rm obs})$)	$\log_{10}(U)$[5]	-3
	α[6]	Logarithmic	$(10^{-2}, 10^3)$		
	β[7]	Logarithmic	$(10^{-2}, 10^3)$		
	A_V	Uniform	(0, 2)	SPS models	BC03
Exponentially Declining:	$\log_{10}(M_{\rm formed}\ /\ \mathrm{M}_\odot)$	Uniform	(1, 13)	IMF	Chabrier (2003)
	$\mathcal{Z}\ /\ \mathcal{Z}_\odot$	Uniform	(0.2, 2.5)	$z_{\rm obs}$	$z_{\rm phot/spec}$
$SFR(t) = C\exp\left(-t/\tau\right)$	τ / Gyr[8]	Uniform	(0.05, 10)	$\log_{10}(U)$	-3
	Age	Uniform	(0, $t(z_{\rm obs})$)		
	A_V	Uniform	(0, 2)	SPS models	BC03[11]
Top-Hat:	$\log_{10}(M_{\rm formed}\ /\ \mathrm{M}_\odot)$	Uniform	(1, 13)	IMF	Chabrier (2003)
	$\mathcal{Z}\ /\ \mathcal{Z}_\odot$	Uniform	(0.2, 2.5)	$z_{\rm obs}$	$z_{\rm phot/spec}$
$SFR(t) = \begin{cases} C & t \in [\mathrm{Age_{min}}, \mathrm{Age_{max}}] \\ 0 & \text{otherwise} \end{cases}$	$\mathrm{Age_{min}}$[9]	Uniform	(0, $t(z_{\rm obs})$)	$\log_{10}(U)$	-3
	$\mathrm{Age_{max}}$[10]	Uniform	(0, $t(z_{\rm obs})$)		

[1] A_v is the Attenuation at 5500Å

[2] M_{formed} is the mass formed

[3] $\mathcal{Z}$ is the metallicity

[4] τ is the peak time for double power law SFH

[5] $\log_{10} U$ is the ionization parameter

[6] α is the rising power

[7] β is the falling power

[8] τ is the exponential decay timescale in τ model SFH

[9] Age_{min} is the initial time for Top-hat SFH

[10] Age_{max} is the final time for the Top-hat SFH

C is the normalization constant

[11] BC03 is the Stellar population synthesis at the resolution of 2003 (Bruzual and Charlot (2003b))

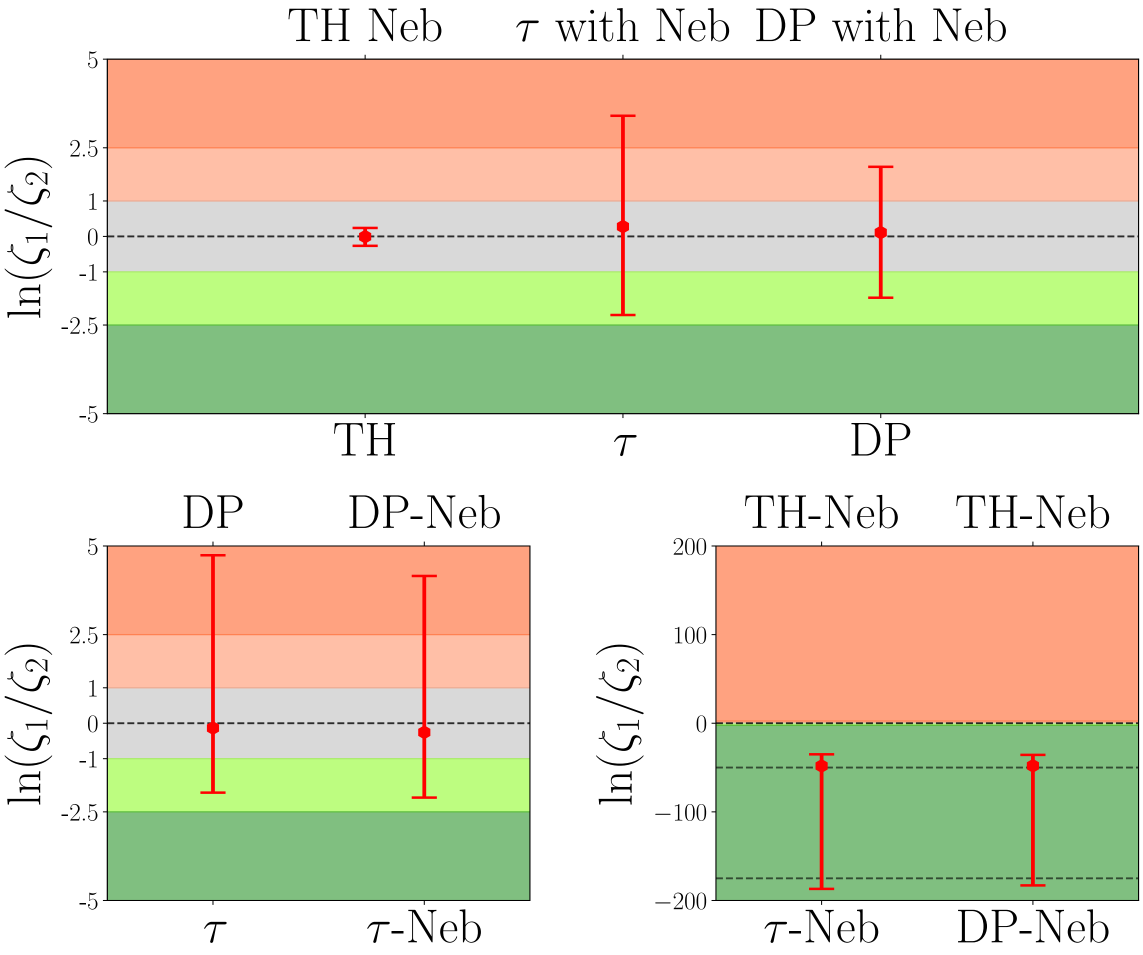


Figure 2.3: Shows the Bayes factor for models with different SFHs and with/without nebular emissions for the redshift and mass-selected sample of galaxies. TH, τ, and DP stand for Top-Hat, exponentially declining, and double power-law model for SFH, respectively. The top panel shows the Bayes factor for models that include nebular emission to those with the same SFH but without emissions. The bottom left shows the relative evidence for Top-hat SFH to τ models and double power law. The bottom right shows the relative evidence for τ models and double power law. The data points show the median, and the error bars are 20, and 80 percentile of the sample. ζ_1 and ζ_2 are evidence for top and bottom models (in each figure). The grey area shows Inconclusive to Weak evidence, the lighter shade shows Weak to Moderate evidence, and darker shade shows Moderate to Strong evidence (Jeffreys 1961). In the case of top hat models, there is strong evidence against the top-hat models when compared with τ and double power law.

We use the Bayesian evidence (marginal likelihood) on different models used for SED-fitting. By doing so, we find the relative evidence for different models according to the data. The definition of the Bayesian evidence is:

$$\zeta = p(D|H) = \int p(D|\theta, H) p(\theta|H) d\theta \tag{2.2}$$

where D, H, and θ represent the data, hypothesis (model), and model parameters, respectively. The $p(D|H)$ is the probability of getting the data D under the assumption of the validity of a specific hypothesis H, which is the Bayesian evidence for H.

We adopt Jefferey's criteria (Jeffreys 1961; also see Kass and Raftery 1995) for interpretation of the relative evidence (Bayes factor) calculated for each galaxy and different models. We use the same prior for fixed parameters such as redshift, and for the rest of the parameters, we use non-informative priors (Kass and Wasserman 1996) across different models. By doing so, the Bayes factor represents the posterior odds of the models ($\frac{p(H_2|D)}{p(H_1|D)} = \frac{p(D|H_1)}{p(D|H_2)}$). In other words, we impose no prior judgment about the validity of a certain model. Results presented in Figure (2.3) express that statistically, none of the models show substantially more evidence. However, the Bayes factors support models that include nebular lines and generally favor double power-law over τ-models. This is consistent with Reddy et al. (2012) regarding the inability of the τ models in reproducing star formation rates from UV+IR measurements in the star-forming population at higher redshifts. Pacifici et al. (2016) showed that double power-law is the best model to describe the SFHs of their samples. Carnall et al. (2019a) also confirmed that a double power-law model could produce relatively stronger evidence. As Figure 2.3 depicts, similar to Belli et al. (2019) investigation of the effect of SFH on age measurements, top-hat models generally fail to produce a comparable level of evidence from data. Therefore, we do not use the results from the Top-Hat models for making a selection.

2.3.4 AGN Contamination

We cross-matched the potential candidates for high redshift massive evolved systems with the publicly available Chandra X-ray catalogs and used Spitzer MIPS 24 μm to detect possible dusty AGNs. We excluded any candidates with a counterpart (shows detection) in either X-ray or MIPS. For the X-ray catalogs we looked for any counterparts closer than 5 arcsec which is 10 times the resolution of the Chandra X-ray Observatory (cross-matched with Laird et al. 2009; Evans et al. 2010; Xue et al. 2016; Nandra et al. 2015; Cappelluti et al. 2016; Civano et al. 2016; Masini et al. 2018) and for Mid-IR we used the catalog published in Barro et al. (2019) for finding any counterpart and flux measurements in the mentioned MIR bands. About 30% of the candidates have an X-ray or MIR counterpart, which were removed from the final sample and not included in the number and stellar mass density measurements.

2.3.5 Final sample

To each galaxy, we assigned a likelihood measure that is the median of the likelihoods estimated for that galaxy from each of the three techniques: the UVJ selection (Section 2.3.1); Observed color selection (Section 2.3.2) and SED fitting method with and without the nebular emission contribution (Section 2.3.3). We combine the results from the SED fitting under different assumptions using the weighted average of the likelihoods for each galaxy using the marginal likelihood calculated for a particular model as their corresponding weights. By doing this, we make sure that we have put more emphasis on likelihoods calculated from the models with higher marginal likelihoods. Then we take the median of all these methods as our final indicator. We select the final sample as those with the median likelihood higher than 0.5. However, we use the 0.3 and 0.7 criteria as our least and most conservative samples, respectively. By using the median indicator, we limit our sample to those galaxies that were assigned at least by two of the methods mentioned to have a high likelihood of being a massive evolved galaxy. Figure 2.8 shows how this threshold changes the number density measurements of each selection method as well as the median value which is taken as the final indicator. The final selected galaxies with their assigned likelihoods and estimated physical parameters are listed in Table 2.7. The galaxies that were selected with the median likelihood higher than 0.3, constitute the most inclusive sample and the one used for finding the upper bounds on the number/mass densities. Also, we control the possible contamination in our final sample, since, we rely on the composite indicator compared to a single measure. Figure 2.4 shows a comparison between different selection methods and how they relate to each other (All the highlighted points are galaxies in our final sample). For example, the selection in the sSFR vs. M_s plane is generally consistent across different methods since the objects with higher BBG/UVJ likelihood are close or inside the selection boundary. For selection in the UVJ colors, we have several candidates that are far from the boundary but show much higher BBG/SED likelihood. This shows the sensitivity of final results to the libraries used, and by changing the SED fitting code and/or the libraries used, we are not necessarily searching through the same part of the models' parameters space. In terms of the selection based on the observed colors, we tend to have higher UVJ/BBG likelihood, but there are a couple of candidates that show higher UVJ/SED likelihoods but fall outside of the criteria. This shows the importance of using information from other bands as well.

2.4 Photometric Uncertainty

All the methods discussed in the previous section are sensitive to uncertainties in photometry at different levels. To quantify the sensitivity of each of the selection methods to photometric uncertainties, we use the Monte Carlo resampling and produce different realizations of each galaxy's photometries. We quantify the confidence level for each candidate galaxy selected from different methods, and the error estimates that propagated into the selections.

Assuming a Gaussian error distribution for each band, we produce realizations for each galaxy by perturbing the photometric measurements within their assigned uncertainties. In other words, we build a bootstrap resampled ensemble of photometric measurement distribution of a galaxy. Then we perform the methods for selecting candidate galaxies and study how these perturbations affect our selections and how sensitive the selected sample is to changes in photometric measurements within their uncertainties. In the case of the selection based on the observed color technique, we only look at several bands, so the uncertainties on the other bands (not used in the selection) are not crucial, but in the case of the UVJ and SED we use all the photometric data which means that the uncertainties in every band can potentially be imperative.

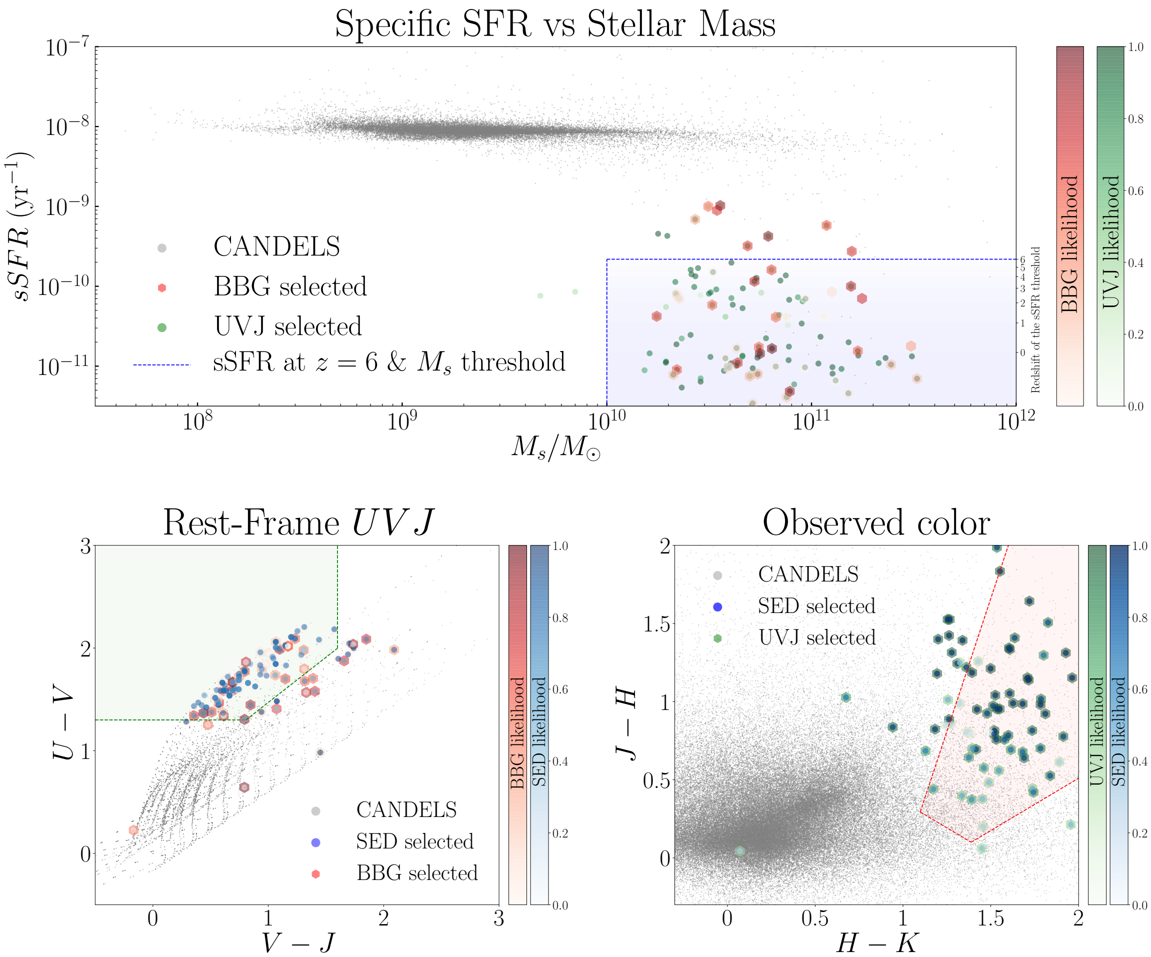


Figure 2.4: The top panel shows the $sSFR$ vs M_s, (from Bagpipes) the bottom left shows the rest-frame UVJ colors (from LePhare), and the bottom right shows the observed colors for the candidate galaxies. The grey points are the CANDELS galaxies (with $2.8 \leq z \leq 5.4$), and the candidates are color-coded with three assigned likelihoods. The green, blue, and red color scales reflect the likelihood measured based on UVJ, SED, and the observed colors, respectively. In order to show the contrast in different selections, we plotted likelihoods on top of each other. So those that have two colors show high likelihood in two methods, and from those, the ones that also fall inside the respective criteria show a higher likelihood in all of the mentioned methods. The shaded area shows the selection in the corresponding plane with the same color-code as mentioned. In the top panel, different limiting sSFR for SED based selection is shown at different redshifts.

2.4.1 Uncertainty of selection based on UVJ

Here we consider the effect of the photometric uncertainties in finding UVJ colors and how that could affect the final selected sample. We first choose a subset of the selected galaxies, construct 50 realizations of each, and run SED fitting on the resampled galaxies. Then we quantify how these uncertainties propagate into our inferred rest-frame colors and how they change the result from the UVJ selection in Section 2.3.1.

Figure 2.5 shows the effect of photometric uncertainty on the UVJ selection. It is clear how the choice of the SFH and whether to include the nebular emissions are pivotal in selecting the quiescent population. We find the UVJ colors from the Bagpipes SED fitting code since we are interested in understanding the effect of the photometric uncertainty on the inference from the SED, which makes a Bayesian posterior estimate more reliable given the clear uncertainty definition based on the posterior. For all but a small subset of galaxies fitted with τ models and nebular lines, the posterior UVJ of the resampled galaxies falls around the posterior from the true photometry with considerable scatter. For a small subset of galaxies (5 out of 28) fitted with τ models, the locus of the posterior varies dramatically, and with higher scatter than the posterior UVJ from the true photometry. The double power-law model is found to be more immune to the photometric uncertainties, as is also the case for physical parameters used in SED based selection.

2.4.2 Uncertainty of selection based on observed colors

To find the effect of photometric uncertainties on the selection based on the observed colors, we generate 10^5 realization for every galaxy and assign a likelihood based on

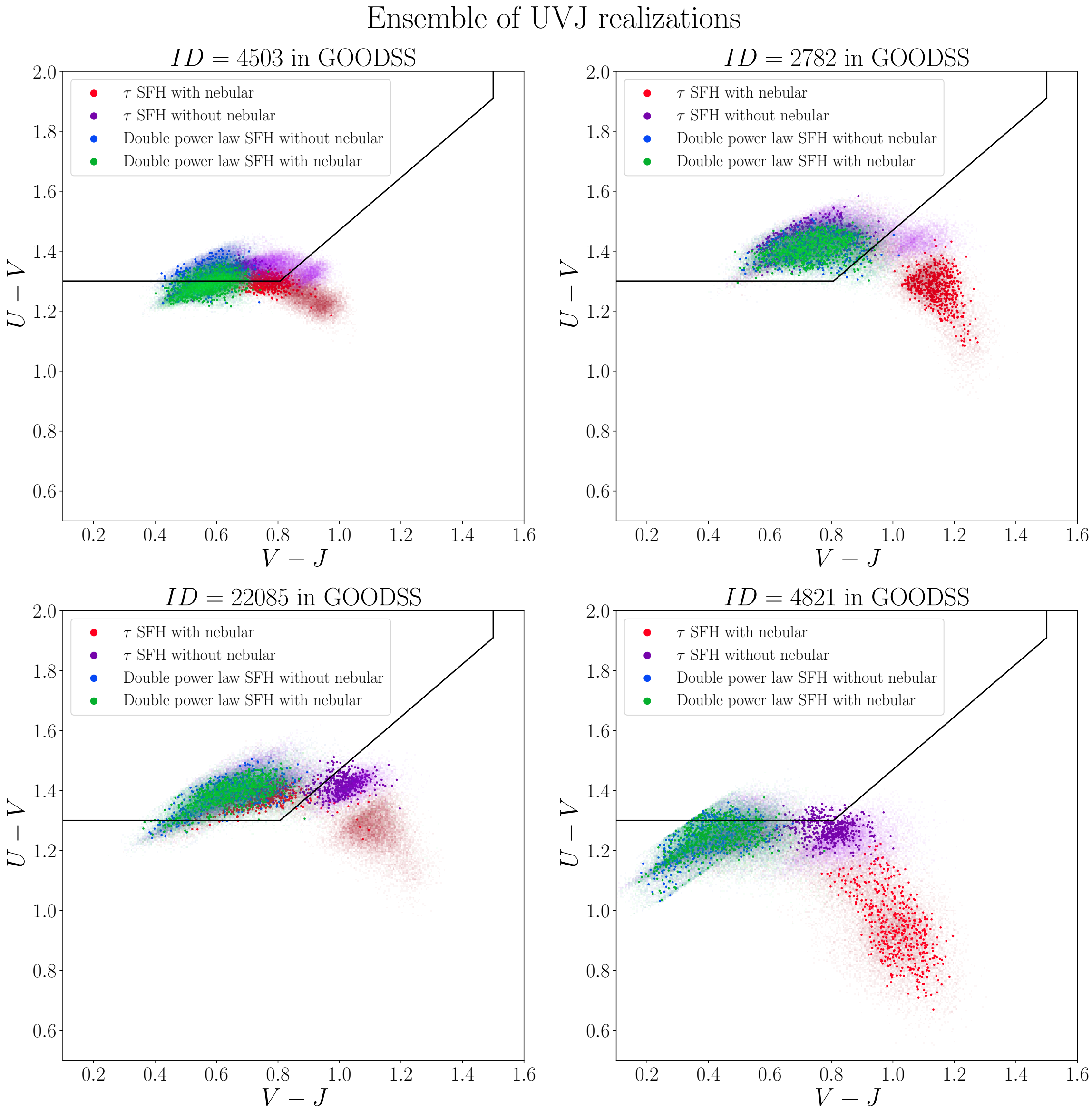


Figure 2.5: The UVJ plane for galaxies resampled ensemble (Monte Carlo simulated photometry). The UVJ colors from the unperturbed photometry are drawn with larger points, and the dimmer points represent the posterior UVJ color for the disturbed photometries on the UVJ color plane. This is a zoomed-in version of the UVJ plot showing the separation of different models. This figure shows how the change in the model assumptions (i.e., star formation histories, including nebular emission) can change the UVJ colors posterior distribution, which can affect our selection based on the UVJ colors.

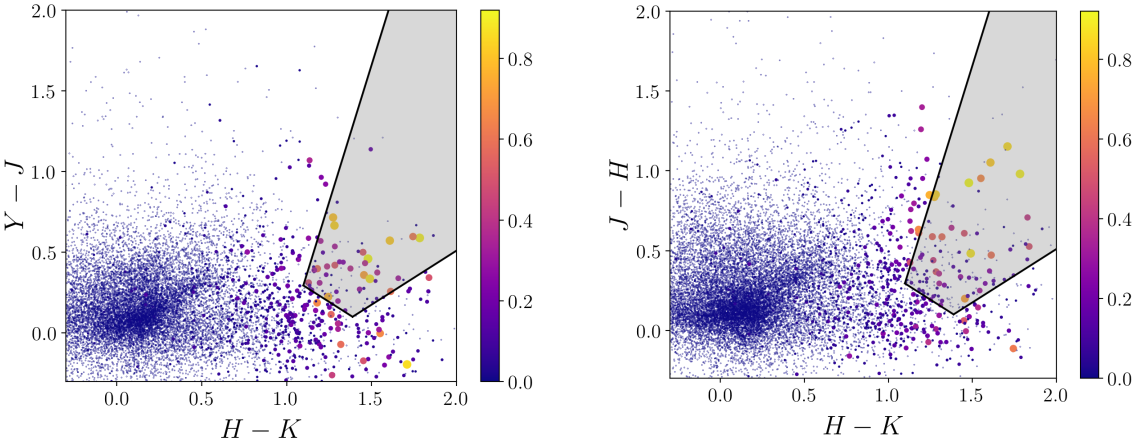


Figure 2.6: The figure shows $J - H$ vs. $H - K$, and $Y - J$ vs. $H - K$ plane of CANDELS in the GOODS-South field. The larger and more yellow data points show more chances of being quiescent given the photometry measurements and their uncertainties. Here we show just the effect of the photometric uncertainty.

the number of realizations that fall into the hard selection criteria to the total number of realizations. The results for the GOODS-S galaxies are shown in Figure 2.6, where they are colored and resized based on the assigned likelihood. Few candidates fall outside of the selection criteria, but when we include the photometric uncertainties, we find realizations that fall in the selected region. On the other hand, few galaxies fall inside the criteria, but because of their photometric uncertainties do not show a clear indication of being quiescent.

2.4.3 Uncertainty of SED fitting based selections

With having a posterior distribution of every physical parameter, we can estimate the uncertainties of these parameters. We use the resampling method mentioned above to get even more reliable error estimates. (more details on the effect of the sampling error in Higson2018). Here we used the same sample from the SED fitting results for finding the sensitivity of the UVJ colors on photometric uncertainties.

We follow the effect on the posterior ψ_{SFR}, which is the critical parameter in the SED based selection discussed in Section 2.3.3. As Figure 2.7 shows, for most of the sample used, the τ-model SFH is more susceptible to photometric uncertainties than double power-law, which has more stable results under photometric perturbations similar to the UVJ colors. Also, including the models' nebular emission leads to models with UVJ colors that are further away from the quiescent region and hence reduce the number densities of the quiescent population.

In this section, we found the effect of the photometric uncertainties on each method's assigned likelihood. As mentioned in Section 2.3.2, for finding the likelihood based on the observed colors for all the galaxies in the CANDELS fields, we assign a likelihood to every photometric realization of a galaxy based on their proximity to the selection boundary, then we take the average of these as the observed color likelihood for that galaxy. By doing this, we can take into account both the photometric uncertainties and closeness to the selection boundary. Since we only calculated the effect of the photometric uncertainties on likelihoods based on the UVJ and SED based selection for a sub-sample of galaxies, we did not incorporate the photometric uncertainties in the final assigned likelihood of these methods.

2.5 Number and Stellar Mass Densities

Using the final sample of massive and evolved galaxies, as presented in Table 2.7, we now estimate their number and mass density across the CANDELS fields. We divide the candidates based on their redshifts into two bins $2.8 \leq z < 4.0$ and $4.0 \leq z < 5.4$, approximately representing the same comoving volumes. To find the most reliable candi-

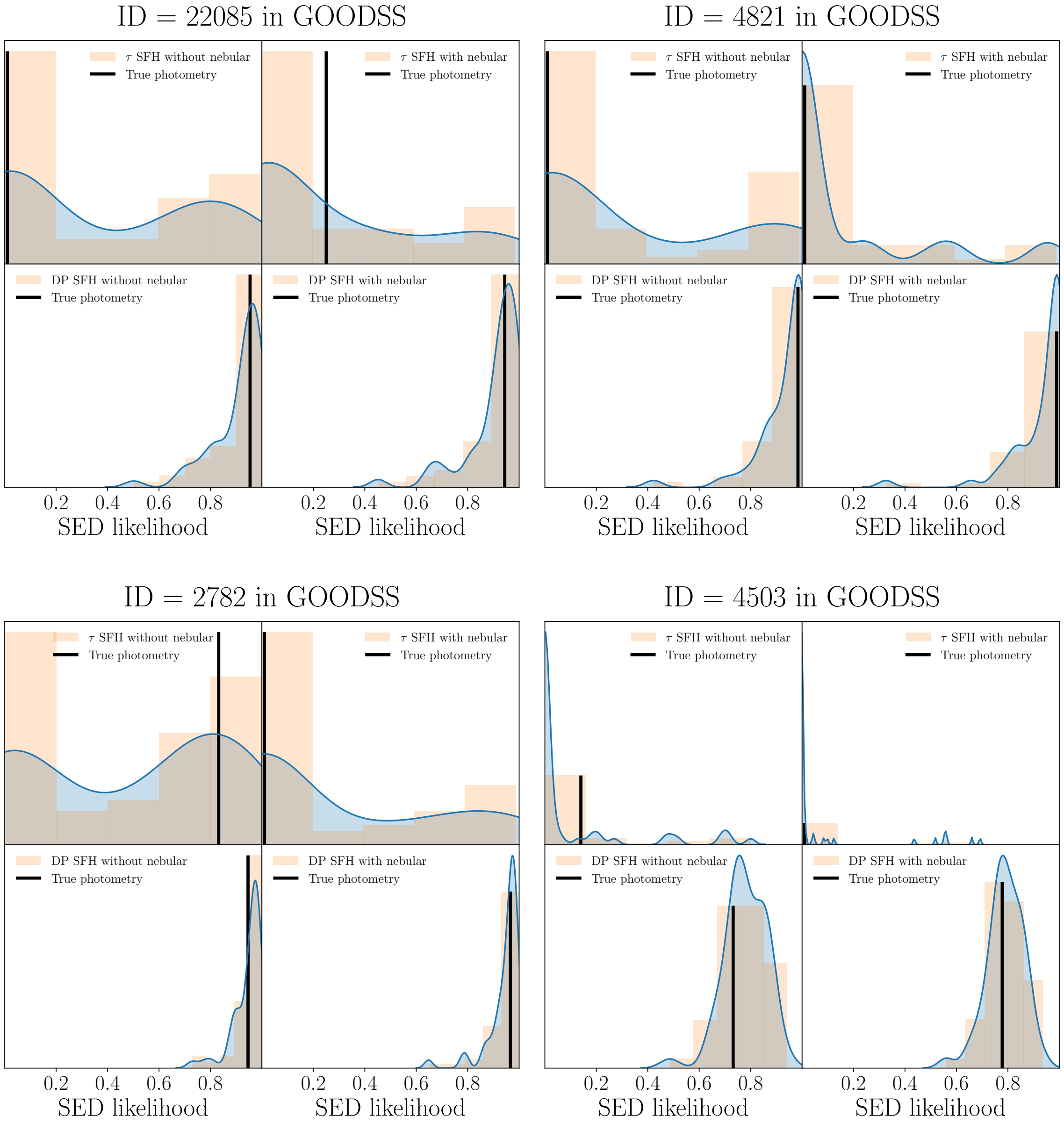


Figure 2.7: Here we show the dependence of the likelihood assigned in the SED selection on photometric uncertainties. ψ_{SFR} likelihood for several galaxies and their bootstrapped photometries assuming Gaussian uncertainties. For each galaxy, the top left and right is τ model without and with nebular emission respectively, and the bottom left and right is double power-law without and with nebular emission, respectively. The histograms show bootstrapped sample photometries, and the curve is their simple kernel density estimate. The black line shows the quiescent likelihood from SED from true photometry. The likelihood estimate from τ models show to be more sensitive to possible variations under photometric uncertainties.

dates, we need to assign a threshold for our likelihood. However, in order to follow this threshold's effect, we impose three different values on the final likelihoods. These thresholds are shown as error bars in Figure 2.9. The upper limits are for galaxies with an assigned likelihood of more than 0.3, the lower limits are for those with likelihood more than 0.7, and the measurement points are for those more than 0.5. The variation caused by changing the selection threshold is larger than the typical Poisson noise. However, we take into account the Poisson noise as an independent source of uncertainty (Figure 2.8). We find the number/stellar mass densities after taking into account the incompleteness of our sample, as discussed in the next section.

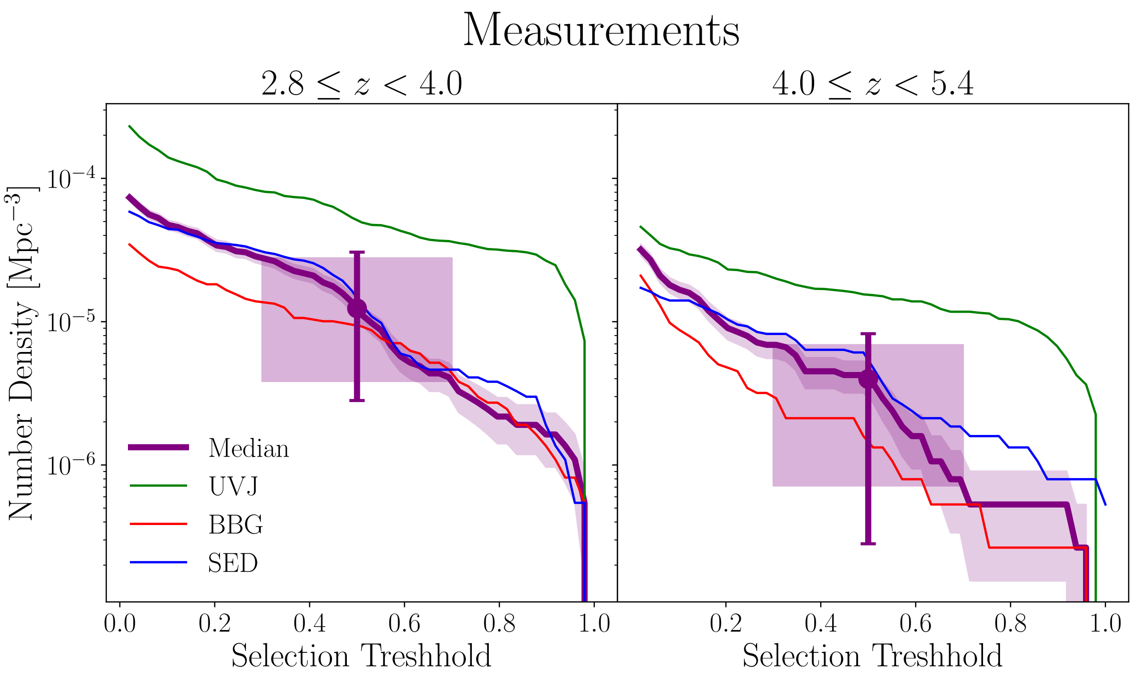


Figure 2.8: The comoving number densities for different selection methods and the median indicator used for the selection of the final sample are plotted. The shaded area around the median line shows the 1 σ Poisson noise at a given threshold. We define the upper bound as the number density at threshold 0.3 and the lower bound as 0.7, including the Poisson noise. The error bars show the uncertainty as upper and lower bounds. The shaded rectangular correspond to variations in the number densities due to variation in the selection threshold ($[0.3, 0.7]$).

2.5.1 Completeness

To estimate the completeness of the sample used, we follow the prescription adopted in Pozzetti et al. (2010). We divide the CANDELS galaxies into two redshift bins mentioned above. We find the 20% of the galaxies with the faintest apparent magnitude in H Band, with specific star formation rates lower than 75% of galaxies in that redshift bin. We calculate the M_{lim} defined as the mass which the galaxy would have if it were at the survey limiting magnitude ($H_{lim} = 26$) (assuming a constant mass-to-light ratios); in other words, we find $\log(M_{lim}/M_*) = 0.4(H - H_{lim})$. Then we define the fraction of the sample with $\log(M_{lim}/M_\odot) > 10$ as the completeness of the quiescent galaxies with that mass range. The effective volume of the survey at each redshift bin is defined to be $V_{eff} = V_{co}\eta(z)$; where V_{co} is the comoving volume at each redshift bin and $\eta(z)$ is the completeness of the sample for $\log(M_*/M_\odot) > 10$ at each redshift bin of the quiescent population.

2.5.2 Comparison to models predictions

In this section, we find the target population predicted from the several existing models listed below and compare their number and stellar mass densities with the sample found in this study.

- **Simulated Infrared Dusty Extragalactic Sky:** We use the simulated catalog presented in Béthermin et al. (2017), which uses the abundance matching technique for occupying dark matter halos from the Bolshoi-Planck simulation (Klypin et al. 2016; Rodríguez-Puebla et al. 2016) and uses an updated version of the 2SFM (two star-formation modes) galaxy evolution model (Sargent et al. 2012; Béthermin et al. 2012 from which a lightcone covering 2 degrees squared was produced.

- **Universe Machine:** We employ the available CANDELS lightcones produced by the Universe Machine semi-empirical model, which uses the same Bolshoi-Planck simulation for halo properties and mass assembly histories. There are eight realizations for each field in CANDELS, and we use all of these realizations. (Behroozi et al. 2019)

- **EAGLE Simulation:** We use the EAGLE cosmological hydro-dynamical simulations of a box with $L = 50$ and 100 co-moving mega-parsec (cMpc) on each side. We use RefL0050N0752, RefL0100N1504, and AGNdT9L0050N0752 in which the first two are the reference physical model used and the last one has higher AGN heating temperature and lower sub-grid black hole accretion disc viscosity. All of these simulations have the same mass resolutions. (Schaye et al. 2015; Crain et al. 2015; McAlpine et al. 2016; The EAGLE team 2017)

- **IllustrisTNG Simulation:** We use the latest IllustrisTNG cosmological hydrodynamical simulations of a box with $L = 100$ and 300 co-moving mega-parsec (cMpc) on each side. There are three resolutions for each simulation box, and here we used the highest resolution. (Nelson et al. 2018; Springel et al. 2018; Pillepich et al. 2018; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2019)

To identify the quiescent galaxies in the simulations and to be consistent with observation of the massive evolved galaxies at $z > 2$ which has been known to be quite compact $R_e \leq 2.5$ kpc (Daddi et al. 2005; Trujillo et al. 2006; Trujillo et al. 2007; van Dokkum et al. 2008; Newman et al. 2010 Damjanov et al. 2011; van der Wel et al. 2014), we use the stellar mass and star formation rates within the twice of half-mass radius. Donnari et al. (2019) has shown that the star-formation main sequence in the Illustris-TNG 100 for $z > 2$ is quite identical when assuming physical properties within an aperture of 5kpc compare to the twice the half-mass radius definition for both $10, 1000$ Myr timescale of SFR measurements. Also, following Merlin et al. (2019) and staying consistent in our comparison within simulations, and since the physical properties of the Eagle simulations are available within certain apertures in the range 1 to 100 kpc, we use the physical properties within an aperture that is closest to four times the half-mass radius. For all the models we select

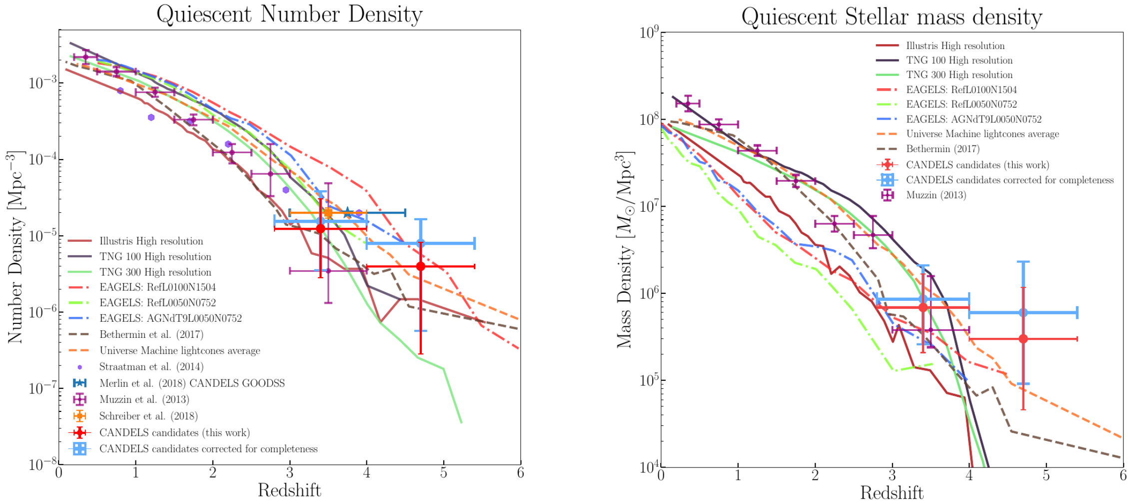
Redshift bin	**Completeness**	**Number Density** (Mpc^{-3})	**Upper bound** (Mpc^{-3})	**Lower bound** (Mpc^{-3})
$2.8 \leq z < 4.0$	0.8	1.2×10^{-5}	3.1×10^{-5}	2.8×10^{-6}
$4.0 \leq z < 5.4$	0.5	4.1×10^{-6}	8.2×10^{-6}	2.7×10^{-7}
$2.8 \leq z < 5.4$	0.7	8.1×10^{-6}	1.9×10^{-5}	1.7×10^{-6}

Table 2.5: The number density measurements and calculated upper and lower bounds (before completeness correction) used in the Figure 2.9 and 2.8. Lower bound and upper bound correspond to the selection threshold of 0.7 and 0.3 respectively. Poisson errors are included in the bounds.

galaxies with stellar masses $\log(M_*/M_\odot) > 10$, and specific star-formation rate within the half-mass radius lower than the evolving sSFR constraint employed in the Pacifici et al. (2016), as $sSFR_{lim} = 0.2/t_U(z)$, where $t_U(z)$ in Gyrs is the age of the universe at redshift z. Carnall et al. (2019a) showed this to be consistent with the ψ_{SFR} measure used in this work across different redshifts.

Figure 2.9 shows the comparison between the observed number density of the quiescent galaxies within the corresponding comoving volumes, corrected for their completeness. The observed and predicted counts from the numerical simulations and empirical models of galaxy evolution are in generally good agreement. However, at the higher-end of the redshift distribution, there are more observed candidates than predicted by models by a factor of ∼ 5. Nevertheless, it is somewhat consistent with the uncertainties shown in Figure 2.9 (especially in the case of Universe Machine). Additionally, the presence of possible contaminants in the sample can ameliorate the tension.

Figure 2.9 further shows the stellar mass density for our sample compared to the simulations and empirical models. We used the mass catalog from CANDELS to measure the stellar mass density. The tension between the stellar mass density for our sample with



(a) The comoving number density

(b) The comoving stellar mass density

Figure 2.9: The comoving number and stellar mass density of the massive quiescent galaxies (defined as those with $M_s \geq 10^{10} M_\odot$ and $sSFR \leq 0.2/t_H(z)$ Gyr^{-1} in which the t_H is the age of the universe at the redshift of z) in two redshift bins from this work are overplotted as red data points on those from the Illustris TNG, and EAGLE simulations for different volumes are shown. Measurements from Behroozi et al. (2019) and Béthermin et al. (2017) is shown with the orange and deep brown dashed line respectively. Measurements from Muzzin et al. (2013), Straatman et al. (2014), Schreiber et al. (2018a), and Merlin et al. (2018) are shown as purple, violet, orange, and blue data points, respectively. The error bars show one σ uncertainty in measurements, except for the error bars of this work, which are upper and lower limits of the measurements, taking into account the dependence of the measurement on artificial selection thresholds and the one σ Poisson uncertainty.

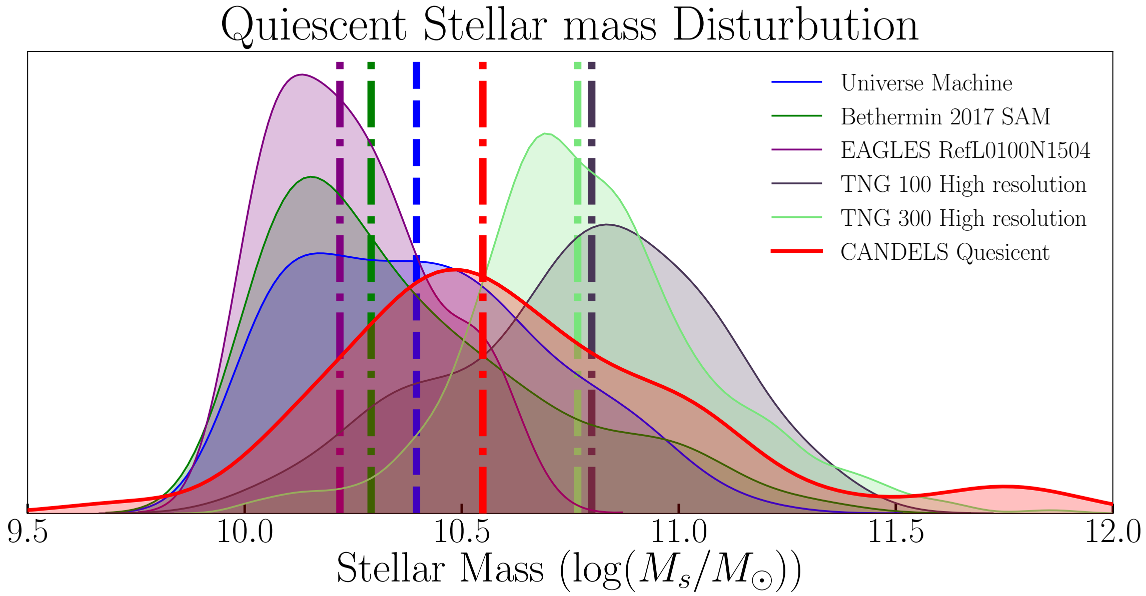


Figure 2.10: Shows the Stellar Mass distribution of the candidates found here compared with semi-analytical models and hydro-dynamical simulations. The y-axis shows the kernel density estimate of the relative frequency. (Using Gaussian kernel and Scott's rule for bandwidth selection)

the simulations seems to be worse; however, the argument above for the number density applies here as well. The stellar mass measurements from the SED fitting results when including nebular line emission can also mitigate the tension as the presence of the nebular lines can reduce the contribution from the continuum part of the SED, which can result in lower inferred stellar masses.

Also, by looking at the mass distribution of the selected samples from models and our CANDELS candidates (Figure 2.10) shows that the CANDELS candidates are relatively more massive than their counterparts in models except for those from the TNG simulations.

The TNG100-1 sample shows to be more massive and using the fact that the stellar mass density at the highest redshift bins is lower than what we found here, we argue that the sample galaxies from the TNG tend to build their stellar mass later and at a higher

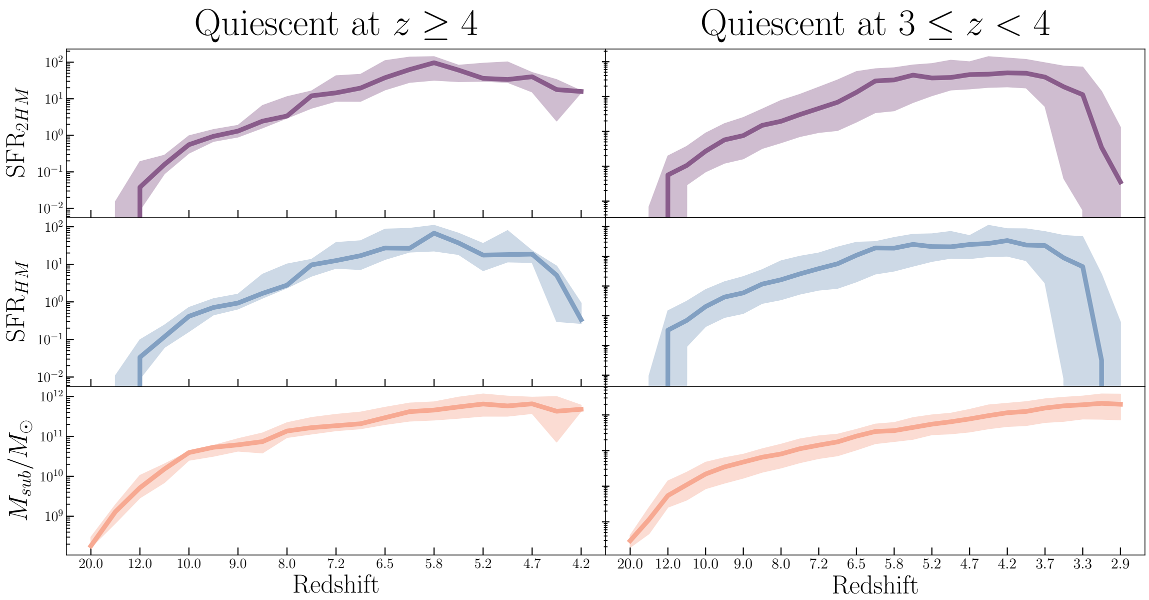


Figure 2.11: Shows the history of total subhalo mass assembly, star formation rate within the half-mass radius (SFR_{HM}), and within twice the half-mass radius (SFR_{2HM}) from the IllustrisTNG100-1 simulation merger trees and following the most massive progenitor branch for the selected sample of quiescent galaxies based on the stellar mass and SFR measurements defined within the half-mass radius. The solid lines show the median history with the shaded area corresponding to the 20% and 80% percentiles. As the figure shows, the higher redshift bin candidates show that the SFR within the half-mass radius is decreasing more rapidly than the star-formation within twice the half-mass radius as they evolve.

rate than what is suggested from the CANDELS observations similar to what Fontana et al. (2009) has found.

Also, by looking at the merger tree information of the selected sample (Rodriguez-Gomez et al. 2015) and following their history back to their formation epoch, we find the mass assembly and star formation histories of the selected sample to check whether it is consistent with their selection as quiescent galaxies. Figure 2.11 shows the median and (20%, 80%) percentile of the mass assembly and star formation rate history of the sample selected. Star formation histories were calculated from two SFR measures, one within the half-mass radius and one within twice the half-mass radius. Figure 2.11 reveals that the

TNG sample at the lowest redshift bin is consistent with quenching even within the twice of the half-mass radius. Although the higher redshift sample shows evidence of quenching within the half-mass radius, the star formation seems to continue outside of the twice of the half-mass radius.

2.6 Discussion and Conclusion

In this study, we used three methods of selecting quiescent galaxies at high redshift. Using the grading scheme explained in Section 2.3, we minimized the effect of selection criteria by assigning a likelihood measure to each galaxy under each selection method. Figure 2.8 shows the number density measurements from different methods and the composite indicator (median) we used in the final selection and their dependence on the selection thresholds. We also investigated the effect of the photometric uncertainties to understand the degree to which the result represented here could be affected. We combined these likelihoods by taking the median across different methods. The results from this study are consistent with similar investigations in CANDELS fields. Nayyeri et al. (2014) found 16 candidates in the GOODS-South of which seven are the same as what we found as candidates. Also, Merlin et al. (2019) found 102 candidates across all the CANDELS, with 36 being the same as those found here. More than half of the proposed candidates that are missing from our sample but reported in Merlin et al. (2019) showed more substantial Bayesian evidence when they were fitted with τ SFH, and double power-law SFH compared to the constant star-formation histories employed in Merlin et al. (2019). This indicates the selection is sensitive to the form of the SFH assumed.

We do not expect a significant difference between the two methods since UVJ colors are expected to follow the sSFR of the model galaxies. However, they react differently to somewhat arbitrary criteria either on the UVJ plane or on the $sSFR$. Moreover, we showed the classification based on the UVJ colors and SEDs to be strongly dependent on the choice of the star formation history and, in the case of τ model library, are particularly susceptible to photometric uncertainties.

We found no strong evidence in favor of any star-formation history, similar to what found in Carnall et al. (2019a). However, double power-law showed more substantial evidence, with less sensitivity to photometric uncertainties. Also, models that included nebular emissions showed stronger evidence.

However, in the grading scheme defined here, we weighted each galaxy's likelihood by its respective evidence so that different models contribute to the likelihoods based on their relative Bayesian evidence.

The selection based on observed colors only uses a few passbands and does not require a search through the parameter space to find the best-fitted model. Therefore, it is computationally more feasible than SED or UVJ based selections when a measure of the sensitivity to photometric uncertainties is needed. Hemmati et al. (2019b) has proposed a faster method of measuring the physical properties of galaxies and their uncertainties using Kohonen's Self-Organizing Map (SOM) (Kohonen 1982). SOM is a neural network that maps the entire color space of galaxies into lower dimensions while preserving the topology of the input space (Kiviluoto 1996; Villmann 1999). This feature ensures that we can have measurements with stable behavior under photometric uncertainties.

We find that the number and stellar mass densities of the candidates are generally consistent with the prediction of the numerical simulations. However, the agreement becomes less pronounced at higher redshifts, where numerical simulations underpredict the number/stellar mass densities of the quiescent galaxies. However, the predictions from numerical simulations are subject to significant Poisson uncertainties.

Although final sample purity requires spectroscopic observations, we employ a conservative approach for the completeness correction used in our number and stellar mass density measurements. In the sense that, if we had used the model galaxies with a quiescent stellar population with FSPS (Conroy et al. 2009; Conroy and Gunn 2010) [2], we get completeness levels that would make the number and mass densities even higher than what we showed here, especially at higher redshift bin. Using the number densities found here and an abundance matching technique, we predict an upper limit for the lowest halo masses of these galaxies and the corresponding lower limit for the highest efficiency of galaxy formation (Vale and Ostriker 2004; Wiklind et al. 2008; Conroy and Gunn 2010; Guo et al. 2010; Moster et al. 2010; Behroozi et al. 2010; Trujillo-Gomez et al. 2011; Reddick et al. 2013; Moster et al. 2013).

Here, we use the abundance matching technique described in Behroozi et al. (2013a), which matches galaxies in the order of decreasing stellar mass to the decreasing peak historical halo mass, according to their accretion history. Therefore, we find the peak historical halo mass corresponding to the number density of candidates, using cumulative peak historical mass function from Behroozi et al. (2013c) at a given redshift. We find the corresponding

---

[2] Using τ model star-formation histories with $\tau = 0.2$ Gyr and the age equal to the age of the universe at each redshift bin, with solar metalicities.

dark matter halos of mass $M_h \approx 4.2, 1.9, 1.3 \times 10^{12} M_{\odot}$ at different redshift bins in Table 2.5 centered at $z = 3.4, 4.1, 4.7$ and $M_h \approx 6.5, 4.7, 4.5 \times 10^{12} M_{\odot}$ for the lower bound of the number densities, respectively. The estimated halo masses are similar to the critical shock-heating halo mass at these redshifts (Birnboim and Dekel 2003; Kereš et al. 2005; Cattaneo et al. 2006), at which the transition from cold-mode accretion into partly hot-mode accretion happens Dekel et al. (2009). This transition could explain part of the quenching mechanism simply by shutting off the accretion of most of the cold gas into the halo. Although the fraction of the cold/hot mode accretion is higher than in the halos of the same size at a lower redshift, the halo masses found here suggest a relatively lower fraction of cold/hot mode accretion than the less massive halos at the same epoch.

This agrees with the "cosmological starvation" scenario as a quenching mechanism, that suggests there is a tight correlation between the star-formation rates and the mass accretion rate onto the halo (e.g., Larson et al. 1980; Balogh et al. 2000; Van Den Bosch et al. 2008) and was shown to be able to explain the relatively quick quenching time for massive central galaxies at high redshift (Argo simulations, Feldmann and Mayer 2015). The relatively high star formation rate required for building the stellar mass within a short timescale is consistent with sub-millimeter galaxies, that exhibit high star-formation rate of order of few 100 $M_{\odot}$ (e.g., Marchesini et al. 2010; Swinbank et al. 2013), as a potential progenitor of these post-starburst systems (e.g., Toft et al. 2014; Wild et al. 2016; Wild et al. 2020).

Therefore, we suspect that a part of the quenching scenario of these systems could be that there was a phase of high accretion of gas into halos, which leads to high levels of star-formation rates and the subsequent mass build-up. Then the accretion of cold gas declines as the starvation phase begins, possibly by transiting to hot-mode accretion of the gas into the halo, followed by an "over-consumption" phase (McGee et al. 2014; Balogh et al. 2016), in which massive galaxies at high redshift use their gas in a depletion time scale which leads to their relatively fast quenching (in $\sim$ few 100 Myrs). Similarly, Estrada-Carpenter et al. (2019) found that massive galaxies at $z \sim 1-2$ enriched quite rapidly to approximately solar metallicities as early as $z \sim 3$. However, radio mode AGN feedback are still required for keeping the galaxy quiescent and suppressing any residual star formation by driving out the remaining gas as well as possible later accretion of a cold gas stream that penetrated the hot halo (Best et al. 2005; Croton et al. 2006; Bower et al. 2006; Feldmann and Mayer 2015; Man et al. 2019; Falkendal et al. 2019). Similar to what has been observed at lower redshift, we find the quiescent sample found here to be consistent with the green valley (quenching) transition stellar mass ($log(M_s/M_\odot) \sim 10.7$) going all the way out to $z \sim 5$. This points to the importance of this stellar mass scale in quenching of the galaxies even at higher redshifts. To find the most important quenching mechanism of these galaxies need additional consideration such as their morphology to find if there are consistent with a compaction phase that was driven by the gas rich major mergers or counter rotating streams from cosmological accretion, which might deliver gas more effectively to the supermassive blackholes and cause phases of AGN activity which could have quenched these galaxies and are mostly have mass scale in which AGN outflows shown to be of great importance (Förster Schreiber et al. 2019).

2.7 Appendix: Quiescent Selection function

To quantify the effect of decision boundaries, we use an extended logistic function to differentiate between points closer to boundaries and those reside well beyond the boundaries. A transformation which takes the distance from the selection boundary and produces a value between $[0, 1]$ in which the value closer to zero represent a galaxy which resides in the star-forming section and far from the selection boundaries and the values close to one is the counterpart inside the quiescent region. However, any squashing function that takes a value from $(-\infty, \infty)$ and maps it to $[0, 1]$ with the limiting behavior discussed above can do the job. We use a generalized logistic function defined below:

$$Y(x) = A + \frac{K - A}{(C + Qe^{-B(x-M)})}$$

in which A, K, B, C, Q, M are free parameters. However, we can fix several of them without losing the needed features, and by the fact that we are interested in maps from $(-\infty, \infty)$ to $[0, 1]$. Changing the parameters of this function can change the values we use to distinguish between different candidates. However, the utility of this method is to compare different candidates given a set of chosen selection parameters, and we do not care about a particular value. A "reasonable" choice of these parameters only generalizes the selection methods without losing information compared to applying a hard selection function.

To combine different criteria consistently, we assume that the function $Y(x_i)$ is the membership likelihood of the i^{th} galaxy in the fuzzy set defined for specific criteria:

$$A = \{(u, m) \mid u \in U \quad \& \quad m = \mu_A(u)\}$$

in which U is the set of all galaxies, and μ_A is the membership likelihood of the galaxies to criterion A. Also for combining different criterion, we use the following rules for t-norm (intersection) and t-conorm (Union):

$$A \cap B = \{(u, m) \mid u \in U \quad \& \quad m = \mu_{A\cap B}(u)\}$$

$$A \cup B = \{(u, m) \mid u \in U \quad \& \quad m = \mu_{A\cup B}(u)\}$$

$$\mu_{A\cap B}(x) = \mu_A(x)\mu_B(x)$$

$$\mu_{A\cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$$

Using the above equations and the parameters are fixed according to the Table 2.6.

Selection	Criteria	Source	Set	Likelihood function	Final criteria
Observed colors	$x_1 = 3.44 - (Y - J) - 3.40 \times (H - K)$	Nayyeri et al. (2014)	A_1	$\mu_1(x_1)$	$((A_1 \cap A_2 \cap A_3)$ $\cup$ $(A_4 \cap A_5 \cap A_6))$ $\cap$ $(A_7 \cap A_8)$
	$x_2 = -1.03 + (Y - J) + 0.67 \times (H - K)$		A_2	$\mu_1(x_2)$	
	$x_3 = 0.83 + (Y - J) - 0.67 \times (H - K)$		A_3	$\mu_1(x_3)$	
	$x_4 = 2.48 + (J - H) - 2.80 \times (H - K)$		A_4	$\mu_1(x_4)$	
	$x_5 = -1.70 + (J - H) + (H - K)$		A_5	$\mu_1(x_5)$	
	$x_6 = 0.16 + (J - H) - 0.43 \times (H - K)$		A_6	$\mu_1(x_6)$	
	$x_7 = S/N(U) - 2$		A_7	$\mu_2(x_7)$	
	$x_8 = S/N(B) - 2$		A_8	$\mu_2(x_8)$	
UVJ	$y_1 = (U - V) - 0.88 \times (V - J) - 0.59$	Whitaker et al. (2012a)	B_1	$\mu_1(y_1)$	$((B_1 \cap B_2 \cap B_3)$ $\cup$ $(C_1 \cap C_2 \cap C_3)$ $\cup$ $(D_1 \cap D_2 \cap D_3))$ $\cap$ B_4
	$y_2 = (U - V) - 1.2$		B_2	$\mu_1(y_2)$	
	$y_3 = 1.4 - (V - J)$		B_3	$\mu_1(y_3)$	
	$y_4 = S/N(J) - 2$		B_4	$\mu_2(y_4)$	
	$y_1 = (U - V) - 0.88 \times (V - J) - 0.59$	Muzzin et al. (2013)	C_1	$\mu_1(y_1)$	
	$y_2 = (U - V) - 1.3$		C_2	$\mu_1(y_2)$	
	$y_3 = 1.5 - (V - J)$		C_3	$\mu_1(y_3)$	
	$y_1 = (U - V) - 0.88 \times (V - J) - 0.56$	Straatman et al. (2014)	D_1	$\mu_1(y_1)$	
	$y_2 = (U - V) - 1.3$		D_2	$\mu_1(y_2)$	
	$y_3 = 1.6 - (V - J)$		D_3	$\mu_1(y_3)$	

Likelihood functions	Logistic parameters
μ_1	A : 0, K : 1, C : 1, Q : 3, B : 20, M: 2
μ_2	A : 1, K : 0, C : 1, Q : 50, B : 1.4, M: 0.25

Table 2.6: The table shows the choice of criteria for the selection functions. Sets A and B are defined $\{(u, \mu(u)) \mid u \in U\}$, in which A is the selected fuzzy set and U is the set of all galaxies. The logistic parameters chosen for each likelihood function is specified.

Table 2.7: Table of the candidate galaxies and their likelihood measures from different methods. The * and $^+$ next to the IDs show the candidates found in the primary and secondary samples reported in Merlin et al. (2019), respectively. The * next to the redshift indicates those with available spectroscopic redshifts. The BBG, UVJ, and SED stand for likelihood based on the observed color, UVJ, and SED fitting, respectively. MIPS and X-ray columns correspond to the detection in the MIPS 24 μm and X-ray. M_* is $\log{(M_s/M_\odot)}$, and Age is the age of the candidate galaxy in Gyrs. The galaxies that show a detection in MIPS 24 μm and X-ray are excluded from the number/stellar mass densities measurements (i.e., Figure 2.9).

Field	**ID**	**RA**	**DEC**	**BBG**	**UVJ**	**SED**	M_*	**Age**	**redshift**	**MIPS**	**X-ray**
COSMOS	3871	150.0839297	2.2235337	0.13	0.97	0.99	10.79	1.61	2.83	False	False
COSMOS	13639	150.0685942	2.3427211	0.35	0.98	0.95	10.86	1.46	2.93	False	False
COSMOS	14284	150.1060909	2.3510462	0.01	0.88	0.73	10.29	1.38	3.51	False	False
COSMOS	14403	150.109033	2.3523803	0.0	0.91	0.56	10.45	1.47	3.73	False	False
COSMOS	14528	150.1033018	2.3536424	0.0	0.46	0.35	10.63	1.11	3.12	False	False
COSMOS	14788	150.0543714	2.356883	0.07	0.33	0.78	11.2	1.11	4.02	False	False
COSMOS	16676*	150.0614902	2.3786845	0.01	0.99	0.56	11.26	1.26	4.13	False	True
COSMOS	16726	150.102881	2.3794029	0.01	0.99	0.56	11.01	0.87	3.65	True	False
COSMOS	16948	150.0667157	2.3823608	0.04	0.91	0.43	10.8	0.87	3.70	False	False
COSMOS	18735	150.1557627	2.4044776	0.0	0.46	0.63	10.27	1.07	3.30	False	False
COSMOS	19502*	150.1308597	2.4135984	0.22	0.86	0.58	10.78	0.84	3.87	False	False
COSMOS	21794	150.1707301	2.4442981	0.1	0.95	0.54	10.51	1.32	3.63	False	False
COSMOS	26635	150.0634065	2.5151526	0.13	0.92	0.61	10.43	1.08	3.24	False	False
COSMOS	26858	150.1648286	2.5190274	0.0	0.47	0.48	10.28	1.94	3.08	True	False
COSMOS	27856	150.0823166	2.5345944	0.18	0.95	0.67	10.79	0.87	4.08	True	True
COSMOS	31823	150.0597769	2.2810259	0.0	0.97	0.41	10.73	1.52	3.72	True	False
COSMOS	32689	150.2035188	2.3120491	0.0	0.55	0.73	10.62	1.47	3.64	False	False
COSMOS	33389	150.1758211	2.3387828	0.0	0.91	0.52	10.37	0.93	4.46	False	False
COSMOS	33927	150.0539523	2.3578909	0.0	0.46	0.4	10.04	1.26	3.25	False	False
COSMOS	33970	150.0799589	2.3592574	0.0	0.9	0.35	10.13	1.04	4.42	False	False
COSMOS	34944	150.0534942	2.3927052	0.0	0.42	0.58	10.19	1.66	3.53	False	False
COSMOS	35033	150.0530481	2.3964053	0.0	0.34	0.92	10.23	1.38	3.67	False	False
COSMOS	35098	150.0530811	2.3992994	0.0	0.46	0.84	10.14	1.54	3.19	False	False
COSMOS	35162	150.0536619	2.4022545	0.01	0.95	0.71	10.38	0.92	4.30	False	False
COSMOS	36674	150.1204151	2.4658729	0.0	0.48	0.84	10.18	1.86	3.19	False	False
COSMOS	37304	150.1030456	2.4952038	0.0	0.93	0.49	10.16	1.6	3.57	False	False
COSMOS	38150	150.0825841	2.5312043	0.0	0.99	0.35	10.2	0.99	4.29	False	False

Field	**ID**	**RA**	**DEC**	**BBG**	**UVJ**	**SED**	M_*	**Age**	**redshift**	**MIPS**	**X-ray**
EGS	16431	215.191335	53.074718	0.0	0.96	0.5	10.45	1.09	3.49	False	False
EGS	21158	214.746219	52.783393	0.0	0.98	0.44	11.0	1.34	3.90	True	False
EGS	21351*	214.673655	52.732542	0.3	0.96	0.5	10.59	0.88	3.61	False	False
EGS	22706	215.122517	53.058015	0.0	0.98	0.51	10.5	1.22	3.25	False	False
EGS	23036*	214.879114	52.88807	0.0	0.73	0.46	10.22	0.98	3.56	False	False
EGS	23572	215.144538	53.078392	0.0	0.98	0.44	10.07	1.64	3.15	False	True
EGS	24177*	214.866081	52.884232	0.0	0.92	0.58	10.98	1.78	3.42	True	False
EGS	24356+	214.620084	52.70959	0.06	0.92	0.47	10.66	1.53	3.43	False	False
EGS	24948	214.767294	52.81771	0.0	0.75	0.52	10.38	1.38	3.44	False	False
EGS	25724*	214.997776	52.986129	0.0	0.96	0.5	10.6	1.43	3.80	False	False
EGS	27491+	214.617755	52.724101	0.56	0.92	0.46	10.55	1.18	3.34	False	False
EGS	29547*	214.695306	52.796871	0.16	0.97	0.51	10.54	1.59	3.15	False	False
EGS	30198	214.966237	52.983055	0.0	0.49	0.42	10.03	1.26	3.01	False	False
EGS	30619	214.981814	52.991238	0.0	0.94	0.57	10.56	1.95	3.07	False	False
EGS	32592	215.080479	52.921568	0.0	0.97	0.53	10.2	0.97	4.28	False	False
EGS	33316	215.252837	53.055545	0.0	0.96	0.54	10.47	1.04	3.70	False	False
EGS	35080	214.922243	52.854748	0.0	0.79	0.59	10.53	0.9	4.00	False	False
EGS	35459	215.231097	53.079181	0.0	0.88	0.32	10.0	1.31	3.95	False	False
EGS	36375	214.778097	52.774151	0.0	0.46	0.34	10.51	1.6	3.27	False	False
EGS	38679	214.692609	52.753274	0.0	0.96	0.65	10.92	1.36	3.91	False	False
EGS	41385	214.701834	52.812333	0.0	0.35	0.33	10.37	0.96	3.63	True	False
GOODSN	599	189.22477876	62.12097508	0.0	0.98	0.37	10.7	1.3	3.77	False	False
GOODSN	3225	188.99068993	62.16051335	0.13	0.45	0.94	10.04	1.98	3.18	False	False
GOODSN	4004*	189.26573822	62.16839485	0.54	0.77	0.63	10.31	1.03	3.81	False	False
GOODSN	4572	189.32906181	62.17385428	0.56	0.63	0.53	10.37	1.0	4.15	True	True
GOODSN	4691+	189.10990126	62.17519651	0.88	0.99	0.48	10.7	1.64	3.18	False	False
GOODSN	5059*	189.1623096	62.17823976	0.01	0.98	0.51	10.79	1.08	3.69	False	False
GOODSN	5744*	189.10012474	62.18362694	0.05	0.98	0.51	10.39	1.11	3.46	False	False
GOODSN	6989	189.22148283	62.19241118	0.48	0.98	0.98	10.22	1.65	2.80	False	False
GOODSN	8074	189.15625504	62.19909581	0.48	0.93	1.0	11.86	0.63	5.11	True	False
GOODSN	8109	189.26660705	62.19930947	0.01	0.47	0.89	10.63	1.68	3.41*	True	False
GOODSN	9083	189.33429106	62.20610777	0.69	0.58	0.2	10.11	1.53	3.62	True	False
GOODSN	9545	188.98263204	62.20890348	0.16	0.34	0.36	10.41	1.42	4.27	False	False
GOODSN	24582	189.40615811	62.34785252	0.5	0.35	0.0	9.2	1.95	2.87	False	False
GOODSN	27305	189.21770656	62.31096041	0.0	0.98	0.51	10.18	2.03	3.01	False	False

Field	**ID**	**RA**	**DEC**	**BBG**	**UVJ**	**SED**	M_*	**Age**	**redshift**	**MIPS**	**X-ray**
GOODSS	2032	53.2293854	-27.8977318	0.0	0.96	0.48	10.07	1.09	3.08	False	False
GOODSS	2717*	53.1893272	-27.8884506	0.0	0.99	0.48	10.92	1.48	3.03	False	False
GOODSS	2782*	53.0835724	-27.8875294	0.75	0.92	0.49	10.48	1.53	3.58	False	False
GOODSS	3912*	53.0622215	-27.8749809	0.35	0.89	0.04	10.17	1.34	3.90	False	False
GOODSS	4503$^+$	53.1132774	-27.869875	0.0	0.86	0.42	10.91	1.29	3.59	False	False
GOODSS	4821	53.0825539	-27.866745	0.04	0.75	0.51	10.24	1.5	3.10	False	True
GOODSS	5479	53.0784645	-27.8598576	0.83	0.71	0.26	11.07	1.15	3.66*	True	False
GOODSS	6131	53.0916061	-27.8533421	0.47	0.96	1.0	11.74	1.1	5.06	True	False
GOODSS	6235	53.1199837	-27.8519554	0.36	0.32	0.0	10.19	1.66	3.50	False	False
GOODSS	7526$^+$	53.0786781	-27.8395462	0.25	0.97	0.44	10.14	1.32	3.32	False	False
GOODSS	8785*	53.0818481	-27.8287373	0.61	0.98	0.48	10.23	1.47	3.85	False	False
GOODSS	9209*	53.1081772	-27.8251228	0.31	0.63	0.49	10.65	1.2	4.49	False	False
GOODSS	12178$^+$	53.0392838	-27.7993088	0.87	0.79	0.04	10.36	1.62	3.29	False	True
GOODSS	12407	53.2218704	-27.7976608	0.0	0.97	0.34	11.06	1.03	4.24	True	False
GOODSS	13299	53.2072105	-27.7913475	0.53	0.43	0.0	10.22	0.91	4.05	False	False
GOODSS	16671	53.1901817	-27.7691402	0.51	0.98	0.77	10.46	1.59	2.87	False	False
GOODSS	17258	53.1411247	-27.7643566	0.31	0.32	0.0	10.4	1.31	4.52	False	False
GOODSS	17749*	53.1968956	-27.7604523	0.99	0.99	0.49	10.69	1.69	3.70	True	False
GOODSS	18180*	53.1812248	-27.756422	0.98	0.96	0.46	10.62	1.01	3.65	False	False
GOODSS	19883*	53.0106544	-27.7416039	0.0	0.97	0.42	10.87	0.97	3.57	True	True
GOODSS	20111	53.0944252	-27.7392502	0.0	0.32	0.56	10.82	1.08	3.13	True	False
GOODSS	22085*	53.0738754	-27.7221718	0.0	0.92	0.48	10.36	1.17	3.47	False	False
GOODSS	32527	53.0305367	-27.7529354	0.0	0.96	0.35	10.48	0.83	4.28	True	False
UDS	164	34.3170586	-5.2759099	0.0	0.44	0.38	10.46	0.92	3.53	False	False
UDS	416	34.5199471	-5.2747688	0.0	0.98	0.46	11.39	1.33	3.38	True	False
UDS	635	34.5062943	-5.273128	0.0	0.99	0.53	11.0	0.92	4.14	False	False
UDS	918	34.2661095	-5.2721262	0.0	0.77	0.44	10.88	1.09	3.17	True	False
UDS	1244*	34.2894669	-5.269805	0.32	0.89	0.01	10.77	0.99	3.79	False	False
UDS	1408	34.5122528	-5.2688479	0.0	0.96	0.49	10.77	0.98	4.16	False	False
UDS	2571*	34.290432	-5.2620749	0.7	0.8	0.08	10.52	1.39	3.70	False	False
UDS	3752	34.5192871	-5.2553701	0.0	0.73	0.52	10.1	1.5	3.11	False	False
UDS	4319	34.4653702	-5.2520308	0.23	0.99	0.51	11.69	0.89	4.48	True	False
UDS	4332$^+$	34.4656906	-5.2519188	0.01	0.93	0.57	10.98	1.7	3.18	True	False
UDS	5256	34.2441864	-5.2458172	0.0	0.34	0.89	10.89	1.86	3.31	False	False
UDS	6218	34.3409538	-5.2405558	0.61	0.76	0.01	10.89	0.93	4.07	False	False
UDS	7520*	34.2558746	-5.23383	0.57	0.96	0.46	11.1	1.48	3.17	True	False
UDS	7779$^+$	34.2588844	-5.2323041	0.54	0.48	0.29	10.66	1.75	3.14	False	False
UDS	8682$^+$	34.2937317	-5.2269621	0.82	0.92	0.26	10.44	1.58	3.46	False	False

Field	**ID**	**RA**	**DEC**	**BBG**	**UVJ**	**SED**	M_*	**Age**	**redshift**	**MIPS**	**X-ray**
UDS	13988	34.3857307	-5.1989851	0.36	0.05	0.63	11.45	1.95	3.03	True	False
UDS	15748	34.5302429	-5.1890779	0.0	0.46	0.36	11.26	2.05	3.08	True	False
UDS	17344	34.3231277	-5.179821	0.35	0.16	0.44	10.88	1.87	3.03	True	False
UDS	17790	34.5422859	-5.1774998	0.34	0.33	0.0	10.54	1.58	3.31	True	False
UDS	18672	34.5668411	-5.1726952	0.92	0.51	0.0	10.32	1.32	3.79	False	False
UDS	19849	34.3381882	-5.1661878	0.07	0.98	0.41	10.41	1.66	3.53	True	False
UDS	20843*	34.4961014	-5.161037	0.64	0.65	0.19	10.76	0.88	3.74	False	False
UDS	22354	34.4276466	-5.1524191	0.72	0.38	0.03	10.93	1.9	3.15	False	False
UDS	23628*	34.2425995	-5.1430721	0.97	0.63	0.21	10.73	0.83	4.25	False	False
UDS	24501	34.5228386	-5.1288252	0.0	0.82	0.34	10.43	1.55	3.40	False	False
UDS	24734	34.5229988	-5.1295991	0.0	0.88	0.49	10.36	1.55	3.47	False	False
UDS	25688*	34.5265884	-5.1360388	0.25	0.99	0.53	11.21	1.33	3.08	False	False
UDS	25893*	34.3996353	-5.1363459	0.08	0.99	0.55	11.05	1.24	4.49	False	False
UDS	32698	34.5237198	-5.1804299	0.0	0.36	0.61	10.99	1.29	4.30	True	False
UDS	35635	34.3170319	-5.127574	0.0	0.93	0.36	10.58	1.36	3.72	False	True

Chapter 3

Statistical Learning Application on the Simulated Lightcones Multiwavelenght Catalog

3.1 Introduction

A serious challenge in astronomy today is to handle the extensive data acquired from multi-waveband galaxy surveys from the ground or space-based observatories. That is, to extract optimum information from the data and look for new statistical trends and correlations within the data in multi-dimensional space, essential to address the scientific questions that instigated these surveys. Given the diversity and the large volume of the data in question, this requires innovative and sophisticated statistical modeling techniques (e.g., Djorgovski et al. 2012; Ménard et al. 2013; Leistedt and Hogg 2017a; Leistedt and

Hogg 2017b; Hogg et al. 2019; Leistedt et al. 2019; Cranmer et al. 2019; also see Ball and Brunner 2010; Baron 2019 and references therein).

Increases in computational power combined with sophisticated algorithms and more realistic modeling of the physics involved have allowed more accurate simulations of galaxies' statistical properties (see Somerville and Davé 2015; Naab and Ostriker 2017 for a review). These simulations can be constrained using the observed data and then used to find hidden complex relations between different physical parameters. The resulting statistical models can be used to define many inference problems, which can then be applied to real data and compared to find potential differences between different models.

For example, simulated data are used to build light cones similar to those observed in photometric surveys of galaxies by matching the geometry, magnitude distributions, limiting magnitudes, and associated uncertainties. Since the outputs from these simulated surveys are known, they are essential to test and compare the models used to predict the observable parameters. The typical approach for measuring galaxies' physical parameters (redshift, mass, star formation rate, metallicity, extinction) is to search in the parameter space of theoretical libraries of the Spectral Energy Distributions (SEDs). The physical parameters of any galaxy are calculated using maximum likelihood method or sampling from the posterior distributions.

However, in this conventional approach, one does not usually consider information from the population-level properties of galaxies. For example, priors are usually chosen to be least informative to avoid imposing any unnecessary constraint on the parameters. Therefore, the resulting posterior distribution on the model parameters only depends on the simplistic priors and input measurements, missing additional information that can be incorporated in the prior, such as conditional probabilities (e.g., probability of observing a star-forming versus quiescent galaxy at the center of a galaxy cluster). The posterior distribution of a parameter can be found using the extra information within the population produced from models, including much more complex and realistic physics but lacking tractable likelihood. In other words, finding the optimal model is done through a Maximum Likelihood Estimate (MLE) or sampling from the parameter space of the explicit models for every galaxy and imposing some selection criteria. Instead, one can approximate the function from the observable to that parameter directly from a large enough simulated sample produced by an implicit and high-fidelity statistical modeling involved in galaxy formation simulations.

Using the simulated data, a decision function is defined based on the physical parameters such as stellar mass, star-formation rates, redshift by calculating/modeling an approximated posterior distribution without having to marginalize over all the simulation internal physical states (i.e., stellar mass, star-formation rates) explicitly through the likelihood. This technique has now been developed using hydrodynamic simulations (Davidzon et al. 2019) and semi-analytic models (SAMs) (Simet et al. 2019).

This work investigates the statistical and machine learning abilities to classify subpopulations of galaxies defined based on their physical properties under realistic conditions for an actual galaxy photometric catalog similar to a combined catalog based on the extensive upcoming surveys. There have been large number of works investigating the use of statistical approaches in variety of applications in astronomy (see Ivezić et al. 2019; Baron 2019 for recent overviews) such as estimating photometric redshifts (e.g., Firth et al. 2003; Vanzella et al. 2004; Collister and Lahav 2004; Carliles et al. 2010), morphological classification (e.g., Storrie-Lombardi et al. 1992; Lahav et al. 1995; Odewahn et al. 1996; Bazell and Aha 2001; Ball et al. 2004; Lintott et al. 2008), and star-galaxy classification (e.g,. Odewahn et al. 1992; Bertin and Arnouts 1996; Henrion et al. 2011; Fadely et al. 2012; Kovács and Szapudi 2015; Kim and Brunner 2017; Sevilla-Noarbe et al. 2018; Beck et al. 2021). Here, we start by statistical decision theory argument for selecting a particular class of galaxies that leaves us with either making a probabilistic model or taking an empirical approach. We start by employing a probabilistic modeling approach with varying degree of assumptions on the joint probability distribution of galaxies magnitude and their labels and move to more empirical approaches without an explicit probabilistic model.

As an example, we focus on a population of high redshift quiescent and relatively massive galaxies. During the last decade, the presence of this population of galaxies has been established (e.g., Cimatti et al. 2004; Labbé et al. 2005; Rodighiero et al. 2007; Grazian et al. 2007;Fontana et al. 2009; Straatman et al. 2014; Nayyeri et al. 2014; Glazebrook et al. 2017; Schreiber et al. 2018b; Merlin et al. 2018; Merlin et al. 2019; Carnall et al. 2019a; Santini et al. 2019; Girelli et al. 2019; Santini et al. 2020); and particularly within the past few years, there was a growing number of confirmations primarily using photometric selection and follow up from near-infrared spectroscopy to sub-mm imaging at progressively higher redshift (e.g., Gobat et al. 2012; Glazebrook et al. 2017; Schreiber et al. 2018b; Schreiber et al. 2018c; Tanaka et al. 2019; Valentino et al. 2020; Forrest et al. 2020b; Forrest et al. 2020c; Santini et al. 2019; Santini et al. 2020; Saracco et al. 2020; D'Eugenio et al. 2021).

The presence of these galaxies at high redshift, and particularly their mass assembly and fast quenching, are challenging to reproduce in the theoretical models of galaxy formation and evolution (e.g., Fontana et al. 2009; Steinhardt et al. 2016; Schreiber et al. 2018c; Santini et al. 2020; Forrest et al. 2020b). Studies of these galaxies and possibly their dust and gas content can put constraints on the efficiency of star formation, and relative effectiveness of different quenching mechanisms and their timescale at early epochs of galaxy evolution within massive dark matter halos (e.g., Schreiber et al. 2018c; Belli et al. 2019; Merlin et al. 2019; Forrest et al. 2020b; Carnall et al. 2020; Whitaker et al. 2021). However, drawing a more general picture of the possible evolutionary pathways of these galaxies requires a statistically representative sample.

While we focus on this specific sample of galaxies, the methods can also be generalized to other objects. We will compare the performance of different techniques with the "ground truth" and, hence, each technique's accuracy and ability in identifying this population of galaxies. The results are compared with those from the conventional SED fitting method applied on the same dataset. In addition, one can turn this problem around to find the most informative parts of the spectrum to observe to maximize the number of interesting objects in question. To address this, a method is used to measure the importance of different photometric bands for selecting a particular class of galaxies. The method can be extended to different categories of objects and has wide applications within astronomy.

§3.2, present details of the SAM used to generate the simulated galaxies and their multi-waveband data. The use of the observed multi-wavelength catalog to add realistic uncertainties and missing values to the simulated counterparts is discussed. Then an estimator for every artificially injected missing value in the dataset is made, and its accuracy and performance are measured based on the actual values in the SAM. In §3.3 the classification problem is defined as a statistical decision problem, and several classification techniques are used on the galaxies' color space with different assumptions and levels of complexity from linear and quadratic to non-linear and non-parametric decision boundaries. In §3.4, the performance of the classification based on SED-fitting is compared to all these methods. The last section provides an overview of the results, possible drawbacks, and ways to improve.

We assume a standard cosmology with $H_0 = 70\,\mathrm{kms^{-1}Mpc^{-1}}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. All magnitudes are in the AB system where $\mathrm{m}_{AB} = 23.9 - 2.5\log(f_\nu/1\mu\mathrm{Jy})$ (Oke and Gunn 1983). We use bold notation for vectors ($\boldsymbol{x}$) and matrices ($\boldsymbol{C}$) throughout the paper.

3.2 Simulated Galaxy Catalog

3.2.1 The Semi-analytic Model

The light cones that are generated for the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS) fields (Grogin et al. 2011; Koekemoer et al. 2011) from an updated version of the Santa Cruz semi-analytic modeling framework are used here (Somerville and Primack 1999; Somerville et al. 2001; Somerville et al. 2008; Somerville et al. 2012; Yung et al. 2019a; Yung et al. 2019b; see Somerville et al. 2021).

The SAM framework uses dark matter merger histories either from dissipationless N-body simulations or Extended Press-Schecter (EPS); Press and Schechter 1974; Lacey and Cole 1993) using formalism described in Somerville and Kolatt (1999) and Somerville et al. (2008). Each branch of the merger tree is traced back in time to the minimum progenitor mass (referred to as the simulation's mass resolution). The SAM used in this study uses the root dark matter halos from the “Bolshoi-Planck" high-resolution cosmological simulation (Klypin et al. 2011) along the past lightcones created to match the CANDELS' field's sky location while covering a significantly larger area. However, the halo merger trees are constructed using the EPS formalism as described in Somerville et al. (2008). The dark matter halos then were populated with galaxies using the SAM framework.

In the merger of dark matter halos, the most massive galaxy is defined as “central” and others are called “satellites". The satellite galaxies lose their angular momentum because of dynamical friction and lose about 30-40 percent of their mass per orbit through tidal stripping (Somerville et al. 2008). Through this process, a galaxy can be tidally destroyed before merging into the central galaxy, in which case the stars are added to the diffuse stellar halo.

Gas accreted onto the halos is shock-heated to the newly formed halo's virial temperature. The gas is cooled down with a cooling rate consistent with a smooth and spherically symmetric model for density profile (White and Frenk 1991). A cooling radius ($r_{\rm cool}$) is defined as the radius within which the hot gas has enough time to cool within the dynamical timescale. Gas accretion is divided into two regimes: a “hot mode" defined as $r_{\rm cool} < r_{\rm virial}$, and a “cold mode" as $r_{\rm cool} > r_{\rm virial}$, where the gas is accreted within the dynamical timescale.

There are two modes of star-formation, quiescent star-formation (or disk mode) and merger-driven starburst. The disk star-formation mode uses an updated version of the classic Kennicutt-Schmidt (KS) relation (Schmidt 1959; Kennicutt 1989; Kennicutt 1998), where instead of the total cold neutral gas surface density, a recipe based on only the H_2 surface density is used (Popping et al. 2014; Somerville et al. 2015). This model predicts an earlier formation time for massive galaxies than that based on the classic KS relation, which is particularly necessary for this study.

The merger-driven starbursts are based on the binary galaxy merger simulations (i.e. Robertson et al. 2006; Hopkins et al. 2009a), and the description of their timescale and amplitude is modeled based on the merger mass ratio and the gas fraction of their progenitors (i.e. Somerville et al. 2008; Somerville et al. 2012; Somerville et al. 2012). The merger is also responsible for the galaxies' morphological evolution by removing angular momentum from the disk and building up the spheroid. This is modeled based on the simulations in Hopkins et al. (2009b). It is updated from the original model described in Somerville et al. (2008).

The production of heavy elements is assumed to follow the simple, instantaneous recycling approximation and is assumed to mix with cold gas in the disk. These elements can be removed by the supernova winds and deposited in the hot gas or outside the halo, which later can re-accrete on the galaxy (see Somerville et al. 2008; Somerville et al. 2015).

Stellar feedback from massive stars and supernovae can drive the cold gas out of the galaxy. The resulting outflow rate is modeled based on the star formation rate (SFR) and the galaxy's escape velocity. The fraction of this outflow deposited into the hot gas reservoir versus the outside of the halo is modeled based on the halo's circular velocity (Somerville et al. 2008), such that lower mass halos lose a higher fraction of their gas compared to their massive counterpart.

Every galaxy is assumed to have a seed central black hole, which can grow in two modes. One is the “quasar/bright/radiatively efficient mode", which is driven by mergers and disk instabilities, and after each merger, the black holes merge instantaneously. The resulting black hole will accrete cold gas and grow at Eddington rate until the pressure-driven outflows deposit enough energy into the central region of the galaxy that significantly slow (or halt) the accretion (Somerville et al. 2008). Additionally, it produces winds that can remove cold gas from the halo. The other mode is the ‘radio mode" and is considered responsible for the luminous jets observed in the radio wavelengths. This growth mode is driven by accretion of hot halo gas assuming the Bondi-Hoyle-Lyttleton model (Hoyle and Lyttleton 1939; Bondi and Hoyle 1944; Bondi 1952), which results in a much lower accretion rate (2-3 orders of magnitudes slower than the Eddington rate) and is assumed to produce radio jets that couple effectively only to the hot halo gas, which heats the gas and increases the cooling timescale.

Finally, the resulting model’s free parameters are fixed to match a set of $z \sim 0$ observations (e.g. see Yung et al. 2019a; Somerville et al. 2021). Using the star-formation and metalicity history of galaxies produced by SAM, synthetic spectral energy distributions (SED) are created by convolving these histories with simple stellar population models of Bruzual and Charlot (2003a) and the models from Padova 1994 isochrones (Bertelli et al. 1994) and Chabrier (2003) initial mass function. The unattenuated synthetic SED is then attenuated by the dust model, assuming a prescription for the V-band optical depth. This depends on the metallicity of the cold gas as well as the mass and radius of the cold gas disk. It also depends on the inclination of each galaxy (that is assigned randomly) and a free parameter that is a function of redshift that is tuned to match the observational constraints imposed by the bright-end of the luminosity functions at $z = 4 - 10$ (Yung et al. 2019a; Somerville et al. 2021). After including the IGM absorption from Madau (1995), the SEDs are convolved with the filters response functions to find the flux/magnitude at each band.

Using the final synthetic SEDs, we then create galaxy magnitudes catalogs with magnitudes uncertainties and missing values fractions resembling the distributions observed in the CANDELS, which we discuss in detail in §3.2.2.

Training, Testing, and Validation Datasets

There are eight publicly available SAM lightcone realizations for every field in the CANDELS from SAM. Throughout this paper, the GOODS-S field is used, and the **"realization zero"**, and **"realization seven"** are treated as the **Training** and **Testing** dataset. Keeping a large amount of data in the validation and testing set ensures that the results are robust to variations in the random selection of the volume for each realization within the larger simulations. One can choose different sets, for example, by combining the realization and separating the training and testing sets afterward; however, different realizations naturally produce this division.

Since many of the theoretical models tend to produce lower number densities of these objects (e.g., Brennan et al. 2015; Santini et al. 2020; Shahidi et al. 2020) and from an observational perspective, their number densities can have significant uncertainties due to lack of extensive and complete spectroscopic sample, we allow a considerable variation in their number densities by artificially introducing these objects from other realizations. We create an additional training set with oversampling of these objects and add the quiescent population from other realizations (one to five) to introduce more examples of this rare population to the training set. However, when applying prior probabilities, the original fraction in the realization zero is used, and when testing on the original test data, there are no such added examples. When this oversampled version of the training set is used to train the model, it is explicitly mentioned.

In all the methods used in the following sections, whenever there is a hyperparameter in the model, 5-fold cross-validation on the **“realization six" as the validation set** is employed to find the optimal hyperparameter (lowest error rate on the validation set). Then these values are fixed, and a model is trained on the training set.

3.2.2 CANDELS observations

Latest photometric catalogs from the CANDELS are used with multi-waveband photometry and measurement of the physical parameters for all galaxies to the survey's flux limit. CANDELS consists of five fields: GOODS-South (GS) (Guo et al. 2013), UDS (Galametz et al. 2013), COSMOS (Nayyeri et al. 2017), EGS (Stefanon et al. 2017), and GOODS-North (Barro et al. 2019).

Photometric catalogs contain observed photometry from the UV to near and mid-infrared wavelengths. Table 3.1 presents the filters used for the GOODS-S field. For each of the CANDELS fields, the photometric catalogs were selected in HST/WFC3 F160W band using SExtractor (Bertin and Arnouts 1996). For low-resolution images, Template FITting (Tfit; Laidler et al. 2007) smooths the high-resolution image to low resolution and fit the best flux consistent with the HST data (Grogin et al. 2011; Koekemoer et al. 2011). In this study, we focus on the GOODS-South field, which has the most Hubble Space telescope

(*HST*) high-quality imaging data available in nine out of seventeen filters, four of which are in the Near-Infrared.

First, galaxies are selected in the CANDELS GS catalog with ‘FLAGS = 0’, which are non-contaminated sources in the F160W band and ‘CLASS_STAR $\leq$ 0.9’, to avoid contamination from stars based on the F160W SExtractor Star/Galaxy classifier output. Then a subset from all of the available bands is chosen based on the number of the available/not missing photometry for the galaxies in the survey (shown with bold font in Table 3.1). For the U-band, the VIMOS photometry is used; in the K-band, ISAAC K_s is used. WFC3 F098M is removed due to a large number of missing values (those without any observed flux), and only IRAC channels 1 and 2 are employed since the longer wavelength channels are not as deep ($\sim$ 3 magnitude shallower).

3.2.3 Flux Uncertainties

The SAM photometric catalog does not have any associated uncertainty model for the flux measurements. To make the magnitude distributions as close as possible to the CANDELS data, for every galaxy, an uncertainty of each band photometry is assigned based on their flux measurement, using the uncertainties in the CANDELS GS.

CANDELS GS observed photometric catalog is used to find the “typical” uncertainties in each band. First, all forms of missing/non-detection values of the CANDELS GS data are separated into two groups:

1. Set the flux to the limiting flux for that band for galaxies with negative fluxes.
2. Bands with no observations are labeled missing.

Table 3.1: The observed bands from UV to Mid-IR SEDs of galaxies in CANDELS GOODS-South field

Field	**Filter set**
GOODS-S (Guo et al. 2013)	Blanco/CTIO U, VLT/**VIMOS U**, HST/**ACS F435W, F606W, F775W, F814W, F850LP,** HST/WFC3 F098M, **F105W, F125W, F160W,** VLT/**ISAAC Ks**, VLT/Hawk-I K_s, Spitzer/**IRAC 3.6 μm, 4.5 μm,** 5.8 μm, 8.0 μm

GOODS-S Bands used	**Missing values fraction**	**Censored values fraction**
1: VLT/VIMOS U	0.0	0.12
2: HST/ACS F435W	0.045	0.04
3: HST/ACS F606W	0.007	0.01
4: HST/ACS F775W	0.019	0.01
5: HST/ACS F814W	0.028	0.01
6: HST/ACS F850LP	0.003	0.1
7: HST/WFC3 F105W	0.37	0.04
8: HST/WFC3 F125W	0.001	0.04
9: HST/WFC3 F160W	0.0	0.0
10: VLT/ISAAC Ks	0.089	0.17
11: Spitzer/IRAC 3.6 μm	0	0.1
12: Spitzer/IRAC 4.5 μm	0	0.1

Doing this is to make sure that the missing values are missing completely at random and their missingness is independent of these galaxies' properties. Since galaxies with values below the detection limit because of their physical properties are assumed to be censored at the limiting magnitude of the survey. An observed magnitudes catalog is created for the SAM in the following steps:

1. Select SAM galaxies using the H-band limiting AB Magnitude ≤ 27.

2. For flux measurements smaller than the flux corresponding to the limiting magnitude

in the CANDELS GS observation of that band, the limiting magnitude/flux for that particular band is used based on the CANDELS GS limiting magnitude at that band.

3. To take into account the heteroscedasticity of flux measurements uncertainties (i.e., those with smaller flux measurements have higher uncertainties), we follow the approach in Simet et al. 2019. AB magnitudes are divided into bins of width 0.1, and the median of the multiplicative uncertainties[1] is found associated with each bin. A simple linear interpolation is applied to estimate the multiplicative uncertainties at a given AB magnitude for a given band (δF). Based on the flux in the SAM catalogs, a noise is added to each flux is assigned by sampling from a Gaussian distribution with a standard deviation of δF.

4. Assuming the non-observed/missing values in each band are missing completely at random, the flux value at each band based is randomly removed based on the fraction of non-observed values in the CANDELS catalog in that band. In other words, the fraction of missing values at each band is made similar to the CANDELS (GS) observations. Table 3.1 shows the fraction of the missing and censored values in the CANDELS (GS) photometric catalog.

The training, test, and validation samples are treated the same way, using the above procedure.

---

[1] Multiplicative uncertainties δF is defined as $\hat{F} = (1 + \delta F)F$, with F being true flux and $\hat{F}$ being the perturbed flux.

3.2.4 Estimation of the non-observed values

The methods used in this study, like many other learning algorithms, cannot handle the missing values on their own, which means that either the band or galaxy with non-observed value should be removed or some estimates for these missing data points must be made. One can use basic summary statistics such as median/mean or linear interpolation approximation in each band as the simplest estimates. However, here using the information in the data set, these missing magnitudes are estimated as a function of the galaxy's non-missing magnitudes. By finding these estimates for the SAMs catalog's missing data points and using the correct values, the performance of the estimators is measured.

There are many approaches for handling the missing data (Little and Rubin 2019) from using mixture models and expectation maximization (EM) technique[2] to find an approximated probability distribution of the missing data given the observation (**?**), to directly estimating the missing values using different estimators (KNNimpute, Troyanskaya et al. 2001; MissPALasso, Städler et al. 2014; MICE, van Buuren and Groothuis-Oudshoorn 2011; MissForest, Stekhoven and Bühlmann (2011)).

Here the iterative multivariate modeling framework of the missing values is used called Multivariate imputation by chained equations (MICE, van Buuren and Groothuis-Oudshoorn 2011). This estimates the conditional regression model of a specific variable (band) given the rest of the variables in the data using a "pseudo-Gibbs sampling" approach which cycles through each variable conditioned on the rest of the variables. Although this approach does not necessarily lead to consistent conditional distributions, it can be useful

---

[2] Look at Bishop 2006 section 9.4, and Do and Batzoglou 2008

in practice (see Gelman and Raghunathan 2001 for more detailed discussion). We have a total of twelve photometric bands in the SAM and CANDELS GS catalogs. From these, we sort the bands based on the frequency of missing values (SAM and CANDELS GS have very similar sorted bands since we set the SAM missing values proportions using CANDELS GS observations; see §3.2.3). Then the following steps are taken:

1. Start with the bands with the greatest number of missing values, and treat it as a "target quantity" (quantity of interest).
2. The data points with observed values in the target are kept as the complete dataset. If these data points have missing values at other bands, a linear approximation in log(wavelength/) between the nearest observed bands is used.
3. An estimator using a regression model is set for the target variable and trained for the target quantity based on the complete dataset.
4. The missing values of the target quantity are estimated based on the trained model.
5. Go through the first step, but for the band with the next most missing values.
6. Go through these steps until every band with missing values has a model and estimates for its missing values. Stop iterating when these estimates converge within some threshold.

For modeling the target quantities at each step, a Random Forest of 50 decision tree regressor is trained (see §3.3.2), which is very effective in estimating the missing values (Stekhoven and Bühlmann 2011; Allingham 2018). The estimated values converge for all the datasets after about 15 iterations (fluctuation in estimated magnitudes is < 0.001 of the standard deviation of that particular band). Figure 3.1 shows the performance of the collection of regression models on the test set in estimating the missing values of the SAM galaxies.

It should be noted that using this Random Forest will require the estimated values from the regression to have smaller values than the limiting magnitudes of each band. In other words, the estimator is a sampling from a distribution similar to the observed distribution where the data are censored at the limiting magnitudes for larger magnitudes. Figure 3.1 presents the effect of this censoring in that there are no predictions larger than the limiting magnitudes, and the regression imposes the limiting magnitude[3]. One can use other censored regression models to allow the prediction of higher values and impose the limiting magnitude afterward.

Using these procedures, estimates of the missing values are made for each magnitude in the training, testing, and validation datasets, as well as the CANDELS GS observation catalog[4].

3.3 Finding the optimal classifier

To address specific questions, it is a common practice that one need to identify and study a particular population of galaxies (e.g., star-forming, Lyman Break galaxies, quies-

---

[3] The Random Forest regression model will learn the distribution similar to the input which has some data censored at the limiting magnitude compared to the true distribution where these censored data are extended to higher values.

[4] One can apply the estimator trained on the training set to estimate the missing values on the other datasets. However, this introduces a risk of adding bias from the training dataset into our testing and validation dataset, which we later use to assess different classifiers' performance and their optimal hyperparameter.

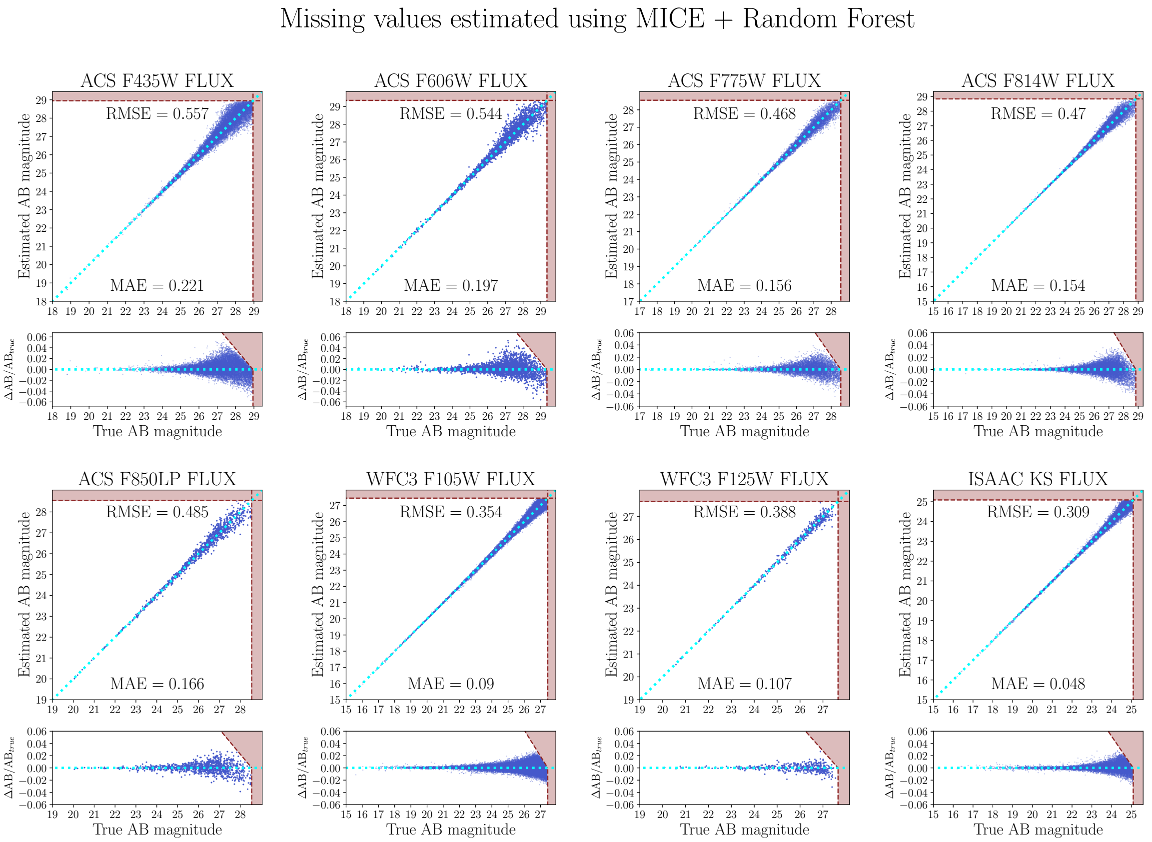


Figure 3.1: Top: Estimated AB magnitudes for the missing values using iterative modeling with Random Forest regressors. The RMSE and MAE are the root mean squared error and mean absolute error of the given band, respectively. The y-axis is the estimated magnitude, and the x-axis is the true magnitude after adding the errors to the SAM lightcones and removing the values artificially. The plot shows those bands with missing values according to the CANDELS GOODS-S photometric catalogs. The shaded regions are the part of the magnitude space where the larger magnitudes are censored due to the limiting magnitudes imposed on the catalog. Bottom: The relative residual of the estimated magnitudes with ΔAB = estimated AB - true AB.

cent, massive quiescent). This section aims to study the effectiveness of different statistical learning algorithms in identifying a specific population of galaxies and comparing them with those from the conventional method. The simulated catalogs that were generated in the last section will be used for this purpose. As a test example here, we focus on the population of **relatively massive quiescent galaxies at high redshift** and use statistical learning techniques to identify them in the simulated catalogs. We also identify these systems using the established technique based on SED-fitting of the observed galaxies' fluxes and compare results.

Identifying a particular population of galaxies such as the quiescent population is a decision problem which requires a decision boundary based on the observed quantities such as colors and fluxes (Cimatti et al. 2002; Daddi et al. 2004; Reddy et al. 2005; Mobasher et al. 2005; Wiklind et al. 2007; Nayyeri et al. 2014), or rest-frame UVJ colors (Labbé et al. 2005; Wuyts et al. 2007; Williams et al. 2009; Whitaker et al. 2013; Barro et al. 2014; Straatman et al. 2014; Merlin et al. 2018) or imposing some criteria on the inferred physical quantities such as star formation rate history (Fontana et al. 2009; Pacifici et al. 2016; Carnall et al. 2018; Carnall et al. 2019b). Given the choices of the decision criteria, one can build a sample of objects. There have been many other approaches to defining the quiescent population in the literature from the original color-magnitude bimodality (e.g., Baldry et al. 2004) to fitting a mixture of Gaussian in the bins of stellar mass for $sSFR$ (e.g., Bisigello et al. 2018; Hahn et al. 2019), and modeling the probability density over the SFR versus M_s plane instead and comparing those directly to simulation and circumvent this definition altogether (e.g., Leja et al. 2021). This study takes a practical approach for defining this population, which helps build large enough candidates for further spectroscopic studies from an extensive photometric catalog.

The population of relatively massive evolved galaxies at high redshift (Q) is defined based on the following criteria on their physical properties:

- $\log_{10}(M_s/M_\odot) \geq \log_{10}(M_{\mathrm{lim}}) = 9.5$
- $sSFR \leq sSFR_{\mathrm{lim}} = \frac{0.2}{t_U(z)}$
- $z \geq 2.5$

where M_s is the stellar mass, $sSFR$ is the star-formation rate per unit stellar mass (in Gyr^{-1}), $t_U(z)$ is the age of the universe at redshift z (in Gyr). We use the evolving $sSFR$ threshold definition based on Pacifici et al. (2016) to find the quiescent population at different epochs, which is consistent with having the current star-formation rate of the galaxy to be less than 10% of its average over galaxy's history.

Using these criteria, we select 286 massive and quiescent galaxies at $z \geq 2.5$ out of a sample of 447,060 galaxies in the simulated catalog. An oversampled training set is made by adding the selected galaxies (based on the above criteria) from magnitude limited realizations one to five to the training set, which results in a dataset with 2159 high-z massive quiescent out of 448,933 galaxies (see section §4.2). The original training set is used unless we state otherwise. The testing and validation sets contain 412 and 330 high-z massive quiescent out of 431,500 and 425,561 galaxies, respectively.

Considering all the galaxies in the catalog, G, and the set of massive quiescent galaxies, Q, the problem of identifying any high-z massive quiescent galaxies at high-z, g, can be formulated by finding the function h,

$$y_g = h(\boldsymbol{x}_g) = \begin{cases} 1 & \text{if } g \in Q \\ 0 & \text{otherwise} \end{cases}$$

where $\boldsymbol{x}_g \in R^d$ is the d-dimensional column vector for galaxy $g \in G$'s magnitudes from the Galaxy's SED. In the following subsections, by applying different statistical learning techniques, this function is either modeled or approximated based on the multi-waveband photometric catalog with all the estimated values for the missing data points in the photometric catalog.

3.3.1 Bayesian Posterior Expected Loss

Selecting a particular type of galaxy is fundamentally a decision problem. This requires quantification of a notion of loss and its dependence on possible decisions h, the data X, and outcomes Y. The loss function depends on two unknown, which are the possible data $\boldsymbol{X}$ and possible outcomes Y, and one wants to find the decision that minimizes the loss. Because of these fundamental constraints, one must make assumptions about the distribution over these random variables and try to minimize the expectation of loss under these models and their assumptions.

One approach to this decision problem is to approximate the Bayes classifier defined as the classifier ($\hat{h}$) that has the minimum expected loss/Bayes risk, $r(h, \pi_Y)$ as the following:

$$r(h, \pi_Y) = E_Y(E_{\boldsymbol{X}|Y}(\text{loss}(h(\boldsymbol{X}), Y)|Y))$$

$$\hat{h} = \arg\min_h \{r(h, \pi_Y)\}$$

with π_Y as the prior probability distribution of the unknown parameter (here is the class label) and $p_{\boldsymbol{X}|Y}$ as the conditional distribution of the galaxies' magnitudes given the labels where $(\boldsymbol{X}, y)$ are random variables associated with possible sets of observations $\{(\boldsymbol{x}_g, y_g)\}$. And E_Y, $E_{\boldsymbol{X}|Y}$ are the expectation over the π_Y and $p_{\boldsymbol{X}|Y}$ respectively. However, instead of finding the optimal decision rule, one could equivalently find the optimal decisions/actions for every single observation. This translates to finding decisions that minimize the posterior expected loss with the expectation taken over the following posterior probability (Berger 1985):

$$p(y_g \mid \boldsymbol{x}_g) = \frac{p_{\boldsymbol{X}|Y}(\boldsymbol{x}_g \mid y_g)\pi_Y(y_g)}{\sum\limits_{y'} p_{\boldsymbol{X}|Y}(\boldsymbol{x}_g \mid y')\pi_Y(y')}$$

$$\hat{h}(\boldsymbol{x}_g) = \arg\min_{h(\boldsymbol{x}_g)}\{\sum_{y'} \text{loss}(h(\boldsymbol{x}_g), y')p(y' \mid \boldsymbol{x}_g)\}$$

where $\boldsymbol{x}_g$ is the column vector of the magnitudes of the galaxy (g), y_g is the label of the galaxy g, and the summation is over y' which are the possible labels for the galaxy g(i.e., whether it is a high-z massive quiescent galaxy or not).

However, since the actual posterior distribution of classifications is unknown, one needs to find an approximation of the posterior by modeling the joint probability distribution $p_{\boldsymbol{X},Y}$. Knowing the posterior probability distribution, one can employ the loss function that is most suitable for the application at hand. For the binary decision here, one needs a two-by-two matrix that provides all the losses under different decisions for different classes and can choose the optimal decision by minimizing the expected loss under the posterior according to the above expression.

The other approach is to approximate the Bayes risk as the empirical risk estimated from some known data and outcome, called the training set (i.e., some samples from the underlying joint distribution) and defining the optimal classification based on minimizing the empirical risk (Vapnik 1992):

$$h^* = \arg\min_h \{ \frac{1}{N} \sum_{i=1}^{N} \text{loss}(h(\boldsymbol{x}_i), y_i) \}$$

In this study, we start by finding approximations to the posterior distribution of class labels. Then we move on to find different approximate mappings from the inputs to the class labels directly assuming some loss function by minimizing the empirical risk with some appropriate regularization (i.e., without making explicit probabilistic models). The performance of different approaches is compared using these notions of uncertainties in each classifier for unseen examples[5] and using different thresholds for what can be accepted as target population, which translate directly to the use of different loss functions/matrices for the binary decision problem at hand.

In the following subsection, we model the Bayes classifier by modeling/estimating the posterior distribution over the classes using the following techniques:

- Spectral Energy Distribution fitting
- Discriminant Analysis
- Gaussian Naïve Bayesian
- Bayesian Logistic Regression

---

[5] Using an empirical estimate of the true error rate of the classifier

Spectral Energy Distribution fitting

Fitting the observed SED of galaxies to a theoretical library of model SEDs is used to predict galaxies' physical characteristics to identify the high-z quiescent population. Here a library of model galaxies SED is made from Bruzual and Charlot (2003a)) assuming a delayed exponentially declining star formation history with an e-folding parameter in the range of $0.02 \leq \tau \leq 30$ (Gyr) (20 uniformly sampled from the logarithmic scale). Models with three metalicities $(0.02, 0.008, 0.004)$ and dust attenuation law from Calzetti et al. (2000) with 20 $E(B-V)$ sampled uniformly between 0 to 1.2 are used with the LePhare SED-fitting code (Arnouts et al. 1999; Ilbert et al. 2006).

The SAM flux catalog is used after adding the proper noise with the uncertainty at each band based on the uncertainty scale for sampling the additive flux measurements noise. To control for how much information is lost due to the missing values, the SED-fitting is performed once on the complete data set and once on the incomplete data set after removing the missing values.

The SED fitting results are found on the test sample while fixing the redshift to their values taken from the SAM catalog (we use the redshift that includes the effect of the peculiar velocities). The median of the marginalized likelihood over nuisance parameters is used to estimate the star formation rates and stellar masses. Fifty realizations are sampled from flux uncertainties as explained in §3.2.3 for every galaxy. The posterior probability $p(y_g|\boldsymbol{x}_g)$ here is defined as the fraction of the realizations that fall into the high-z massive quiescent definition[6].

---

[6] In principle, one can find this probability utilizing the joint posterior distribution over the related parameters and finding the fraction of this joint distribution which falls into the selection criteria.

In order to find have a more representative libraries of galaxies, a SAM-based SED-fitting approach is performed, where the galaxies in the training set are treated as the libraries and fifty closest galaxies in the training set defined based on the $\chi^2 = \sum_{\text{band}} \left(\frac{F_{\text{test}} - F_{\text{train}}}{\sigma_{\text{test}}}\right)^2$ are found for every galaxy in the testing set. Then the weighted count of the number of high-z quiescent solutions based on the χ^2 is used to assign the probabilities. It should be noted that this method will have the advantage compared to the SED-fitting simply because the libraries are by design more closely follows the testing set.

Discriminant Analysis

In this section, we explain Linear Discrimination Analysis (LDA) and Quadratic Discriminant Analysis (QDA) to select the population of quiescent galaxies (Fisher 1936; Duda and Hart 1973; Hastie et al. 2009). In both methods, the likelihood is modeled as a Gaussian distribution:

$$p(\boldsymbol{x}_g \mid y_g = k) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}_g - \boldsymbol{\mu}_k)^T \boldsymbol{C_k}^{-1} (\boldsymbol{x}_g - \boldsymbol{\mu}_k)\right)}{(2\pi)^{\frac{d}{2}} \sqrt{\mid \boldsymbol{C_k} \mid}}$$

where $\boldsymbol{C_k}$ is a $d \times d$ covariance matrix of magnitudes for class k and $\mid \boldsymbol{C_k} \mid$ is its determinant:

$$\boldsymbol{C_k} = \frac{1}{N_k - 1} \sum_{g \in G_k} (\boldsymbol{x}_g - \boldsymbol{\mu}_k)(\boldsymbol{x}_g - \boldsymbol{\mu}_k)^T$$

$$\begin{cases} G_k = \{g \in G \mid y_g = k\} : \text{Set of galaxies in class } k \\ N_k = |G_k| : \text{Number of galaxies in class } k \end{cases}$$

where $\boldsymbol{x}_g$ is the magnitude vector for galaxy g, and $\boldsymbol{\mu}_k$ is the average magnitude vector for class k. Now the following posterior distribution is formed:

$$p(y_g = k \mid \boldsymbol{x}_g) = \frac{p(\boldsymbol{x}_g \mid y_g = k)p(y_g = k)}{\sum_i p(\boldsymbol{x}_g \mid y_g = i)p(y_g = i)}$$

which gives the probability of each class given the galaxy's magnitudes.

Assuming the same covariance matrix for all the classes changes the classifier to Linear Discriminant Analysis (LDA), which has a linear decision boundary:

$$\boldsymbol{C} = \frac{\sum\limits_k (N_k - 1)\boldsymbol{C_k}}{\sum\limits_k (N_k - 1)}$$

To ensure enough samples for the quiescent population, all the quiescent galaxies from realizations zero to five is used to calculate the quiescent population covariance matrix. However, only the realization zero is used for calculating the prior probabilities to be consistent with the rest of the classifier trained.

Naïve Bayesian

The Naïve Bayesian is another generative classifier with a likelihood function consisting of independent variables (Zhang 2004; Hastie et al. 2009)[7]. Here we find the colors calculated from the adjacent bands' magnitudes perform much better than the magnitudes themselves.

---

[7] One should note that the above assumptions put heavy constraints on the joint probability distribution of galaxy colors given its class. This is not a valid assumption in general since we know that the colors of galaxies are dependent. However, even with this naïve assumption on the independence structure of the colors given the class, one can build a useful/competitive classifier.

For a given galaxy with colors $\boldsymbol{c}_g = (c_1, c_2, \ldots, c_{d-1})$, the likelihood is the following:

$$p(\boldsymbol{c}_g \mid y_g) = \prod_{j \in \{1,2,\ldots,d-1\}} p(c_j \mid y_g)$$

assuming a Gaussian function over each color dimension normalized and centered by average μ_{jk} and standard deviation of σ_{jk} over color j of the class k ($y = k$), the posterior distribution over the galaxies class label is formed after multiplying the class prior and normalizing the posterior probability. Also, one can show that the decision boundary of the Gaussian naïve Bayes binary classifier is quadratic (similar to QDA) if different covariance matrices for each class are used similar to what is done here[8].

Here all the high-z quiescent candidates in the realizations zero to five are used for calculating the likelihoods, but for finding the prior probabilities, only the realization zero is used.

Bayesian Logistic Regression

Another way to model the Bayes classifier is to use a Bayesian Logistic Regression model. Using logistic regression, one can model the probability of interest using a linear combination of the variables. However, one can also introduce higher-order interaction terms between these variables by introducing two-way, three-way, and higher-order terms to capture non-linear structures.

Here we discuss the scaling of the input features, introductions of the higher term interaction, and finally, the full probabilistic model employed for finding the posterior pre-

---

[8] Assuming the same covariance matrices for both classes, results in the posterior with a logistic functional form with a linear decision boundary.

dictions for the testing data averaged over the posterior of the model parameters (averaging over uncertainties of parameters).

First, to make the computations more stable and the specification of the prior in our model easier, we transform the galaxy magnitudes within samples to have a mean of zero and a standard deviation of one in any given band.

Second, the following probabilistic model is created in order to identify higher most important second-order interaction between galaxies magnitudes at different bands (e.g., terms such as $x_{g,j}x_{g,l}$) using average F-score (harmonic mean of the purity and completeness[9]) for the quiescent galaxies across 5-fold cross-validation based on the predictions of the following model:

$$\beta_i \sim \mathrm{Normal}^{(i)}(0,1), \quad i = 0, \ldots d+1$$

$$\log(\frac{p_g}{1-p_g}) = \beta_0 + \sum_{i=1}^{d} \beta_i x_{g,i} + \beta_{d+1} x_{g,j} x_{g,l}$$

$$p(y_g \mid p_g) = \mathrm{Bernoulli}(y_g | p_g)$$

$\sim$ means that the random variable on the left side is sampled from the probability distribution on the right side. The p_g is the parameter of the Bernoulli trial.

For finding the optimal weights $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{d+1})$ in every cross-validation, the Maximum A Posteriori estimate (MAP) given the data is used. Because of the choice of the Gaussian prior, which is equivalent to classical logistic regression with l_2 regularization

---

[9] The purity is defined as the number of true quiescent galaxies selected by the classifier (true positive) to all the galaxies selected as quiescent (true positive + false positive). The completeness is defined as the number of truly quiescent galaxies selected by the classifier (true positive) to all the truly quiescent galaxies (true positive + false negative). The F-score is the harmonic mean of the purity and completeness which is $\frac{2 \times \mathrm{purity} \times \mathrm{completeness}}{\mathrm{purity} + \mathrm{completeness}}$.

on the parameters, one can find the MAP estimates using optimization techniques. Then the following prediction $\hat{y}$ is made on each validation set based on the MAP estimates:

$$\hat{y} = \arg\max_k \{p(y_g = k \mid \boldsymbol{x}_g, \boldsymbol{\beta}_{MAP})\}$$

Using this, we calculate the average F-score of the 5-fold cross-validation sets and identify those interaction terms that lead to a better F-score. We note that this approach could be less than optimal if there is a significant correlation between some of the interaction terms; however, we use this simple approach since the number of possible models can get combinatorially large. Introducing some interaction terms, mainly those that include bands near the Balmer break significantly increases the score. Therefore the fifteen most essential interaction terms $\{z_{g,i}\}_{i=d+1}^{p}$ (out of seventy eight possible second-order terms) are included in the following probabilistic model:

$$\beta_0 \sim \text{Student-t}(\nu = 5, \mu = 0, \sigma^2 = 4)$$

$$\beta_i \sim \text{Student-t}(\nu = 5, \mu = 0, \sigma^2 = 1), \quad i = 1, \ldots p$$

$$\log(\frac{p_g}{1 - p_g}) = \beta_0 + \sum_{i=1}^{d} \beta_i x_{g,i} + \sum_{i=d+1}^{p} \beta_i z_{g,i}$$

$$p(y_g \mid p_g) = \text{Bernoulli}(y_g | p_g)$$

where ν is the degree of freedom[10], μ and σ are the location and scale parameters, respectively. A Student's t-distribution for the prior probabilities over the regression coefficients is used following Gelman et al. (2008). These priors allow for larger absolute values of regres-

---

[10] $\nu = 1$ is the Cauchy–Lorentz distribution, and $\nu \to \infty$ is the Gaussian distribution.

sion coefficients than a Gaussian distribution but still have enough weight near zero that can lead to regularization of the regression coefficients where needed. Also, using the scaling only for the original inputs (i.e., magnitudes) and using the same prior across the regression coefficients, including the interactions, makes the effects of the interactions more minor on average compared to the original variables. The parameters of the priors are selected to be weakly informative on the transformed data scales (see Gelman et al. 2008, and Ghosh et al. 2018 for a thorough discussion on using different priors for logistics parameters).

`Stan` probabilistic programming language (Carpenter et al. 2017) is utilized to sample from the posterior distribution of the above model. `Stan` uses a Dynamic Hamiltonian Monte Carlo algorithm with no-U-turn adaptive samplers. Here four Markov chains are made according to the model above, each consisting of 4000 posterior samples after 2000 warm-up samples. The resulting posteriors are insensitive to the specific choice of the prior parameters, as long as they are not highly restrictive or have a hard constraint that excludes some possible value of the parameters, which is expected given a large number of examples for this model.

Using the posterior probability of the model parameters $\boldsymbol{\theta} = (\beta_0, \beta_1, \ldots, \beta_p)$ given the training data, a prediction can be made on the testing data by finding the predictive posterior distribution:

$$p(y_g|\boldsymbol{x}_g, \text{training set}) = \int d\boldsymbol{\theta} p(y_g|\boldsymbol{x}_g, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\text{training set})$$

where $p(\boldsymbol{\theta}|\text{training set})$ is the posterior distributions over all the parameters. In order to calculate the above expression, the average of the $p(y_g|\boldsymbol{x}_g, \boldsymbol{\theta})$ using the posterior samples is found, and, the predictive posterior probability of the testing data are calculated and used to measure the performance.

3.3.2 Approximating the True Classification Function

In the following subsection, instead of estimating the Bayes classifier by modeling the posterior distribution over the classes, the function h is directly learned from the data by empirical risk minimization using the following techniques:

- Random Forest of Decision Trees
- Support Vector Machines
- Feed Forward Neural Network

Random Forest of Decision Trees

Here we use a structure called Decision Tree (Breiman et al. 2017, for reviews see Murthy 1998; Loh 2011; Kotsiantis 2013; Loh 2014) that is trained to classify simulated galaxies in the Semi-analytic models.

When dividing the galaxies into sub-populations, a decision tree separates the population based on each feature's importance (i.e., observed magnitudes). The most important feature is found by computing the Gini impurity, Entropy (equivalently Information Gain criterion), or some other metric for deciding which band/feature is more effective in separating the sub-populations. This method is similar to imposing a threshold on the observed colors and magnitudes to separate the population. However, every decision is made based on the quantitative measure of how each decision at every step could separate the sub-populations more clearly.

The decision tree's basic structure divides the input (feature) space into small parts and allows us to find a simple model in each of these regions. In other words, it turns the problem of finding a global and often complicated decision function into the problem of finding many simple decisions for small and local regions in the data. To reduce the overfitting caused by using a single tree, we employ an ensemble method called Random Forest (Tin Kam Ho 1995; Breiman 1996; Amit and Geman 1997; Tin Kam Ho 1998; Breiman 2001; Hastie et al. 2009) that uses multiple decision trees to make the same decision. This is demonstrated to outperform results based on a single tree (Breiman 2001; Hastie et al. 2009).

In this approach, for every model decision tree, a random sample of galaxies in the training data is drawn with replacement (bagging or bootstrap aggregating; Breiman 1996) as well as some random subset of its bands (random subspaces; Tin Kam Ho 1998).

Here is the step-by-step process of implementing the Random Forest method:

1. Randomly select a sample of the galaxies in the training data (two-third of the training data with replacement) using bootstrap resampling. Then at every step we find the most important feature (j^*) in a subspace of bands, randomly chosen from full set of bands[11] and a threshold (t^*) from a set of possible magnitudes threshold $\{t_{ij}\}$ (which is at most as large as the number of the galaxies in our sample, and these possible thresholds are enumerated by subscript i) that splits the space into two regions defined

---

[11] The dimension of the random subspace of bands are set following the rule of thumb of $p = \lfloor\sqrt{d}\rfloor$ from Hastie et al. 2009, where p is dimension of the subspaces, and d is the dimension of the space.

as $R_1 = \{g \in G \mid x_{gj^*} \leq t^*\}$ and $R_2 = \{g \in G \mid x_{gj^*} > t^*\}$:

$$(j^*, t^*) = \arg \min_{j \in \{1,\dots,d\}, t \in \{t_{ij}\}} \{\mathcal{C}(R_1) + \mathcal{C}(R_2)\}$$

where the x_{gj^*} is the j^* th feature of the galaxy $g \in G$, and $\mathcal{C}$ is the cost function which for this study, we use entropy:

$$\mathcal{C}(R_1) = -\sum_{k \in \{0,1\}} \pi_k(R_1) \ln\left(\pi_k(R_1)\right)$$

$$\pi_k(R_1) = p(y = k \mid g \in R_1) = \frac{1}{|R_1|} \sum_{g \in R_1} \delta_{y_g,k}$$

where $\delta_{y_g,k}$ is the Kronecker delta function. And the same equation applies for R_2.

In the regression used for finding the missing values in section 3.2.4, the following cost/lost is used:

$$\mathcal{C}(R_1) = \frac{1}{|R_1|} \sum_{g \in R_1} (y_g - \bar{y})^2$$

$$\bar{y} = \frac{1}{|R_1|} \sum_{g \in R_1} y_g$$

The most important features and the corresponding thresholds (j^*, t^*) is found using these cost functions.

2. By doing the above splits consecutively, the tree is grown until any further split in a leaf (region) results in a leaf with the number of training data smaller than some value N_{min}. The resulting structure (tree $\mathcal{T}$) is one of the many models in our ensemble.

3. The new data points are classified by assigning the data to a leaf of the tree that corresponds to a class by going through the tree structure. This defines the classification function as $\hat{y} = \hat{h}_{\mathcal{T}}(\boldsymbol{x}_g)$. m trees are made following the same steps to get the Random Forest ($\{\mathcal{T}_i\}_{i=1}^{m}$).

Using m decision trees ensures that the final classifier does not depend on some superficial set of features or examples in the training data. The final estimator is calculated using the average of the classification decision made by each of these m trees, basically, what fraction of the decision trees classifies particular galaxies into specific sub-populations:

$$p(y_g = 1 \mid \boldsymbol{x}_g) = \frac{1}{m} \sum_{j=1}^{m} \hat{h}_{\mathcal{T}_j}(\boldsymbol{x}_g)$$

To find the optimal hyperparameters m and N_{min}, a grid search is performed according to the Table 3.2 and use a 5-fold cross-validation on the validation set. We find the $m = 200$, N_{min} to have optimal performance in terms of F-score of the classifier.

Table 3.2: Table of variables used for grid-search in various models for finding the best hyper-parameter for the model using the average of 5-fold cross-validation. m represents the number of decision trees used in a Random Forest, N_{min} is the minimum number of samples allowed at the end leaf of the tree structure. C controls the trade-off between the miss-classification permitted rate and the distance of the hyper plane's margins defined by support vectors. γ is the parameter that controls the width of the Gaussian kernel used in the Support Vector Machines.

Model	**Hyper-parameter**	**Values**
Random Forest	m	5, 10, 25, 50, 100, 200, 300, 400, 500
	N_{min}	1, 2, 5, 10, 20, 50, 100
Support Vector Machine	C	0.01, 0.1, 1, 10, 100, 500, 1000, 10000
	γ	0.0001, 0.001, 0.01, 0.1, 1, 10, 100

Support Vector Machines

There are several approaches to find the decision boundary necessary for classifying massive quiescent galaxies at high redshift. However, the true selection function in question could be highly non-linear. To capture full variability, one can search for decision boundaries in the higher dimensional space than the original magnitudes' space by creating a feature map that takes the input space into higher dimensions. Finding a linear decision boundary in this high-dimensional space (hyper-plane) is analogous to finding a curved decision manifold in lower dimensions. Using a kernel function (i.e., a similarity measure) of our data points in the lower dimension, we find the decision hyperplane in higher dimensions without explicitly mapping the data points to a higher dimension (for a review, see Cristianini et al. 2000; Hofmann et al. 2008).

Using a Support Vector Machine (Support Vector Machines), one defines the problem as finding a linear hyperplane that separates the two classes of the galaxies while maximizing the margin based on each class's closest instances to the hyperplane. However, since there is no guarantee that the classes are linearly separable even in the high-dimensional space, a method called soft-margin Support Vector Machines is used (Boser et al. 1992; Cortes and Vapnik 1995). This method can deal with linearly non-separable classes by adding a parameter for each instance called the *slack parameter* (ξ_g). The hyperplane for the binary classification can be found by solving the following optimization problem and forming the corresponding Lagrangian:

$$\mathcal{Q}(\boldsymbol{w}, \boldsymbol{\xi}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{g\in G}\xi_g$$

$$\begin{cases} \text{arg} \min\limits_{\boldsymbol{w}, w_0, \vec{\xi}} \mathcal{Q}(\boldsymbol{w}, \vec{\xi}) \\ \text{Subject to:} \\ 1 - \xi_g - (-1)^{1-y_g}(\boldsymbol{w}^T\boldsymbol{x}_g + w_0) \leq 0, \quad \forall g \in G \\ \quad -\xi_g \leq 0, \quad \forall g \in G \end{cases}$$

where ξ_g is the slack parameter that allows misclassification of data points, $(\boldsymbol{w}, w_0)$ determines the normal vector of the separating hyperplane $(\boldsymbol{x}, 1)$, and C is the regularization parameter which controls the trade-off between the importance of the misclassification versus larger margins for the hyperplane $(1/ \parallel \boldsymbol{w} \parallel)$[12]. In other word the above Lagrangian equation consist of two terms: $\boldsymbol{w}^T\boldsymbol{w}$ controls the size of the margins (lower the value, larger the margins) and $\sum_{g\in G}\xi_g$ is the allowed misclassification of some galaxies and C controls the importance of each term. For the points that are classified correctly but are inside the margin of the hyperplane, we have $0 < \xi_i \leq 1$, and for misclassified points, we have $\xi_i > 1$, and for correctly classified, we have $\xi_i = 0$, which is considered in the above optimization as the constraints. The Lagrangian of the problem is as the following:

$$\mathcal{L}(\boldsymbol{w}, w_0, \{\xi_g\}_{g\in G}, \{\alpha_g\}_{g\in G}, \{\beta_g\}_{g\in G}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{g\in G}\xi_g - \sum_{g\in G}\beta_g\xi_g$$
$$+ \sum_{g\in G}\alpha_g(1 - \xi_g - (-1)^{1-y_g}(\boldsymbol{w}^T\boldsymbol{x}_g + w_0))$$

---

[12] The equation $\boldsymbol{w}^T\boldsymbol{x} + w_0 = 0$ defines the hyper-plane and since one can always re-normalize the hyper-plane (one can multiply this equation by any scalar without changing the hyper-plane), and redefine the margins hyper-planes to be $\boldsymbol{w}^T\boldsymbol{x} + w_0 = \pm 1$. Therefore, the distance between these margins is $(2/ \parallel \boldsymbol{w} \parallel)$

in which $\beta_g, \alpha_g \geq 0$ are the Lagrange multipliers (or dual variables). The optimization problem becomes:

$$\min_{\boldsymbol{w}, w_0, \xi_g} \max_{\beta_g, \alpha_g \geq 0} \mathcal{L}(\boldsymbol{w}, w_0, \{\xi_g\}_{g\in G}, \{\alpha_g\}_{g\in G}, \{\beta_g\}_{g\in G})$$

This is the primal form of the problem however using the max-min inequality we can form the dual problem by switching the order of optimization:

$$\max_{\beta_g, \alpha_g \geq 0} \min_{\boldsymbol{w}, w_0, \xi_g} \mathcal{L}(\boldsymbol{w}, w_0, \{\xi_g\}_{g\in G}, \{\alpha_g\}_{g\in G}, \{\beta_g\}_{g\in G})$$

And one can find the optimization for the dual problem and since the primal optimization is a case of quadratic programming problems (quadratic optimization with linear constraints) and the Slater's condition holds, we have a strong duality which means that the solution to the primal problem is also the solution to the primal problem (duality gap is zero). Now we can find the dual optimization problem:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} &= \boldsymbol{w} - \sum_{g\in G} \alpha_g (-1)^{1-y_g} \boldsymbol{x}_g = 0 \\
\frac{\partial \mathcal{L}}{\partial w_0} &= -\sum_{g\in G} \alpha_g (-1)^{1-y_g} = 0 \\
\frac{\partial \mathcal{L}}{\xi_g} &= C - \alpha_g - \beta_g = 0
\end{aligned}$$

The optimal optimization problem can be rewritten using the above:

$$\mathcal{Q}(\{\alpha_g\}_{g\in G}) = \sum_{g\in G} \alpha_g - \frac{1}{2}\sum_{g\in G}\sum_{g'\in G} \alpha_g \alpha_{g'} (-1)^{1-y_g}(-1)^{1-y_{g'}} \boldsymbol{x}_g^T \boldsymbol{x}_{g'}$$

$$\begin{cases} \underset{\{\alpha_g\}}{\arg\max}\, \mathcal{Q}(\{\alpha_g\}_{g\in G}) \\ \text{s.t.} \quad \sum\limits_{g\in G} \alpha_g (-1)^{1-y_g} = 0 \\ \text{s.t.} \quad 0 \le \alpha_g \le C, \quad \forall g \in G \end{cases}$$

This is a quadratic programming optimization problem that can be solve using standard CVXPY python library (Diamond and Boyd 2016; Agrawal et al. 2018). After finding the solution we use the following equation to find the classification:

$$\hat{y} = \frac{1 + \text{sign}(w_0 + \boldsymbol{w}^T \boldsymbol{x}_g)}{2} = \frac{1}{2} + \frac{1}{2}\text{sign}(w_0 + \sum_{g'\in G} \alpha_{g'} (-1)^{1-y_{g'}} \boldsymbol{x}_{g'}^T \boldsymbol{x}_g)$$

A kernelized version of the optimization problem is used, which employs an implicit map to higher dimensional space prior to optimization in order to be able to find a non-linear decision boundary, using the following Gaussian kernel:

$$K(\boldsymbol{x}, \boldsymbol{x'}) = \exp(-\gamma |\boldsymbol{x} - \boldsymbol{x'}|^2)$$

where γ is the constant determining the width of the Gaussian kernel and $\boldsymbol{x}, \boldsymbol{x'} \in R^d$. Using the kernel changes the above linear decision boundary and classification estimate into the following (based on the solution of the optimization problem):

$$f(\boldsymbol{x}_g) = w_0 + \sum_{g' \in G} \alpha_{g'} (-1)^{1-y_{g'}} K(\boldsymbol{x}_{g'}, \boldsymbol{x}_g)$$
$$\hat{y} = \frac{1 + \text{sign}(f(\boldsymbol{x}_g))}{2}$$

where $\{\alpha_g\}_{g \in G}$ are the solutions to the dual problem formed from the constrained optimization problem.

In order to have a measure for the probability assigned by this solutions we use a heuristic technique proposed by Platt (1999) in which we assume $f(\boldsymbol{x}_g) = \log(\frac{p(y_g=1|\boldsymbol{x}_g}{p(y_g=0|\boldsymbol{x}_g})$ to have a logistic regression problem with the following posterior:

$$p(y_g = 1 \mid \boldsymbol{x}_g) = \sigma(Af(\boldsymbol{x}_g) + B) \tag{3.1}$$

where σ is the logistic function, and we can find the optimal value for A and B by maximum likelihood estimation on the validation set.

For the classification problem presented here, we find a Gaussian kernel to perform the best, and the hyperparameters are set to $\gamma = 0.1$ and $C = 100$ by 5-fold cross-validation on the training data. However, the final model, which uses the above parameters, uses all of the training data. Also, for finding the function f in the algorithm, we use all the candidates in the realizations zero to five, and since the algorithm itself is only sensitive to those critical points in which $\xi_g \neq 0$, this does not bias the results in terms of the massive quiescent and high-z population relative abundance. Therefore, by adding more candidates, we learn a more effective hyperplane that separates the classes. Also, for finding the constant A and

B, we use another 5-fold cross-validation on the training set, and we find an estimate for $f(\boldsymbol{x}_g)$ in which $\boldsymbol{x}_g$ are in the 5th set. We do this all possible five models, and then use the unions of all these estimates to find and make a set $\{f(\boldsymbol{x}_g) \mid \boldsymbol{x}_g \in G\}$. We find the best fit to the expression above (3.1) based on these values and the maximum likelihood estimates, which resulted in $A = -3.7$ and $B = -1.8$.

Feed-Forward Neural Network

The Feed-Forward Neural Network (Multi-layer perceptron), is a universal function approximator (e.g., Cybenko 1989; Hornik et al. 1990; Hornik 1991) that has been extensively used in a range of applications in astronomy from morphological classification to photometric redshift estimation and measurement of the physical parameters in galaxies.

The neural network is a multilayer perceptron (McCulloch and Pitts 1943; Rosenblatt 1958; Minsky and Papert 1969) that mimics the multilayered logistic regressions (if the activation function is logistic). In a fully connected neural network, we build a layered structure, and at layer i, we have a set of neurons with edges that connect them to every neuron in the previous $(i-1)$ and next layer $(i+1)$. We find the linear combinations of inputs to layer $i-1$, considering their assigned weight to every associated neuron to layer i. Then we apply a non-linear function (σ) to this value, which gives us the value for neurons in layer i, and we move through the network layer by layer. We start by assigning random weights and then passing the input through the network. We can find values in the last layer (output layer), and based on the task at hand, whether we have a classification or regression problem, we define a loss function that is calculated using the training data. The final value depends on every single weight in the network. One could use back-propagation to find the gradient of the cost function with respect to the weights to find the weights that minimize the function (e.g., Rumelhart et al. 1986; Hastie et al. 2009) in an iterative process. This means we can find the weights up to the last layer, and we update the weights layer by layer using gradient descent by moving in the opposite direction of the gradient of the cost function. At the final layer, we use the following Boltzmann distribution (also called softmax activation function) to convert the final layer output to probabilities:

$$p(y_g = 1 \mid \boldsymbol{x}_g) = \frac{\exp(z_1)}{\sum\limits_{k \in \{0,1\}} \exp(z_k)}$$

$$z_k = f_k(\boldsymbol{x}_g; \{w_{nm}, C_n, a_n\}_{nm})$$

where z_1, z_0 are inputs of the last layer of the network consisting of two neurons representing high-z quiescent classes and otherwise, respectively, and $\{f_k\}$ represent the function from the input layer to each output neuron of the network. w_{nm} is the weight of the input at n'th layer, for m'th neuron, $a_n m$ is the activation function at n'th layer, and C_n is the bias node for the n'th layer other than the output, which acts as a scalar value that offset the input of the activation functions in the next layer (similar to intercept term in linear regression), and is only connected to the next layer (opposed to other neurons which have both incoming and outgoing connections). We set this bias term to have a value of one, which can produce different offsets for the neurons' input in the following layer when multiplied by its associated weights to the next layer.

We use the cross-entropy loss function as:

$$J(w;\{(\boldsymbol{x}_g, y_g)\}_{g\in G}) = -\frac{1}{N_{\text{batch}}}\sum_{i=1}^{N_{\text{batch}}}\{y_i \log(p(y_i = 1|\boldsymbol{x}_i))$$

$$+(1-y_i)\log(1-p(y_i = 1|\boldsymbol{x}_i))\}$$

where N_{batch} is the number of training data used for calculating the cost function, and y_i is the true label of the galaxy. We divide the training set into several mini-batches, and we find the cost function for this subset. For the next update, another mini-batch is used until all the training data are used once.

To find the best structure for the neural network, we start with two hidden layers with 16 neurons at each layer, and we increase the depth of the network and the number of neurons until we do not see any significant change in the network's performance. Here we find the architecture with 12 Neurons in the first layer (dimension of the magnitudes' space), 32 Neurons in the second layer + one bias node, 16 Neurons in the third layer + one bias node, 16 Neurons in the fourth layer + one bias node, 16 Neurons in the fifth layer + one bias node, 2 Neurons in the output layer (corresponding to the binary outcome) to be sufficient.

All of the neurons except the last layer use the rectifier linear unit or "ReLu" activation function, defined as $\max(x, 0)$ for input x, which does not suffer from the vanishing gradient problem since its gradient is either zero or one opposed to other activation functions such as logistic and *tanh*. Furthermore, we use an adaptive stochastic gradient descent algorithm for minimizing the cost function called "Adam", which is a very effective method for stochastic optimization (Kingma and Ba 2014). We use the default configuration of Adam proposed in Kingma and Ba (2014). At each iteration, we use batches of 4096 examples from the data. We find that the validation loss function (a surrogate for the unknown true error rate) become stable after 150 iterations through the data. For building the neural network and training it, we use the `TensorFlow` python package (Abadi et al. 2015).

3.3.3 Dimensionality Reduction Techniques

Here we present dimensionality reduction techniques to approximate the structure in high dimensional magnitudes' space and convert it into lower-dimensional representations optimized to conserve the original space's structure. The lower-dimensional representations can be used to identify any class of galaxies or even estimate their physical properties. The dimensionality reduction techniques have shown to be effective for finding the physical properties of galaxies (e.g., Hemmati et al. 2019c; Davidzon et al. 2019) and photometric redshift measurements (i.e, Masters et al. 2019; Hemmati et al. 2019a).

In the following sections, we use the following algorithms to find a lower-dimensional representation of the data, and we train a classifier on these lower-dimensional embedding:

- Supervised Principal Component Analysis
- t-distributed Stochastic Neighbor Embedding
- Uniform Manifold Approximation and Projection

Supervised Principal Component Analysis

This section explores supervised principal component analysis (S-PCA; Barshan et al. 2011). PCA is used for finding the transformation that maximizes the variance of the data. S-PCA extends this idea to find the transformation that maximizes the statistical dependence of the transformed data and target variables, which could be of any dimension.

For finding the maximum dependence between transformed data and the target variable, S-PCA uses the Hilbert-Schmidt Independence Criteria (HSIC) (Gretton et al. 2005), which is a tool to find how independent two random variables are. HSIC is defined as the sum of the squared singular values of the cross-covariance operator.

This algorithm can be applied directly to the original data or their transformation to a Hilbert space (kernel S-PCA) for the ability to capture non-linear structures and allow a much more flexible notion of independence. The empirical definition of the HSIC for the sample of $\{(\boldsymbol{U}^T\boldsymbol{x}_i, y_i)\}_{i=1}^N$ in which $\boldsymbol{U}$ is the transformation matrix applied to the original input data $\{\boldsymbol{x}_i\}_{i=1}^N$:

$$\mathrm{HSIC}(\{(\boldsymbol{U}^T\boldsymbol{x}_i, y_i)\}_{i=1}^N, \boldsymbol{K}, \boldsymbol{L}) = (N-1)^{-2}\,\mathrm{tr}(\boldsymbol{HKHL})$$

in which $\boldsymbol{K}_{ij} = k(\boldsymbol{U}^T\boldsymbol{x}_i, \boldsymbol{U}^T\boldsymbol{x}_j)$ and $\boldsymbol{L}_{ij} = l(y_i, y_j)$ are kernels on the transformed input and target data respectively, also $\boldsymbol{H}_{ij} = \boldsymbol{\delta}_{ij} - N^{-1}$ is the centering matrix.

The supervised PCA algorithm maximizes the Hilbert-Schmidt Independence Criteria, or, in other words, maximizes the dependence of the transformed data ($\boldsymbol{U}^T\boldsymbol{X}$ in which $\boldsymbol{U}_{d\times d}$ is the transformation matrix and $\boldsymbol{X}_{d\times N}$ is the data matrix) and the target variable which could have any dimensions.

Using the following optimization and assuming $\boldsymbol{K} = \boldsymbol{X}^T\boldsymbol{U}\boldsymbol{U}^T\boldsymbol{X}$, the optimal transformation is found:

$$\begin{cases} \arg\max\limits_{\boldsymbol{U}}\{\mathrm{tr}(\boldsymbol{HKHL}) = \mathrm{tr}(\boldsymbol{U}^T\boldsymbol{XHLHX}^T\boldsymbol{U})\} \\ \text{Subject to: } \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I} \end{cases}$$

This optimization problem and kernelized version can be solved analytically using its Lagrangian, which yields a closed-form solution (p largest eigenvectors of the $\boldsymbol{XHLHX}^T$, for p dimensional representation of the data). We use the kernelized version of this algorithm, in which we use an implicit map to a Hilbert space, similar to what we used in Support Vector Machines, before finding the optimal transformation. For the kernel on the input variables, we use a Gaussian kernel with $\gamma = 0.1$, and on the binary target variable, we have chosen the following linear kernel:

$$L(\boldsymbol{y}, \boldsymbol{y}') = \boldsymbol{y}^T\boldsymbol{y}'$$

where $\boldsymbol{y}$ is the vector of labels in the dataset.

Finding the kernel and eigenvectors can become computationally expensive as the number of data points increases. In order to circumvent this problem, we use the Nyström approximation for both calculating the kernel and finding the eigenvectors for constructing the transformation $\boldsymbol{U}$ (i.e. Eckart and Young 1936; Williams and Seeger 2001; Bach and Jordan 2005; Drineas and Mahoney 2005; Nemtsov et al. 2013). This approximation uses the reasonable assumption that the rank of the full kernel matrix $\boldsymbol{K}^{(N\times N)}$ is much smaller than N, which allows us to approximate the full matrix without having to use $N \times N$ kernel

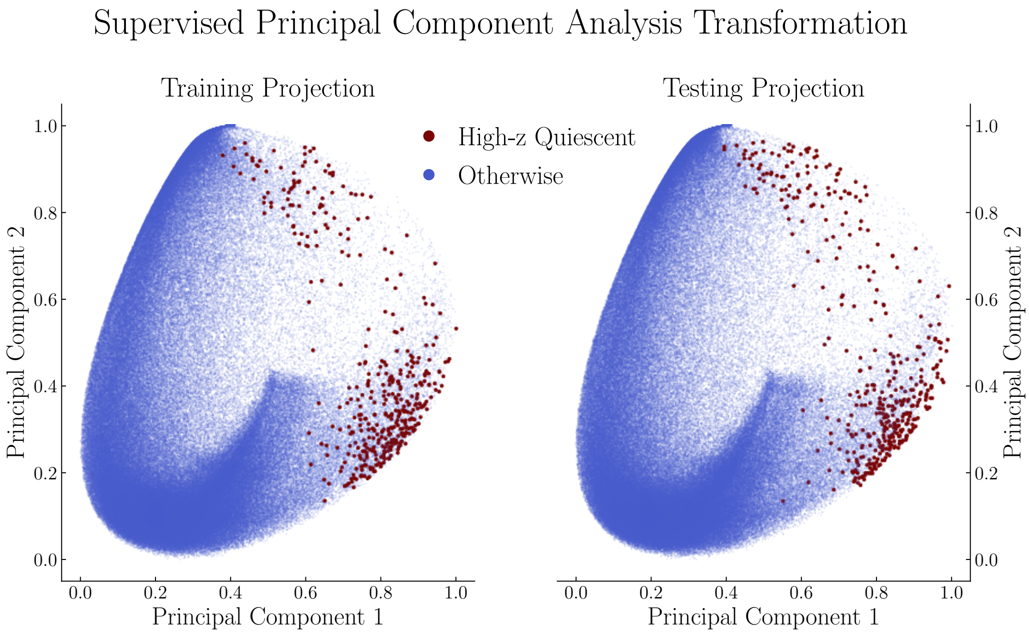


Figure 3.2: Embedding of the training and testing data set under the Kernel Supervised PCA transformation into two dimensions.

matrix. For the problem presented here, we use the $m = 5000$ randomly chosen data points (we only need $m \times N$ values of the original matrix to approximate the full $N \times N$ matrix). We use this approximation to calculate the kernel and solve the eigenvector problem using the singular value decomposition (SVD).

Figure 3.2 shows the two-dimensional representation of the data using S-PCA, colored by whether they are part of the target population. As seen in the figure, the S-PCA tries to maximize the dependence between magnitude representation and the class labels of galaxies and correctly separate many of the galaxies with the correct trend, in which the galaxies on the left side of the embedding are not high-z quiescent galaxies. However, many galaxies near every high-z quiescent galaxies are not part of this population, leading to contaminants in a classifier's predictions trained on this space.

t-distributed Stochastic Neighbor Embedding

The t-SNE technique can separate quiescent and dusty star-forming populations efficiently (Steinhardt et al. 2020). The basic idea here is to define a similarity measure on the original data space and then optimize some distribution on a lower-dimensional space that conserves that similarity measure as much as possible. This is a variation of the Stochastic Neighbor Embedding technique (Hinton and Roweis 2003). In the t-SNE, the similarity measure for any pair of $(\boldsymbol{x}_i, \boldsymbol{x}_j) \in R^d$ is defined as:

$$p_{j|i} = \frac{(1-\delta_{ij})\exp(-\frac{|\boldsymbol{x}_i - \boldsymbol{x}_j|^2}{2\sigma_i^2})}{\sum_{k\neq i}\exp(-\frac{|\boldsymbol{x}_i - \boldsymbol{x}_k|^2}{2\sigma_i^2})}, \quad p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N}$$

which defines the asymmetric probability measure $p_{j|i}$ that the $\boldsymbol{x}_i$ has $\boldsymbol{x}_j$ as its neighbor, N represent the number of data points, and the p_{ij} is the symmetric version of this probability measure. Here σ_i is the standard deviation of the Gaussian distribution around the point $\boldsymbol{x}_i$, which is set by finding the value that makes the entropy of the above distribution equal to $\log k$ with k defined as the "effective neighbors" or "perplexity". The "perplexity" is one of the hyperparameters of this technique and, based on its definition above, controls the trade-off between preserving the global representation versus the local representation of the data[13].

In its original version, the same probability distribution is defined on the low-dimensional representation, and the standard deviation of the Gaussian distribution is kept the same globally. However, in student's t-distributed Stochastic Neighbor Embedding (Maaten and Hinton 2008), the similarity measure of any pair of $(\boldsymbol{z}_i, \boldsymbol{z}_j) \in R^p$ in the low

---

[13] Higher perplexity results in a more global picture of the data compared to the lower values that tends to preserve the local structures of the data

dimensional representation is defined by student's t-distribution with one degree of freedom (i.e., Cauchy–Lorentz distribution):

$$q_{ji} = \frac{(1 + |\boldsymbol{z}_i - \boldsymbol{z}_j|^2)^{-1}}{\sum_{k \neq i}(1 + |\boldsymbol{z}_i - \boldsymbol{z}_k|^2)^{-1}} \tag{3.2}$$

The above, heavily tailed student's t-distribution of the lower dimensional embedding allows more accessible volume in the space, which allows the dissimilar data points in the higher dimensional space to be separated more efficiently in the final representation.

Since we define the similarity measure in terms of the probability distributions, we can define an optimization problem based on the distance between these two probability distributions. The cost function used here is the Kullback-Leibler (KL) divergence[14], also called the relative entropy, between the joint distribution in the original space P and the joint distribution in the low dimensional space Q, which is defined as follows and needs to be minimized:

$$\mathcal{C}_{\text{t-SNE}} = KL(P||Q) = \sum_{i,j} p_{ij} \log(\frac{p_{ij}}{q_{ij}})$$

The above optimization problem can be solved by finding the gradient of the cost function with respect to all $\boldsymbol{z}_i$. Here, we use the `openTSNE` python package developed in Poličar et al. 2019 to find the best transformation, with 600 iterations for optimization and perplexity of 15 for t-SNE. Choosing lower values means one wants to preserve the local structure versus larger values in which more of the global structure is preserved. Here the

---

[14] One feature of this embedding understood based on the asymmetric nature of the KL divergence. Here the optimization tries to find the optimal solution that keeps the original space's similarities' structure but might introduce artificial similarities in the lower dimensional space that do not represent the original space's intrinsic structure.

perplexity is set to maximize the F-score of the classifier trained to separate the high-z massive quiescent population. The new data is transformed based on a learned embedding without running the entire optimization problem (only calculate and optimize probabilities for the new data points concerning the learned embedding without changing the learned embedding).

Figure 3.3 shows the resulting two-dimensional embedding of the training and testing set projected onto the embedding, colored with different datasets' physical properties. There are locations with a higher probability of finding the target galaxies, which allows the construction of a classifier. The physical properties have relatively structured distribution over the embedding, which shows that this embedding is learning some of the underlying structure of the data and can be used to estimate physical properties such as redshift, stellar mass, and star-formation rates.

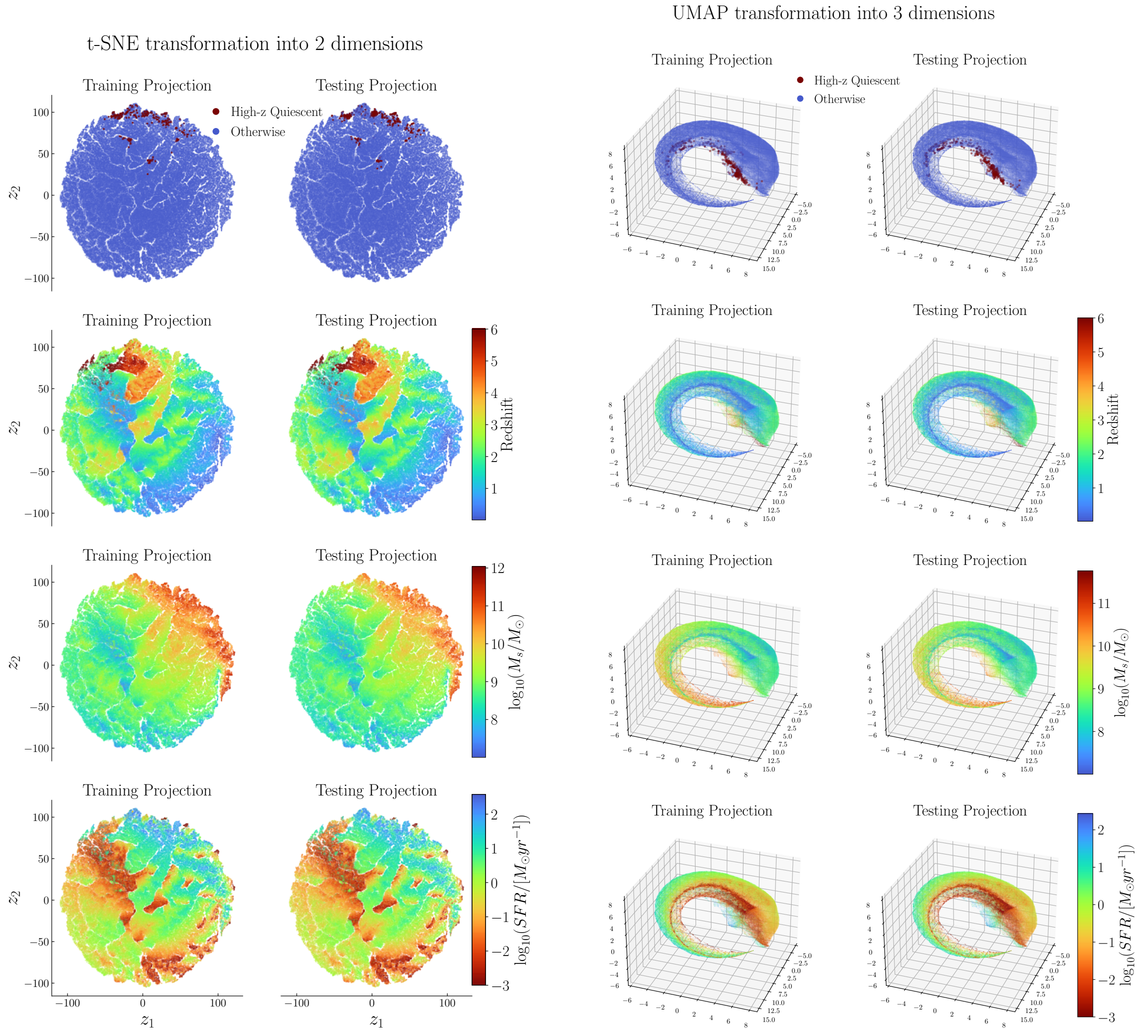
Uniform Manifold Approximation and Projection

Another dimensionality reduction technique is the Uniform manifold approximation and projection (UMAP) (McInnes et al. 2018). UMAP finds an approximation of the manifold that represents the data in $\boldsymbol{x}_g \in \mathrm{IR}^{\mathrm{d}}$ at every point (local approximation) and, based on this, builds a local fuzzy simplicial set[15], where every simplex has an associated measure from zero to one (Figure 3.4 shows a depiction of such construction).

A similar simplicial set is constructed over the lower-dimensional representation, assuming a manifold for the lower-dimensional representation of the data. The dimension-

---

[15] Simplicial set are combinatorial objects which can represent (model) the topological structure of the data



(a) Embedding of the training and testing data set under the t-SNE transformation with the perplexity of 15 into two dimensions. The x-axis and y-axis are embedding components of the data trained.

(b) Representation of the training and testing data set under the UMAP transformation (with 20 nearest neighbors graph) into three dimensions. The axes are embedding components of the data trained.

Figure 3.3: Dimensionality Reduction using t-SNE, and UMAP

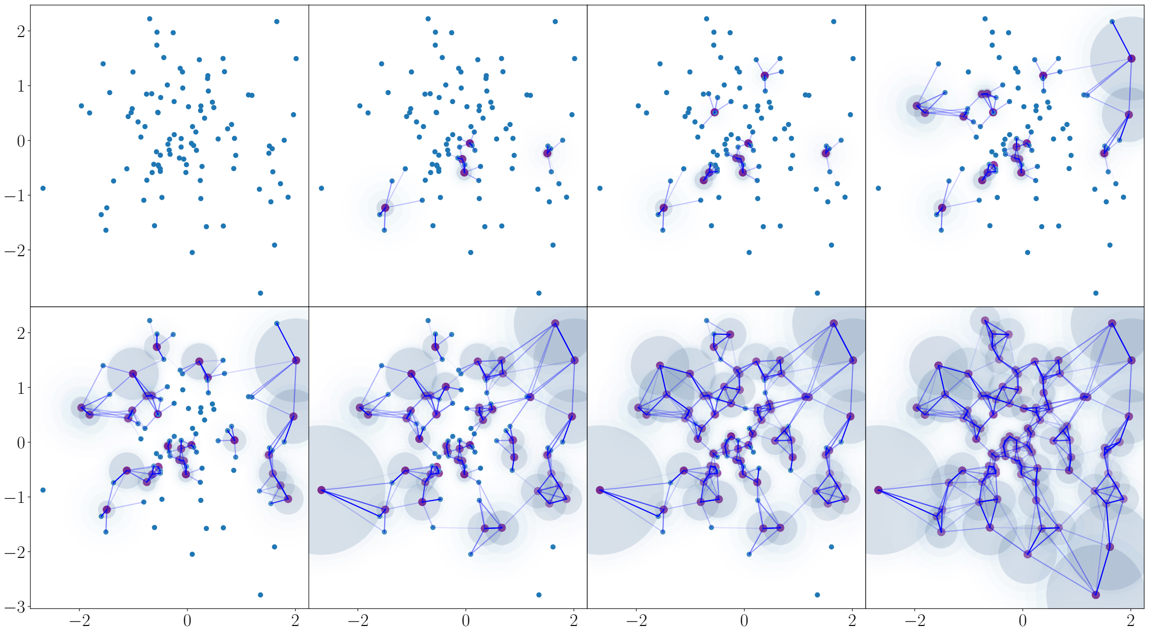

Construction of the weighted graph

Figure 3.4: Shows the construction of graph in UMAP in 2-dimension and how the position of each point affect its sphere of influence. The upper left figure shows the original data points, and as we move across the figures, we show the neighborhood graph of more data points. The opacity of each edge in the graph is the based on the fuzzy union of the fuzzy simplicial sets on the data defined based on the 5 nearest neighbors. The spheres around each point is based on the value of σ solved for that point and the decrease in its opacity shows the decrease in the similarity measure.

ality reduction aims to minimize the cross-entropy between the topological structure in the high and low dimensions.

UMAP assumes a uniform distribution of the data over the manifold with its intrinsic (unknown) metric to find the manifold's best local approximation. Furthermore, this assumption implies that there should be the same number of data points around any given data points within the same volumes (defined based on the intrinsic metric). Using this and choosing k'th nearest neighbor of each point, a geodesic distance from the data $\boldsymbol{x}_g$ to its neighbors can be approximated. However, this implies a family of N (number of data points) metric spaces, which then can be combined using the fuzzy simplicial sets. This is done by making a fuzzy simplicial set for every metric space and finding their resulting fuzzy set union[16], which results in a final fuzzy simplicial set representation of the data. Moreover, similar to t-SNE, a similarity measure, which is the local fuzzy simplicial set membership, can be defined based on the metric:

$$p_{j|i} = \exp(\frac{\rho_i - d_i(\boldsymbol{x}_j)}{\sigma_i}), \quad p_{ij} = p_{i|j} + p_{j|i} - p_{i|j}p_{j|i}$$

in which the asymmetric probability measure $p_{j|i}$ is the membership of the outgoing graph edge from the point i to j, $d_i(\boldsymbol{x}_j)$ is the distance between $\boldsymbol{x}_j$ from point i (metric defined based on the point i), ρ_i is the distance of point i from its nearest neighbor, and σ_i is set similar to the perplexity defined in t-SNE. Furthermore, the similarity in the low-dimensional representation is defined:

$$q_{ji} = (1 + a|\boldsymbol{z}_i - \boldsymbol{z}_j|^{2b})^{-1}$$

where $\boldsymbol{z}_i, \boldsymbol{z}_j$ are the lower-dimensional representation of each point, and a, b are the positive and constant hyperparameter, which we set to $a = 1.4$ and $b = 0.9$ for finding the resulting embedding. If we assume $a = 1, b = 1$, we have the student's t-distribution used in t-SNE.

Assuming only 1-skeleton for the fuzzy simplicial set representation, which means considering a graph with membership grade for its edges, high and low dimensional representation is reduced to fuzzy sets of graph edges. Given these fuzzy sets, the following

---

[16] probabilistic t-conorm

cross-entropy cost function is defined and minimized:

$$\mathcal{C}_{\mathrm{UMAP}} = \sum_{i \neq j} \{p_{ij} \log(\frac{p_{ij}}{q_{ij}}) + (1 - p_{ij}) \log(\frac{1 - p_{ij}}{1 - q_{ij}}))\}$$

The first term in the cost function is very similar to the KL divergence, used in t-SNE, representing the clustering of similar points. However, the second term, missing from the t-SNE cost function, introduces cost for making the originally dissimilar points similar in the lower-dimensional embedding. Also, similar to t-SNE, the optimization problem can be solved using gradient descent algorithms. The algorithm described above is unsupervised. However, it can be turned into a supervised (or even semi-supervised) by adding a term that describes the simplicial set over the target variables. Here we use the unsupervised version, with $k = 20$ to build the k nearest neighbor graph of the data using the Euclidean metric. We use the `UMAP` python package (McInnes et al. 2018) to find the three-dimensional embedding of the data with Euclidean metric on the three-dimensional space with 500 iterations for the optimization problem.

Figure 3.3 shows the resulting three-dimensional embedding of the training and testing set projected onto the embedding, colored with different physical properties of the galaxies. Similar to t-SNE, we find structures based on the values of different physical properties, which shows the utility of this embedding in estimating these physical properties.

K-Nearest Neighbors Classifier

To classify galaxies in the lower dimensional embedding produced from S-PCA, t-SNE, and UMAP, we use the K-Nearest Neighbor (KNN) algorithm[17] on these representations of the original data. Based on the fraction of these k neighbors that are high-z massive quiescent galaxies in the training set, a probability is assigned to data from the testing set. Also, to find the optimal number of nearest neighbors, we use 5-fold cross-validation on the validation set, assuming a Euclidean metric for all of the lower dimensional representations. $k \sim 50$ was found to be optimal based on the F-score for the high-z massive and quiescent class of galaxies. The notion of dissimilarity/distance on the embedding depends on the hyperparameters used and is not usually well defined. For example, one of the main drawbacks of the t-SNE algorithm is that the cluster size and distances between each cluster are not a well-defined notion (Wattenberg et al. 2016). However, as the embedding results from Figure 3.3 show, the compression of the galaxy magnitude space is approximately continuous and is not separated into multiple clusters. Therefore, we assume that the KNN classifier can be applied since we do not deal with multiple clusters.

3.4 Results

After training the statistical procedures on the semi-analytic model, their performances are compared on the test set using multiple criteria. The first criterion used is the *Receiver Operating Characteristic* (ROC) curve, which is defined as an empirical curve

---

[17] The KNN classifier on the lower-dimensional embedding is chosen because of its simplicity based on the distance measure and its similarity to the t-SNE and UMAP algorithms way of building the embedding; however, one can choose different classifiers which can have different performances.

showing the True positive rate (those with correct classification as quiescent) dependence on the False positive rate (those classified as quiescent but are not). This curve shows how many of the actual quiescent galaxies we can recover at the expense of having falsely classified galaxies. Since the best-case scenario is that our classifier is always selecting all the quiescent galaxies at any false positive rate, we expect better models to have a higher area under the ROC curve (ΣROC).

According to this measure, all the statistical methods used outperform the SED-fitting on the complete and incomplete data. Figure 3.5 shows the ROC curves for different models. It should be noted that in a problem in which one class of galaxies is infrequent compared to the other class (very imbalanced), this curve is not going to distinguish between models since the false positive rate can be small compared to the total number of galaxies, but the number of misclassified galaxies could be much higher than those classified correctly.

To distinguish the models further than their performance on the ROC curve, we look at the *Purity-Completeness curve*[18] of the quiescent population. The purity is defined as the number of true quiescent galaxies selected by the classifier (true positive) to all the galaxies selected as quiescent (true positive + false positive). Completeness is defined as the number of truly quiescent galaxies selected by the classifier (true positive) to all the truly quiescent galaxies (true positive + false negative). This performance measure shows the trade-off between purity versus completeness of the sample. The perfect scenario is to have the highest purity at all the completeness rates.

We find that the Gaussian Naive Bayes and LDA classifier perform worse than SED-fittings on the incomplete data. However, the uncertainty in redshifts can change this

---

[18] precision-recall curve

conclusion since we only focused on the SED-fitting procedures at a fixed and accurate redshift. The KNN classifiers trained on the lower dimensional embedding of the data showed very similar results in terms of their performance based on the purity-completeness and ROC curves to SED-fitting on incomplete data other than S-PCA, which performed worse based on the purity-completeness curve. However, the SED-fitting on the complete data performs better than these methods. We also find that the QDA classifier performs somewhere between the SED-fitting on the complete and incomplete data sets. Bayesian Logistic Regression, SAM-based SED-fitting, Random Forest, Support Vector Machines, and neural network classifier perform better than both SED-fitting procedures reflected from the area under purity-completeness and ROC curves are shown in Table 3.3.

As shown in the left panel of Figure 3.5, the SED-fitting on the complete data can be a better classifier than Bayesian Logistic Regression at very high completeness rates (about 0.8), but it fails at other regimes. Overall, we find that the Random Forest, Neural network, and Support Vector Machines classifiers have higher purity than the SED-fitting on complete data with fixed and accurate redshift at almost every completeness rate. The Bayesian Logistic Regression classifier has better purity than the SED-fitting on incomplete data up to the completeness rate of 0.8[19].

Figures 3.6, and 3.7 show the result of the classification on the SFR versus stellar-mass plane and UVJ plane respectively. The SED fitting results on the complete data (i.e., no non-observed values) show that the SED-fitting can correctly classify most of the

---

[19] Given that there are uncertainties in determining a redshift for a galaxy survey, we would suspect that Bayesian Logistic Regression, Random Forest, Support Vector Machines, and neural network classification widen their superiority over SED-fitting that uses generic libraries of SEDs. On the other hand, a reliable training set becomes essential for these models.

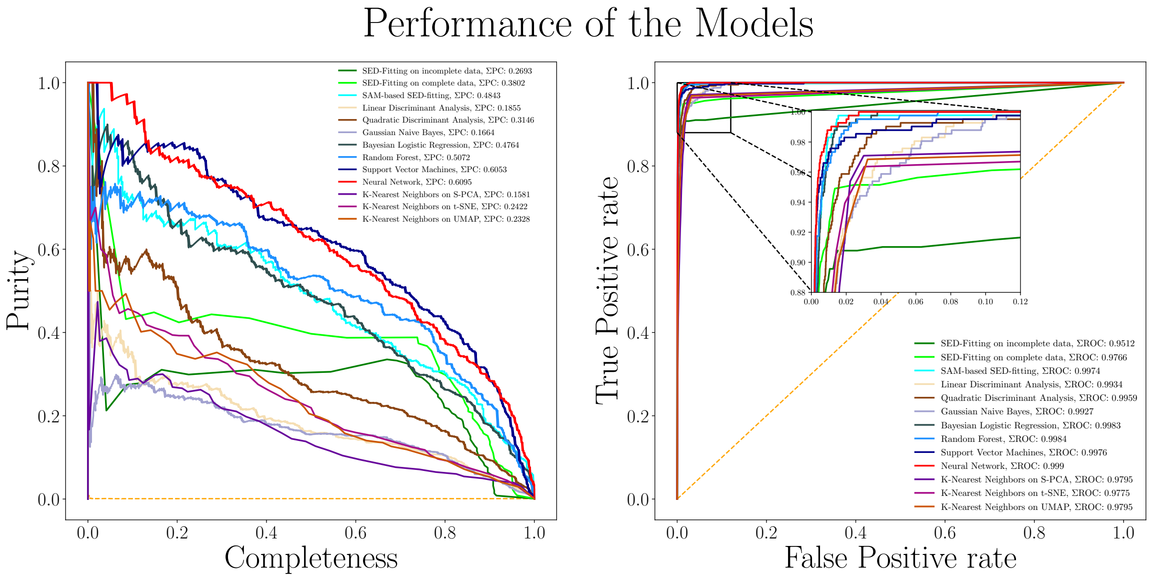


Figure 3.5: The purity-completeness curve on the left and Receiver operating characteristic (ROC) is on the right. The figure inside the ROC curve is zoomed in to distinguish the models better. As both figures show, some of the statistical models used here are superior compared to SED-fitting. The dashed orange line is a random classifier that assigns the labels at random. The ΣPR and ΣROC are the areas under purity-completeness and ROC curves, respectively, and for both figures, the higher values translate roughly into a better model. However, for a specific application, one can select the model based on the performance at a certain level (e.g., the best model with the highest purity at a completeness rate of 80%). As the figure suggests, the Neural Network, Support Vector Machines, Random Forest, and Bayesian Logistic Regression have the best performance and outperform the generic SED-fittings-based classification, with Bayesian Logistic Regression and Random Forest matching the performance of the SAM-based SED-fitting classification procedure.

target galaxies; however, many candidates are far from the decision boundary, including some dusty star-forming galaxies, and are classified as the high-probability objects, which introduces many contaminants into the sample. In SED-fitting on the incomplete data, SED-fitting misses a higher fraction of the high-z quiescent galaxies and suffers from more contamination.

The SAM-based method improves this shortcoming to a large degree. It can remove the majority of the contaminant but still misclassify some of the galaxies and, for many of the target population, underestimate the probability of being from the target population. The out-performance of the SAM-based method reveals the importance of using a more diverse set of SED libraries and the information from the population distribution of the galaxies that are not shared with the generic individual-based SED-fitting.

LDA does a reasonable job in classification compared to the SED-fitting given the simplicity of its assumption and the fact that it has a linear decision boundary in the magnitude space. However, the LDA has a high error rate for very dusty galaxies apparent on the UVJ plane (redder $U - V$ and $V - J$ colors). The QDA model does not show such behavior and can distinguish dusty versus quiescent galaxies better. Both of these models have contaminants that are relatively close to the boundaries compared to classification based on the SED-fitting results; however, they miss many of the target galaxies near the boundaries.

Gaussian Naïve Bayesian classifier can assign non-zero probabilities to the target galaxies in the testing data set; however, there is significant contamination from the dusty star-forming population, as seen in the UVJ plane, similar to the behavior of the LDA.

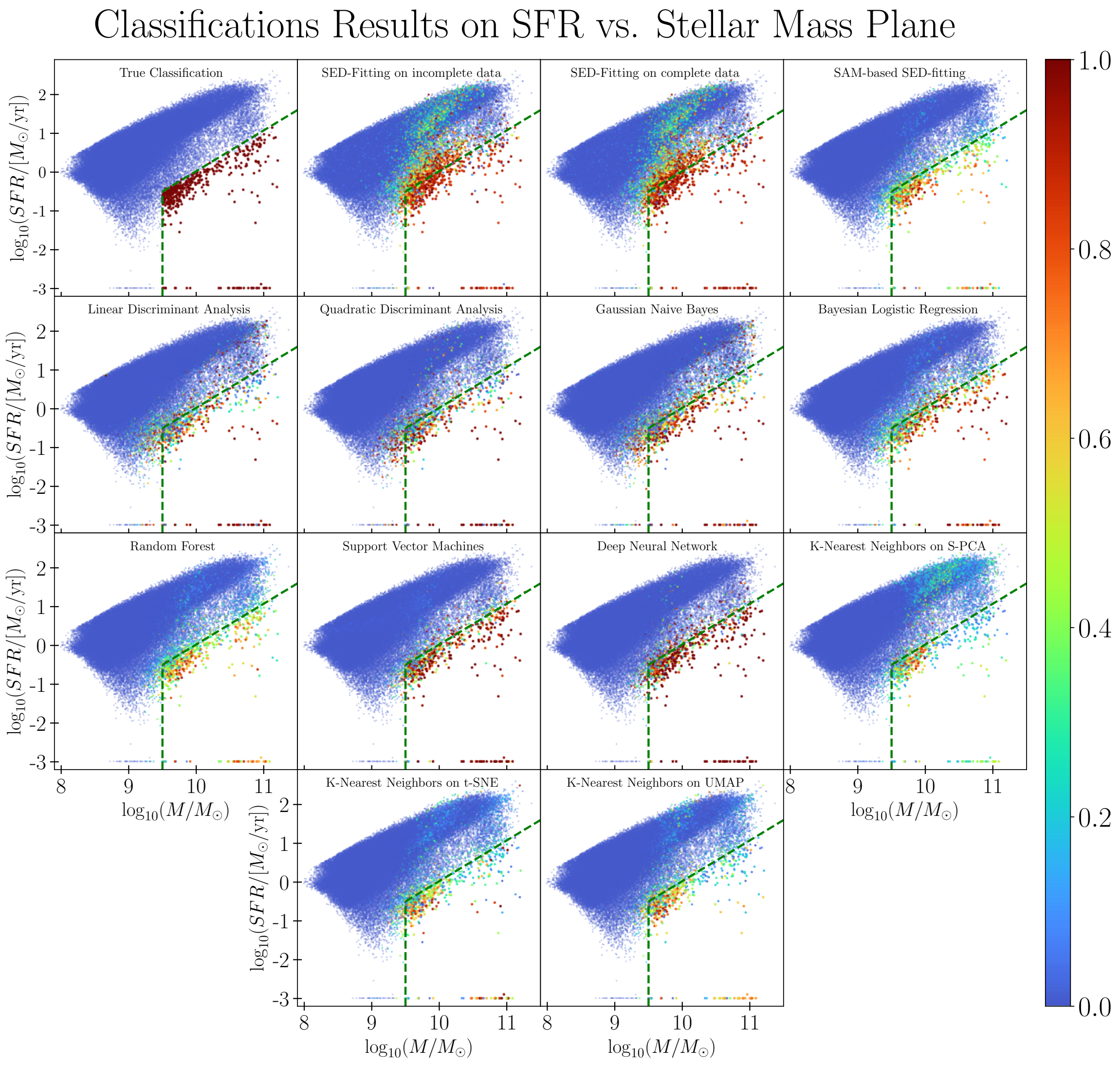


Figure 3.6: Classification performance of the models on star-formation rate vs. stellar mass plane for the galaxies with $z \geq 2.5$ in the test set. The color bar shows the probability of being a high-z quiescent. The boundaries shown with green dashed lines in the SFR vs. stellar mass plane are from the imposed criteria for selecting the sample defined in §3.3. The first figure on the top left shows the true label of the galaxies in the testing set.

Overall, both SED-fitting approaches are superior in selecting the target galaxies compared to the Naive Bayes. However, one should note that the Gaussian Naïve Bayesian classifier is a simplistic model but still can capture some of the underlying structures of the actual posterior distribution.

The Bayesian Logistic Regression, Random Forest, Support Vector Machines, and Neural Network can separate the massive quiescent galaxies quite well. The classification probabilities for Bayesian Logistic Regression and Random Forest continuously change from low to high when moving towards the region of interest in the UVJ and SFR vs. M_s planes, which shows more uncertainty near the selection boundary as expected. Also, there is a significant reduction in contaminants from the star-forming population compared to SED-fitting and other approaches studied here.

In order to further analyze the performance of the most optimum models compared to the SED-fitting based procedures, the purity-completeness of different models as a function of redshift and stellar mass are shown in the Figure 3.8. According to the top panel, the overall trends in the redshift show that the models are performing worse when we go to the higher redshift. However, the performance of the SED-fittings for the highest redshift bin does not drop significantly for completeness rates of less than 0.5, which some of it might be explained by availability of the accurate redshifts during the SED-fitting. On average, the models have more trouble in the highest redshift. This is expected since the higher redshifts sources are systematically dimmer, and have lower number densities (i.e., lower representations in the sample), and the effect of the uncertainties and censoring becomes more significant.

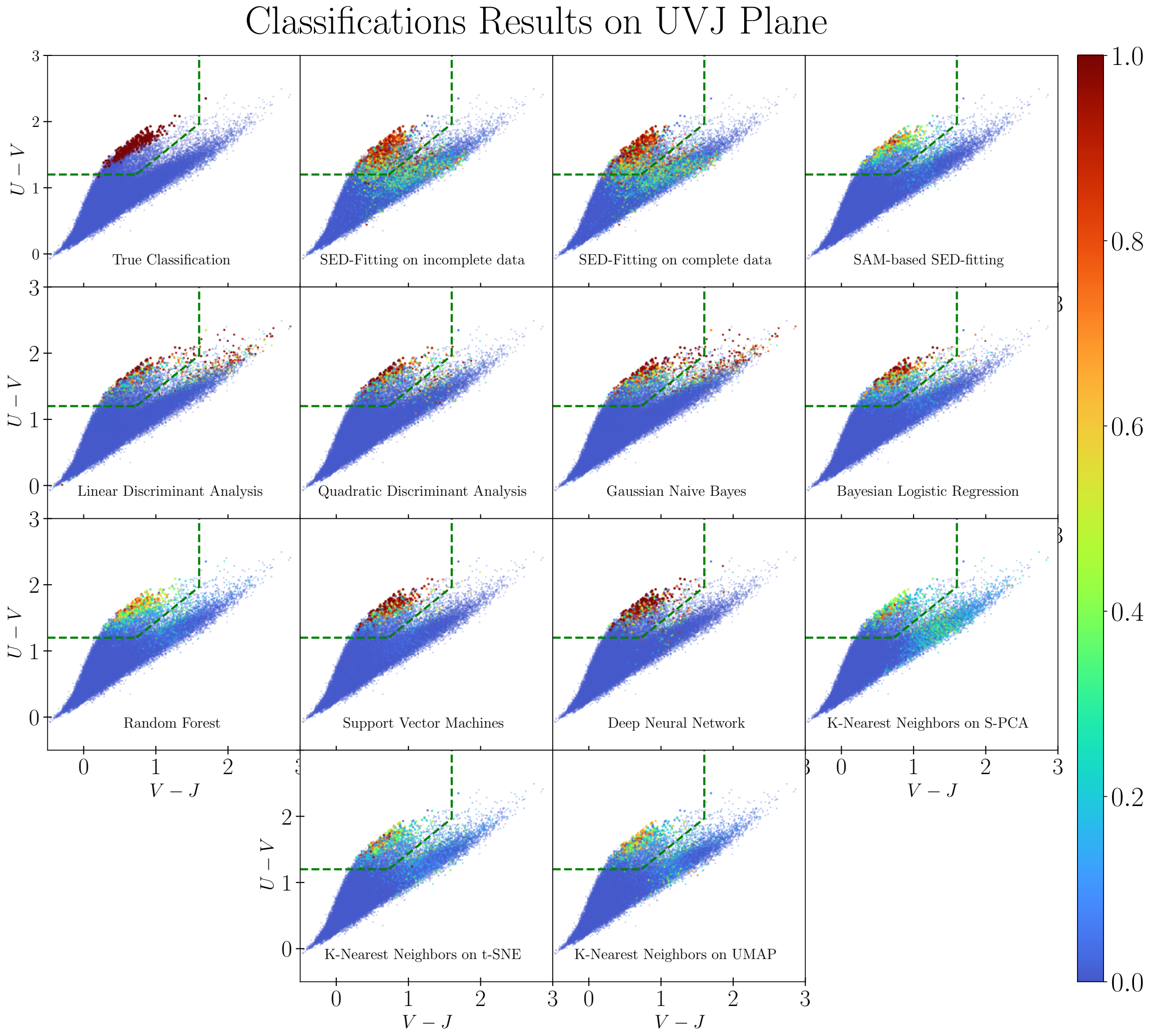


Figure 3.7: Classification performance of the models on the UVJ plane for the galaxies with $z \geq 2.5$ in the test set. The color bar shows the probability of being a high-z quiescent. The boundaries shown with green dashed lines on the UVJ plane are from selection criteria in Straatman et al. (2014). The first figure on the top left from the left shows the true label of the galaxies in the testing set.

A similar trend is seen in the bottom panel of Figure 3.8 showing the purity-completeness as a function of stellar mass, where the lower stellar masses are more difficult to predict even though there are more numerous compared to the highest mass bins. The SED-fitting approach works quite well on the galaxies with stellar masses of more than $\log_{10}(M_s/M_\odot) > 10.5$, only slightly worse than Bayesian Logistic Regression, Support Vector Machine and Neural Network performance, but for the lower mass objects performance significantly drops and makes the overall performance of the SED-fitting much worse. This again points us to the importance of the flux, flux uncertainties, and censoring for the dimmer objects which are much more numerous than the very massive objects.

In order to find the different classifiers ability to recover the overall properties of the target population, we investigate their predictions for the quiescent fraction as a function of redshift and stellar mass for galaxies with a stellar mass larger than $10^{9.5} M_\odot$ and $2.5 \leq z \leq 3.75$. The top panel in Figure 3.9 shows the quiescent fraction as a function of redshift for a given completeness level. This fraction is compared to what is expected from the true fraction in the testing set and the variations across different SAM realizations. For completeness lower than 0.5, the models underpredict quiescent fraction, and at the same time, the SED-fitting overpredict this value which reflects the lower purity of the classification at given completeness. For completeness higher than 0.6, all the methods over predict the quiescent fraction, which is expected from the lower purity of the resulting sample at high completeness.

The lower panel in Figure 3.9 shows the quiescent fraction as a function of stellar mass for a given completeness level. For completeness level lower/higher than 0.4/0.6, all the methods under/over predict the quiescent fraction. However, for the completeness between 0.4-0.6, all the methods reasonably predict the quiescent fraction. The SED-fitting on the incomplete data overpredicts this fraction at lower masses.

We note that the resulting samples from different models have different purity levels at given completeness, so the resulting quiescent fractions include those contaminations as well. However, on average, for completeness level of around 0.5 for Bayesian Logistic Regression, SAM-based SED-fitting, Random Forest, Support Vector Machines, and Neural Network can recover the behavior of the quiescent fraction as a function of redshift and stellar mass.

Table 3.3: Performance of the models based on the area under the Receiver operating characteristic (ΣROC), and purity-completeness (ΣPC) curves. The best four performances on each measure are identified in bold.

Model	**ΣPC**	**ΣROC**
Random Classifier	0.001	0.5
SED-Fitting with missing values	0.2569	0.9731
SED-Fitting w/o missing values	0.4193	0.9948
SAM-based SED-Fitting	**0.4843**	0.9974
Linear Discriminant Analysis	0.1855	0.9934
Quadratic Discriminant Analysis	0.3146	0.9959
Gaussian Naïve Bayesian	0.1664	0.9927
Bayesian Logistic Regression	0.4764	**0.9983**
Random Forest	**0.5072**	**0.9984**
Support Vector Machines	**0.6053**	**0.9976**
Feed Forward Neural Network	**0.6200**	**0.9990**
k-Nearest Neighbors on S-PCA	0.1680	0.9783
k-Nearest Neighbors on t-SNE	0.2422	0.9775
k-Nearest Neighbors on UMAP	0.2328	0.9795

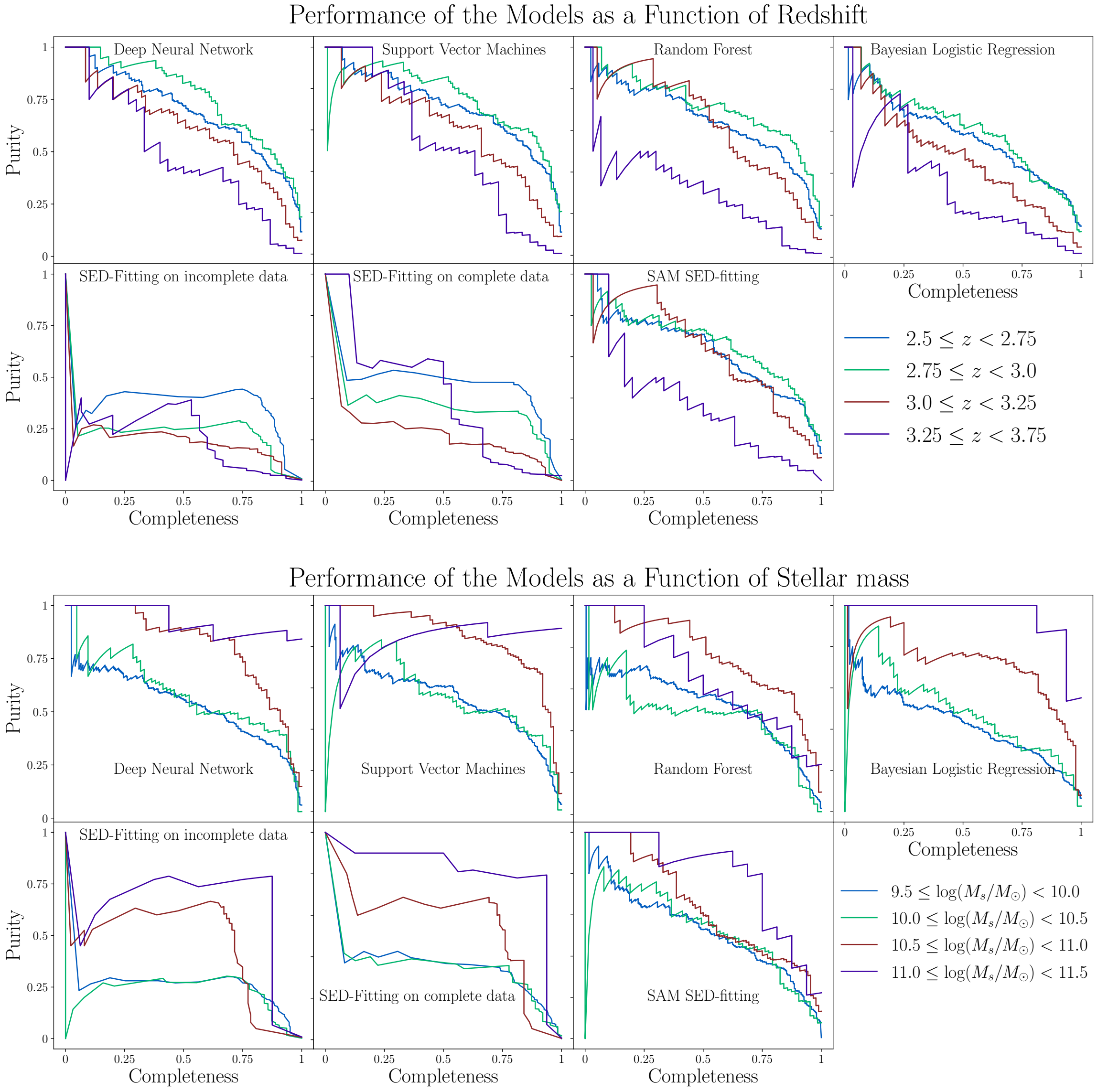


Figure 3.8: Top panel: The performance of the models seen as the purity-completeness curves as a function of redshift. Bottom panel: The performance of the models seen as the purity-completeness curves as a function of stellar mass.

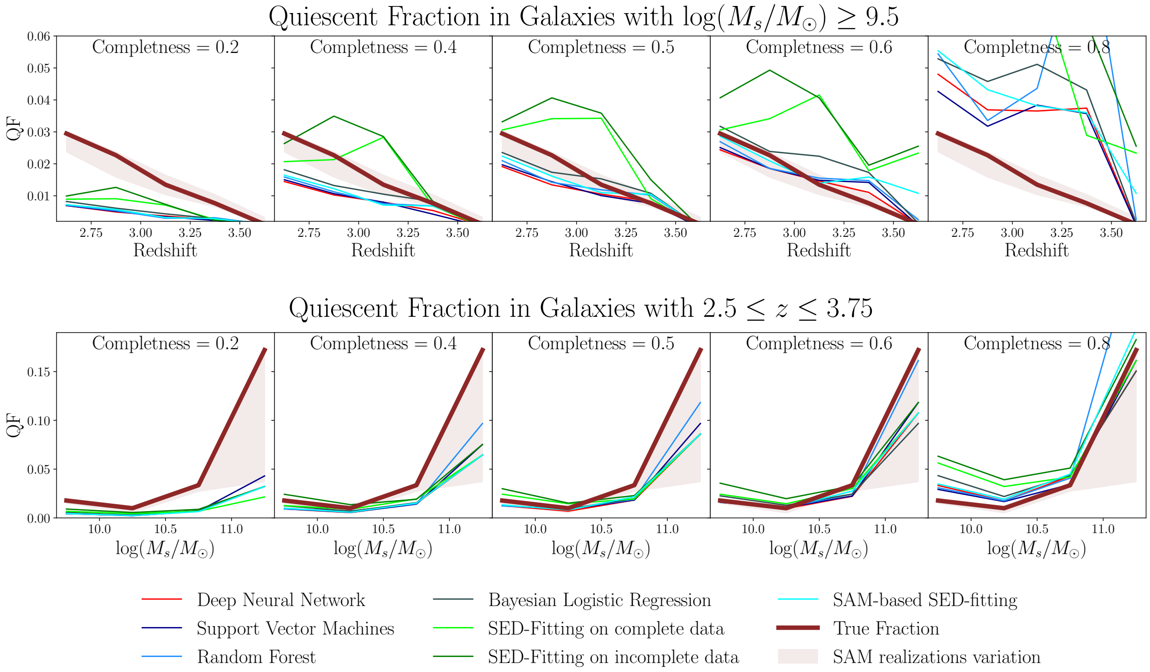


Figure 3.9: Top panel: The performance of the models in recovering the true quiescent fraction as function of redshift in bins of 0.25 at a given completeness. Bottom panel: The performance of the models in recovering the true quiescent fraction as function of stellar mass in bins of 0.5 at a given completeness level. The true fraction in the testing set is shown as the dark red line and the total variations within the SAM realizations shown with light red region. The quiescent fraction is defined for sample of the galaxies with $\log(M_s/M_\odot) \geq 9.5$ and $2.5 \leq z \leq 3.75$.

3.5 Discussion

In this study, we use statistical learning techniques and build statistical models to classify galaxies using simulations from the Santa Cruz semi-analytical modeling framework. We then use these models to identify high redshift massive quiescent galaxies. First, we ensure that the semi-analytical model's lightcone is similar to the CANDELS GOODS-S by assigning limiting magnitudes to each band to match the observations. We artificially create missing values for each flux at random to replicate the fraction of the data without observation in that particular waveband in CANDELS GOODS-S. To replicate the uncertainties in the flux measurements, we add random noise at levels that were interpolated from the CANDELS GOODS-S observations.

We introduced and used a method named MICE to estimate the missing magnitudes using an iterative approach by finding an estimate for missing magnitudes conditional on the known magnitudes at every iteration. We show that this method can recover the missing values with reasonable accuracy compared to the ground truth. Then we trained the models using the training set and applied 5-fold cross-validation to set each model's hyperparameters where needed.

We compared the models' performances based on the ROC and purity-completeness curves. We find that the Bayesian Logistic Regression, Random Forest, Support Vector Machines, and neural network classifier do a much better job than the usual SED-fitting technique in identifying the high-z massive quiescent population, even when we fix the redshifts and do not include any missing values (the noise where included). Bayesian Logistic Regression and Random Forest have similar performance to the SAM-based SED-fitting procedure.

This result is similar to that from Steinhardt et al. (2020), that the t-SNE algorithm applied on the UltraVISTA catalog (McCracken et al. 2012) was able to classify quiescent galaxies better than UVJ and sSFR selection. Here, we find Gaussian naive Bayes, LDA, QDA, and the KNN classifiers trained on three different lower-dimensional data representations to be sub-optimal compared to the SED-fitting without missing values. However, we find the KNN classifiers on t-SNE and UMAP have a similar performance as the SED-fitting with missing values, and we find that QDA outperforms the SED-fitting in this case. Given the posterior probabilities of all these models, one can decide on what threshold to use to minimize the posterior expected loss after properly calibrating the posterior probabilities and using a loss function appropriate for the particular application.

There are many possible reasons behind the lack of high purity in the SED-fitting model. One of the most probable reasons is that we use simplified models for star-formation histories when fitting the galaxies compared to the much more complex star-formation histories present in the SAM catalog. This problem with the SED-fitting classifier is exacerbated by the fact that the population of the galaxies we are interested in are either became quiescent recently or are in the final process of this transition, which means that the SED is more susceptible to recent star-formation episodes (type A stars still significantly contribute to the total SED). The importance of using a more complex set of libraries can be seen in simply treating the galaxies in the training set as a library, and performing the “fitting” will significantly improve the classifier’s performance. In work presented here, we did not use Bayesian SED-fitting codes (e.g., Carnall et al. 2018; Johnson et al. 2021) that can be more robust given their posterior over physical parameters, which can directly turn into a classifier and improve the results presented here. Some codes also consider more realistic star-formation histories using SAMs or hydrodynamical simulations or having a non-parametric approach to model the star formation history and have star-formation at different stages as free parameters inferred from the data, enhancing the results from the SED-fitting method (e.g., Iyer and Gawiser 2017; Iyer et al. 2019; Leja et al. 2019). However, they become increasingly computationally demanding as the number of sources and the dimensions of the physical parameter grows and including more physical parameters can lead to more significant uncertainties in the inferences of these physical parameters and down the line classification, especially if only limited photometric bands are available.

Results show that learning a classifier directly from a population either by modeling the Bayes classifier or approximating it using empirical risk minimization can provide much more robust samples, mainly when dealing with a rare population of galaxies, by circumventing the estimation of the physical properties. There are some advantages to these models compared to the case-by-case classification used based on SED fitting results. One is that some of these models incorporate an implicit prior on the population level for a given galaxy classification, which means that we can provide the model a baseline for how much we should expect for the classification of a given galaxy during the learning process. Another advantage comes from the fact that the models trained on the entire training data can share information about the relevance of a given band at the population level so that a band with little information about the labels (at the population level) will have a more negligible contribution to the classifier compared to the SED-fitting which can only access the individual flux measurements and their uncertainties. Also, the modeling and optimization done with these models are precisely set to make the best classification decision based on the labels we provide for the whole population rather than inferring the physical properties and classifying a galaxy. In other words, the problem of classifying galaxies presented here is more straightforward than inferring the joint probability distribution of the relevant physical properties that fall into the selection criteria after marginalizing over the rest of the nuisance parameters correctly.

However, we note that some fraction of this improvement comes from the fact that the testing set and training set for the SAM-based method and other learning-based models are coming from having the same population distribution in the training and testing set, which is not necessarily expected when applied on the real galaxy population.

3.5.1 Prediction on CANDELS GOODS-S observation

As a general test, we apply the trained model to the CANDELS GOODS-S magnitudes catalogs. After applying the procedure detailed in §3.2.4, we estimate all the missing values in the CANDELS observation. The models were then used to classify galaxies in the CANDELS GOODS-S field to quiescent and star-forming populations and assign a probability to each galaxy from a given model. In the models trained on the lower-dimensional data embeddings, the embedding trained on the SAM catalog is used to transform the CANDELS data and apply the K-Nearest Neighbors to classify the galaxies. Then we compared the results with the previous studies done on the CANDELS fields. Merlin et al. (2018), and Merlin et al. (2019) use an approach based on the SED-fitting of the galaxies using constant star formation histories and putting some constraint on the probability of the star-forming versus quiescent solutions. Shahidi et al. (2020) uses a combination of UVJ, observed colors, and SED-fitting selection, in which different star-formation histories are assumed and combined using Bayesian model averaging.

We find a general agreement with candidates selected in those studies, and most of the models used here assigned a non-zero probability to many of the galaxies chosen in those studies (with many having ALMA confirmations; see Santini et al. 2019; Santini et al. 2020); however, there are some cases in which the models assign a high probability, where they are not selected in their expected categories and vice versa. There is also a general consistency between the models' predictions in the sense that if one predicts a higher probability, the other will do so similarly.

Table 3.4 shows the probability assigned by the models with best performances to the set of the candidates that are confirmed based on the upper bound in their ALMA fluxes from Santini et al. (2020). We find that 7/17 of the objects studied in Santini et al. (2020) have very high probability across different models, 6/17 objects were not selected in Shahidi et al. (2020), and 4/17 were selected in both of those studies and are assigned a low probability based on the models here. Nevertheless, it is difficult to draw a general conclusion on the true error rate of the classifiers discussed here, and a larger and more complete sample of spectroscopic confirmation of these populations is required to better calibrate the models used here.

3.5.2 Importance of each photometric band

In addition to the classification task, we use the Random Forest to measure the relative importance of different bands in classifying the galaxies.

Table 3.4: Prediction of the Bayesian Logistic Regression (BLR), Random Forest (RF), Neural Network (NN), and Support Vector Machine (SVM). The last columns indicate if a given candidate was selected in Shahidi et al. (2020).

Candidate ID	BLR	RF	NN	SVM	SH20
GOODSS-2782	0.34	0.42	0.96	0.80	Y
GOODSS-3897	0.00	0.00	0.00	0.00	N
GOODSS-3973	0.99	0.24	0.99	0.73	N
GOODSS-4503	0.33	0.72	0.99	0.98	Y
GOODSS-7526	0.01	0.03	0.05	0.02	Y
GOODSS-7688	0.17	0.59	0.02	0.45	N
GOODSS-9209	0.00	0.08	0.00	0.02	Y
GOODSS-10578	0.80	0.59	0.99	0.90	N
GOODSS-12178	0.00	0.07	0.02	0.01	Y
GOODSS-13394	0.00	0.00	0.00	0.00	N
GOODSS-16526	0.00	0.00	0.00	0.00	N
GOODSS-17749	0.69	0.29	0.99	0.93	Y
GOODSS-18180	0.89	0.28	0.99	0.99	Y
GOODSS-19301	0.04	0.08	0.01	0.03	N
GOODSS-19446	0.01	0.00	0.00	0.00	N
GOODSS-19505	0.00	0.00	0.00	0.00	N
GOODSS-19883	0.00	0.03	0.00	0.00	Y

In the first approach, we use the entropy decrease at every decision node in the tree calculated during the training. For finding the importance of some band x_j, the weighted average of the reduction of entropy summed over all the nodes related to the band x_j is calculated. The weights are simply the number of training examples that reach the particular node. We do this for all the trees in Random Forest $\{\mathcal{T}_i\}_{i=1}^{m}$, and find the average and standard deviations for the importance of the band x_j.

In the second approach, we first predict the labels and find the testing dataset's misclassification rates. We then go through each magnitude x_j one by one and randomly permute the values of magnitude x_j, which effectively removes its relevance in the classification task. Then the increase in the mean squared error of the predicted probabilities compared with the actual labels of the galaxies is calculated[20] (binary labels of 0 and 1) for all the bands. We go through these steps fifty times and find the average and standard deviation of the increase in this measure. This procedure underlines the importance of each magnitude in reducing the classifier's predicted probabilities error and can be applied to any classifier.

Figure 3.10 shows the relative importance of each variable (magnitude) in learning the model that predicts the training set classification. We find that the bands: IRAC2, IRAC1, Y, K, J, H are the most important. As the figure shows, the second approach tends to produce more uniform importance measures[21], however, the ranking of the importance of the features does not change significantly. The result is somewhat expected since these are the bands mainly used for the photometric selection of the high-z quiescent population (e.g.,

---

[20] This is called the Brier score loss

[21] This is a generic behavior of the permutation method, compared to the decrease in entropy (See Hastie et al. 2009 section 15.3).

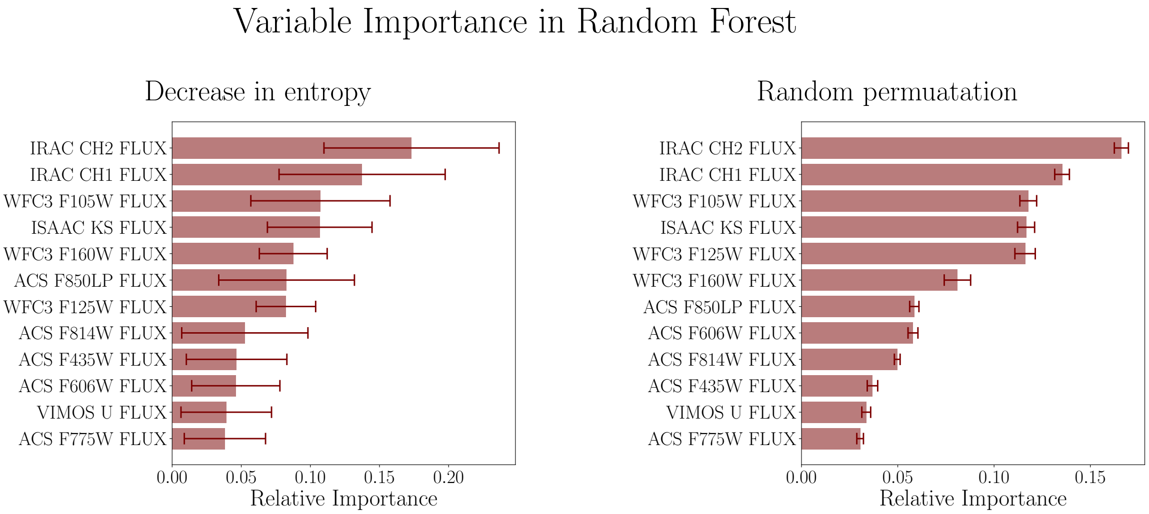


Figure 3.10: Left: Shows the ensemble average of each band's importance defined based on the decrease in entropy. The error bar indicates the standard deviation within the ensemble of decision trees. Right: Shows the average increase in the error of probability predictions after 50 permutations for each band. The error bar indicates the standard deviation of this measure over permutations. As expected, the bands to which the Balmer break is shifted to (at the observed redshift of the galaxy) are more critical. The feature/variable importance measures for each band shown here are relative and normalized to sum up to one.

Nayyeri et al. 2014; Shahidi et al. 2020) and the importance of the $\mathrm{IRAC2}, \mathrm{IRAC1}, K$ has been discussed before with details in Merlin et al. (2019) and Santini et al. (2020).

3.5.3 Caveats and possible improvements

One of the central assumptions needed for reproducing the techniques' enhanced performance is to have a representative training set. Although the application of the models trained on the SAM and applied on the CANDELS GOODS-S observation can give us confidence that using a training set made from the simulations is successful to a degree, a better approach would be to have a confirmed representative sample for a subset of observations and using those in training. Then one can use these models to apply on the more extensive survey where careful selection and confirmation of these objects using comprehensive checks would not be feasible. An easier way to increase the confidence in the trained models would be to use and check catalogs from different simulations to validate and test the trained models.

Also, it should be noted that we have not employed all the simulated lightcones for the training, since, we wanted to use more of this rare population of galaxies (by oversampling from other lightcones) and their number density is not very well calibrated because of the lack of complete spectroscopic samples, due to their faint nature and lack of prominent emission lines (see i.e., Newman et al. 2018b; Valentino et al. 2020; Forrest et al. 2020b; Forrest et al. 2020c). Therefore, including all these lightcones and oversampling from different error realizations rather than different lightcones could increase the classifier's performance, at least on the testing set.

The t-SNE and UMAP algorithms are originally unsupervised learning algorithms that can be used as a general preprocessing step for other estimators because of their ability to learn the lower-dimensional representation of the manifold the data is sampled from. This makes them good knowledge discovery tools. However, they can turn into supervised algorithms by adding the relevant terms based on their labels to the cost functions. One can further improve the resulting embedding in favor of particular labeling. Besides, the S-PCA algorithm can be used to find the embedding that maximizes the dependence of magnitudes and physical properties, which can be the high dimensional set of physical properties. By defining a meaningful kernel (similarity measure), one can use this transformation as another alternative to the SED fitting techniques for finding any set of physical properties. One should note that this algorithm strongly depends on the notion of dissimilarity/distance, and changing those can affect the final embedding, resulting in the classifiers with different performances.

We have also used different neighbors to construct the neighbor graph in UMAP and different perplexities for t-SNE (from 5 to 50). We do not find any substantial structural changes in the resulting maps in their topology, probably because of the simplicity of the underlying manifold that allows the global structure to be captured with even a few neighbors.

3.6 Conclusion

This work investigates statistical learning techniques to classify galaxies based on their direct observables (i.e., magnitudes) using a simulated catalog from a semi-analytical model after including observational effects such as measurement uncertainties and the presence of censored and missing values. We defined the true classification based on galaxies' actual physical properties and turned our focus on the rare population of high-z massive quiescent galaxies as an example.

Two SED-fitting procedures are performed utilizing LePhare code with 50 realizations of the Gaussian noise for galaxies' photometry, assuming known redshifts, and one with and one without missing values in the catalog. Then classification is made on the space of the inferred physical properties similar to the original definition of the class.

We demonstrated that modeling conditional distributions for missing bands using an iterative sampling approach can be useful for estimating the missing values. Furthermore, after estimating these missing values, the completed catalog was used for learning different classifications.

The K-nearest neighbors (KNN) classification on the lower-dimensional representation of the data based on t-SNE and UMAP transformations perform similar to the SED-fitting with missing values, and classification based on the linear discriminant analysis, Gaussian naïve Bayes, and KNN on the supervised-PCA representation showed to perform worse than SED-fitting according to the area under the purity-completeness curve. According to the area, under the purity-completeness and ROC curves, *Bayesian Logistic Regression*, *Random Forest*, *Support Vector Machines*, and *neural network* classifications outperform both SED-fitting classification procedures. The *Support Vector Machines* and *neural network* classifiers have better purity at almost every completeness rate.

In short, we have demonstrated the application of several statistical learning algorithms trained on the SAMs lightcones as a powerful tool, which can provide many applications involving classification or inferring physical properties, particularly for upcoming extensive surveys where SED-fitting methods can be very time-consuming. Using the methods discussed in this study, one can select a sub-sample of the candidates for spectroscopic follow-ups or perform more suitable SED-fitting using more informative prior information based on their classifications.

Chapter 4

Probabilistic Modeling of Star-Forming Sequence: Quiescent Fraction Dependence on Mass and Environment

4.1 Introduction

Observations of the population of galaxies at high redshifts, combined with predictions from the standard ΛCDM models have helped us understand the complex process of galaxy evolution (Thoul and Weinberg 1995; Benson et al. 2003). Galaxies build up their stellar masses over time with a rate that largely depends on the amount of available cold gas, which is determined by the cosmological accretion rate, the efficiency in cooling of the gas, the baryonic effects involved with delivering these gas reservoirs from halo size scale to much smaller scales of the galaxies, in addition to feedback processes.

Feedback processes make galaxies only relatively efficient at forming stars within relatively narrow range of the halo mass (Moster et al. 2010; Behroozi et al. 2013b). At the lowest mass regime of the mass function, the photoionizing UV radiation during and after the reionization suppresses the galaxy formation (e.g., Rees 1986; Quinn et al. 1996; Efstathiou 1992; Barkana and Loeb 1999; Bullock et al. 2000; Shaviv and Dekel 2003; Gnedin and Kravtsov 2006); at more intermediate stellar mass, the stellar feedback processes decrease the efficiency of galaxy formation (e.g., Larson 1974; Dekel and Silk 1986; White and Frenk 1991; Cole 1991; Dekel and Woo 2003; Hopkins et al. 2014). Finally, at higher masses, the effects of the virial shock heating of the accreted gas onto the massive halo (Rees and Ostriker 1977; Silk 1977; Binney 1977; White and Rees 1978; Birnboim and Dekel 2003; Kereš et al. 2005; Cattaneo et al. 2006; Dekel and Birnboim 2006) and the Active Galactic Nuclei (AGN) (Silk and Rees 1998; Haehnelt et al. 1998; Granato et al. 2004; Croton et al. 2006) becomes important that can suppress the star-formation rate by halting the accretion of the gas to the galaxy and/or remove the gas from the galaxy or halo (Fabian 2012; Harrison et al. 2018).

Despite the non-linearity of these processes and complex interaction between them, the observed star-forming galaxy population appears to be in a quasi-equilibrium state on the star formation rate-stellar mass plane. This implies an emergent feedback process that scales the star formation rate to its stellar and halo mass to keep galaxies along the sequence (Bouché et al. 2010; Davé et al. 2011; Davé et al. 2012; Lilly et al. 2013; Dekel and Mandelker

2014). However, not all galaxies live on the star-forming sequence, and other than a few bursts of star formation and up and down of the sequence, they evolve to build up the quiescent population with little or no ongoing star formation rate.

Additionally, the local environment has been shown to affect the galaxy properties in addition to the in situ evolution (Dressler 1980; Balogh et al. 2004; Kauffmann et al. 2004; Blanton et al. 2005). There have been many studies showing the amount of star-formation rate in different environments is not showing any trend (e.g., Peng et al. 2010; Wijesinghe et al. 2012; Muzzin et al. 2012; Darvish et al. 2014; Darvish et al. 2016; Leslie et al. 2020) or a modest trend was reported (e.g., Patel et al. 2011; Haines et al. 2013; Old et al. 2020). However, the probability of finding a quiescent galaxy increases in the densely populated regions mostly attributed to the external quenching mechanism for the satellite galaxies (e.g., Balogh et al. 2004; Baldry et al. 2006; Peng et al. 2010; Darvish et al. 2014; Darvish et al. 2016).

In this study, an empirical model for the "quiescent/transitioning" fraction as a function environment and stellar mass is used that consistently model the star-formation rate as a function of stellar mass and galaxy type based in the presence of the noisy and censored measurements.

The multi-waveband data and the observed catalog is discussed in §4.2. In §4.3 the probabilistic model is defined, and different constituents of this model is described. In §4.4, the result of the inference on the model parameters is represented. In §4.5 an overview of the results are discussed in addition to possible drawbacks, and ways to improve.

4.2 Data

We use the latest photometric catalogs from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) (Koekemoer et al. (2011); Grogin et al. (2011)) with consistent multi-waveband photometry, and physical parameters for all galaxies to the flux limit of the sample [1]. Details about the selection and photometry at different bands for all the CANDELS fields are given in the following: GOODS-South (Guo et al. (2013); Santini et al. (2015)), UDS (Galametz et al. (2013);Santini et al. (2015)) , COSMOS (Nayyeri et al. (2017)) , EGS (Stefanon et al. (2017)), and GOODS-North (Barro et al. (2019)). These catalogs contain, for each galaxy, the observed photometry from the UV to near and mid-infrared wavelengths in many broadband and narrowband filters as well as inferred physical parameters.

Galaxies are selected in the CANDELS catalog with 'FLAGS = 0', which are non-contaminated sources in the F160W band and 'CLASS_STAR $\leq$ 0.9', remove potential stellar objects using the SExtractor Star/Galaxy classifier output.

We use an approach introduced by Pozzetti et al. (2010) to find the mass complete sample of galaxies at different redshift bins. This method find the limiting mass (M_{lim}) defined as the mass if an individual galaxy, if it were at the limiting magnitude of the survey with the same mass-to-light ratio (M/L).

Therefore, for any galaxy, i, we keep the M_i/L_i ratio constant and use it to find the mass-to-flux ratio and hence, the magnitude of the galaxy. Using this procedure, one can find the limiting mass for all galaxies in the survey, using the relation: $\log(M_{lim,i}) =$

---

[1] http://arcoiris.ucolick.org/candels/

$\log(\hat{M}_i) + 0.4 \times (m_i - m_{lim})$, where m_i, and m_{lim} are the magnitude of the galaxy and limiting magnitude of the survey respectively.

4.2.1 Star-formation Rates Estimates

The star formation rate (SFR) estimations are from Barro et al. (2019), which uses a consistent empirical approach using emissions in ultraviolet, mid-infrared, and far-infrared. The IR fluxes in *Spitzer* MIPS, *Herschel* PACS, and SPIRE sources were consistently matched to the WFC3-F160 sources in the CANDELS.

The UV SFR is based on the monochromatic luminosities and the total emission from UV after bolometric correction and estimated using Kennicutt (1998) relation. Two different methods were used to measure infrared-based SFRs: (1). for sources with detection in MIPS and at least one Herschel band, thermal IR emission were fitted to dust emission models from libraries and after integrating the IR luminosity in best fitted model the bolometric IR calibration of the Kennicutt (1998) was used. (2). For galaxies with only a MIPS detection, the total IR emission were estimated using the Wuyts et al. (2008) and Wuyts et al. (2011) and from which the SFRs were calculated. The total SFR for each galaxy is then estimated as the sum of the UV-based and IR-based SFRs. If no IR data were available, the UV-based SFR was used after applying correction for dust extinction using were used after correcting for the dust using the UV spectral index (i.e., β_{UV}). We assume that the SFR measurements are censored at the lowest possible values one can estimate (i.e., have upper limits) from the limiting magnitude of the observations based on the bands that covers the rest-frame UV for each redshift bin.

4.2.2 Redshift Estimates and Stellar Mass Estimates

The redshifts are taken from the photometric redshift catalogs for all the CANDELS fields from Kodra 2019. The redshifts for galaxies in each field, were estimated from six independent measurements using template-based methods. The probability density functions (PDFs) for individual galaxies from each method were estimated and were corrected for bias and their variance. The corrected PDFs were then combined using the minimum Frechet distance. The final catalog consist of the spectroscopic/3D-*HST* grism redshifts and the photometric redshifts (Kodra et al. in prep.). The redshifts used in this work show a normalized median absolute deviation of $\sigma_{\rm NMAD} \sim 0.02$.

The stellar masses were measured through SED fitting by fixing the redshifts at their best values. A library of model galaxies SED is made from Bruzual and Charlot (2003a)) assuming an exponentially declining star formation history with an e-folding parameter in the range of $0.02 \leq \tau \leq 30$ (Gyr) (20 uniformly sampled from the logarithmic scale). Models with three metalicities $(0.02, 0.008, 0.004)$ and dust attenuation law from Calzetti et al. (2000) with 20 $E(B-V)$ sampled uniformly between 0 to 1.2 are used. We use the LePhare SED-fitting code (Arnouts et al. 1999; Ilbert et al. 2006) which include the nebular emission as described in Ilbert et al. (2009).

4.2.3 Local Density Estimates

To estimate the local density contrast where each of the galaxies reside, we use the catalog from Chartab et al. 2020 for the CANDELS field. These catalogs construct a two-dimensional structure within a narrow redshift slice in the redshift catalog and the density field is estimated using a non-parametric weighted kernel density estimation. The density field is defined as a weighted mixture of von Mises kernel that is fixed at every galaxy's position. The weights were assigned to each galaxy as the integral of their redshift PDF within the redshift bin. For finding the overall bandwidth of the kernel at a given redshift bin, a likelihood cross-validation method is used. These bandwidth are then adapted locally to reduce the bias of the densely populated regions by reducing the bandwidths and over-smoothing using the local density. Also, the variance in the under-dense regions is reduced by increasing the bandwidth to avoid under smoothing in these regions. The measurements of the density at the boundary of the survey is underestimated and for correcting this a boundary correction is calculated and applied to the density measurements for all the galaxies. Throughout this work we use the density contrast measurements defined as the ratio of the local number density to background number density at a given redshift bin minus one.

4.3 Probabilistic Model

To describe the multi-dimensional joint distribution, we find the joint probability over all the galaxies within the population observed in our sample; assuming that the full distribution over the population can be separated[2] into the individual probability distributions as the following:

---

[2] identical and independently distributed (i.i.d)

$$p(\{M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \Delta_i, \eta_i\}_{i=1obs}^{N} | \boldsymbol{\theta}) = \prod_{i=1}^{N_{obs}} p(M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \Delta_i, \eta_i | \boldsymbol{\theta})$$

where $\{\hat{M}_i, \hat{\psi}_i, \Delta_i\}_{i=1}^{N_{obs}}$ is the set of *observed physical properties* of N_{obs} galaxies in the sample with $\hat{M}_i$, $\hat{\psi}_i, \Delta_i$ being respectively the stellar mass, star-formation rate, and local density contrast of galaxy i respectively. η_i is a categorical variable that indicates the galaxy type of the i-th galaxy (i.e., star-forming, star-burst, quiescent). $\boldsymbol{\theta}$ refers to all the parameters used in the model and $\{M_i, \psi_i\}$ are the *latent variables of the model for true physical properties* of the galaxies that are not directly observed and we only have access to them through a noisy measurements and need to be evaluated within the probabilistic model.

In order to find the likelihood of the parameters based on some conditional independence of physical properties to specify our model as the following:

$$p(M, \psi, \hat{M}, \hat{\psi}, \Delta, \eta | \boldsymbol{\theta}) = p(\hat{M}, \hat{\psi} | M, \psi) p(\psi | M, \eta, \boldsymbol{\xi}) p(\eta | M, \Delta, \boldsymbol{\zeta}) p(M | \boldsymbol{\alpha}) p(\Delta | \boldsymbol{\mu})$$

$p(\hat{M}, \hat{\psi} | M, \psi)$ describes the measurement process (i.e., the relation between the measured and true physical properties) and can modeled using the full covariance of the measurement uncertainties, $p(\psi | M, \eta, \boldsymbol{\xi})$ is the conditional probability distribution of the galaxies true star-formation rate given their true stellar masses and galaxy type which can be modeled by some set of parameter $\boldsymbol{\zeta}$, $p(\eta | M, \Delta, \boldsymbol{\zeta})$ is the conditional probability of a galaxy type given their stellar mass and local density which can be modeled using parameters $\boldsymbol{\zeta}$, $p(M | \boldsymbol{\alpha})$ is the probability distribution over the true stellar masses and can be modeled using a Schechter distribution (Schechter 1976) with some parameters called $\boldsymbol{\alpha}$, and finally $p(\Delta|\boldsymbol{\mu})$ is the distribution of the local density parameterized by $\boldsymbol{\mu}$.

For modeling the joint distribution of the $p(M, \Delta)$, we have chosen to assume an independent assumption which is not necessarily true (see van der Burg et al. 2020; Weigel et al. 2016) and this choice was to make the full model well specified. Consequently one can drop from the likelihood and condition on their observed values since we do not have a measure of uncertainties over these. The parameters of the model are encoded as $\boldsymbol{\theta} = \{\boldsymbol{\zeta}, \boldsymbol{\xi}, \boldsymbol{\alpha}\}$ after dropping the probability distribution over the local density contrast.

We define the likelihood function of the parameters for the observed galaxies after marginalizing the categorical variable of galaxy type:

$$\mathcal{L}_{obs} = p(\{M_i, \psi_i, \hat{M}_i, \hat{\psi}_i\}_{i=1}^{N_{obs}} | \{\Delta_i, \}_{i=1}^{N_{obs}} \boldsymbol{\theta}) = \prod_{i=1}^{N_{obs}} \sum_{\eta} p(M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \eta_i = \eta | \Delta_i, \boldsymbol{\theta})$$

Now we can describe the joint probability distribution as their individual constituents.

4.3.1 Galaxy Types and the Quiescent Fraction

In order to model galaxies on the star-formation rate and stellar mass plane, we assume that the galaxies can be modeled separately given their type. The formulation of the above model can be used to model galaxies within distinct groups of star-burst, star-forming, and “quiescent”. Here we refer “quiescent” to all galaxies that are not part of the star-forming sequence and have lower star-formation given the stellar mass. These class of galaxies will include all the objects that are in middle of transition to fully quiescent population, however, there are indication on some of these galaxies are in the bursting phase and not necessarily becoming quench (see Martin et al. 2017; Darvish et al. 2018). This problem can be alleviated by using an asymptotic relation in SFR vs. Mass. The simultaneous modeling of different galaxy types is known as mixture model where we describe each galaxy type as a component of the larger model. This modeling approach to the SFR vs Mass plane is closely related to the method introduced in Renzini and Peng (2015), with the difference of using explicit probabilistic modeling. Here we need to model two parts, one for the description of each component and one for the probability of belonging to each galaxy type. The probability of belonging to each component in the mixture of three component model is equivalent to modeling the Multinomial distribution[3] parameter (i.e., the fraction of star-burst, star-forming, and quiescent). We assume that the parameter for star-forming and quiescent depends on the stellar mass and local density of each galaxy, while the star-burst are modeled simply as outliers because of their rarity which makes a more detailed model to become unidentifiable without strong prior.

Using a generalized linear model with Gibbs distribution, one can parameterize the Multinomial parameter π_Q in terms of the log-odds as the following:

---

[3] Generalization of the Bernoulli distribution for more than two outcomes

$$\pi_{SB} + \pi_Q + \pi_{SF} = 1$$

$$\pi_{SB} = p(\eta = \text{Star-Burst}|M, \Delta, \boldsymbol{\xi}) = p(\eta = \text{Star-Burst}),$$

$$\pi_Q = \pi_Q(M, \Delta) = p(\eta = \text{Quiescent}|M, \Delta, \boldsymbol{\xi})$$

$$\ln(\frac{\pi_Q}{\pi_{SF}}) = \xi_0 + \xi_1 \log(\frac{M}{10^{10} M_\odot}) + \xi_2 \log(1 + \Delta) + \xi_3 \log(\frac{M}{10^{10} M_\odot}) \times \log(1 + \Delta)$$

This is a simple model describing the log odds of quiescent to star-forming given the galaxy stellar mass and local density in terms of their linear combination as well their first order interaction. This parameterization, similar to what was proposed in Baldry et al. 2006, and Peng et al. 2010, is capable of producing the behaviour of quiescent fraction seen in the plane of stellar mass and local denisty for SDSS sample. However, since the quenching efficiency of SDSS sample showed little to slight dependence of the mass quenching efficiency on environment and environment quenching efficiency on the stellar mass, they proposed a complementary log-log link function to model the quenching efficiency. Since there has been couple of studies showing that these quenching efficiency might have noticeable dependence on each other at high redshift (e.g., Lin et al. 2014; Balogh et al. 2016; Kawinwanichakij et al. 2017; Chartab et al. 2020), we use a logit link function that can produce such dependency on top of the behavior seen on the stellar mass and local denisty plane.

4.3.2 Star-formation rate versus Stellar Mass

The observed relation between the SFR and stellar mass has been extensively studied particulary when describing the star-forming sequence, where a linear relation with an intrinsic scatter has been used to model them (e.g., Noeske et al. 2007; Salim et al. 2007; Elbaz et al. 2007; Daddi et al. 2007; Santini et al. 2009; Whitaker et al. 2012b; Speagle et al. 2014; Schreiber et al. 2017; Pearson et al. 2018). There have been many studies suggesting that the slope of this relation depends on the stellar mass particularly at higher mass end and lower redshifts, which led to use of quadratic form, piecewise linear, and asymptotic relations (e.g., Whitaker et al. 2014a; Schreiber et al. 2015; Lee et al. 2015; Leslie et al. 2020). Also, there is evidence that the galaxies below the star-forming sequence linear relation at higher masses to be in bursting phase defined as the time derivative of the NUV − i color (Martin et al. 2017; Darvish et al. 2018). However, after trying to model our galaxies using these models, we find that at higher redshift the solutions prefer the linear part of the model and the asymptotic part falls outside of the data, which might behave differently if additional data is available. In order to have a homogeneous modeling for all of our redshift bins we limit this study to a simple linear relation, while noting its possible drawbacks.

$$
\begin{gathered}
\mu_{SF} = \zeta_{0,SF} + \zeta_{1,SF}\log(\frac{M}{10^{10}M_\odot}) \\
\mu_{Q} = \zeta_{0,Q} + \zeta_{1,Q}\log(\frac{M}{10^{10}M_\odot}) \\
\mu_{SB} = \zeta_{0,SB} + \zeta_{1,SB}\log(\frac{M}{10^{10}M_\odot}) \\
p\left(\log(\frac{\psi}{M_\odot/yr})\middle|M, \eta = \text{Star-forming}\right) = \mathcal{N}\left(\log(\frac{\psi}{M_\odot/yr})\middle|\mu_{SF}, \sigma^2_{int,SF}\right) \\
p\left(\log(\frac{\psi}{M_\odot/yr})\middle|M, \eta = \text{Quiescent}\right) = \mathcal{N}\left(\log(\frac{\psi}{M_\odot/yr})\middle|\mu_{Q}, \sigma^2_{int,Q}\right) \\
p\left(\log(\frac{\psi}{M_\odot/yr})\middle|M, \eta = \text{Star-burst}\right) = \mathcal{N}\left(\log(\frac{\psi}{M_\odot/yr})\middle|\mu_{SB}, \sigma^2_{int,SB}\right)
\end{gathered}
$$

where ζ_0 is the intercept in SFR at $10^{10} M_\odot$, ζ_1 is the slope of the mean relation, and $\sigma_{int,SF}$, and $\sigma_{int,Q}$ are the intrinsic scatter of the $\log(\frac{\psi}{M_\odot/yr})$ from it's mean relation for star-forming and quiescent population respectively. It should be noted that whenever we change the variable in the model, a proper Jacobian factor is introduced into the larger model to transform the volume element in which the probability density is defined.

4.3.3 Probability distribution over the stellar mass and over-densities

Another way to describe the overall properties of the galaxy population is through the study of number density of the galaxies given their stellar masses i.e., the stellar mass function. There any many different approaches used to measuring this distribution (see Weigel et al. 2016 for a recent overview of the methods). The $1/V_{max}$ method (Schmidt 1968), a parametric maximum likelihood method (Sandage et al. 1979 and a non-parametric maximum likelihood method (Efstathiou et al. 1988)) are some of the main approaches. Different approaches employ different methods for correcting the Malmquist bias (Malmquist 1922, Malmquist 1925) and the Eddington-Jefferys bias (Eddington 1913; Jeffreys 1938). In this study the measurement uncertainties is consistently modeling the Eddington-Jeffery and given the mass completeness condition the Malmquist bias should be minimal.

To model the stellar mass probability distribution, we adopt the Schechter parameterization (Schechter 1976) after turning it into a proper probability distribution using incomplete gamma function:

$$\Phi(M) = \frac{N}{V} p(M|\boldsymbol{\alpha})$$

where the $V = \int_{z_j}^{z_{j+1}} \frac{dV}{dz} dz$ is the volume of the survey within the adopted redshift bin (i.e., $[z_j, z_{j+1}]$) in which we assume the stellar mass function parameters do not evolve, and N is the number of galaxies in the population above some mass ${M_{min}}$[4], where we truncate the probability distribution to make it proper for values of $\alpha \leq -1$. The probability distribution of the galaxies stellar mass follows a Schechter distribution is defined as the following:

$$p(M|\alpha, M_*) = \frac{\left(\frac{M}{M_*}\right)^{\alpha} e^{-\frac{M}{M_*}}}{M_* \Gamma(\alpha + 1, \frac{M_{min}}{M_*})}$$

where the parameters $\{\alpha, M_*\}$ which are the lower mass slope (shape parameter) and exponential cut-off scale parameter respectively. $\Gamma(\alpha + 1, \frac{M_{min}}{M_*})$ is the upper incomplete gamma function defined as the following:

$$\Gamma(s, x) = \int_x^{\infty} t^{s-1} e^{-t} dt$$

The conditional distribution of galaxy's stellar mass is calculated given the galaxies overdensity and type as the following:

$$p(M|\Delta, \eta, \boldsymbol{\alpha}, \boldsymbol{\zeta}) = \frac{p(\eta|M, \Delta, \boldsymbol{\zeta}) p(M|\boldsymbol{\alpha})}{\int p(\eta|M, \Delta, \boldsymbol{\zeta}) p(M|\boldsymbol{\alpha}) dM}$$

---

[4] The lowest stellar mass in the observed catalog at a given bin.

To model the distribution over the over-densities needed for taking averages over density, the following skew normal distribution is adopted:

$$p(\log(1+\Delta)|\boldsymbol{\mu}) = \text{Skew-}\mathcal{N}(\log(1+\Delta)|\mu_\Delta, \sigma_\Delta, \alpha_\Delta)$$

where $\boldsymbol{\mu} = \{\mu_\Delta, \sigma_\Delta\}$ are the location, scale, and shape parameters of the skew normal probability distribution over $\log(1+\Delta)$ defined as the following:

$$\text{Skew-}\mathcal{N}(x|\mu, \sigma, \alpha) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \int_{-\infty}^{\alpha\left(\frac{x-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\ dt$$

One can adopt a more general approach for modeling the distribution over the density such as a mixture of normal distribution, however, we find the skew normal distribution to be adequate for this model.

4.3.4 Measurements Uncertainties

To account for measurement uncertainties, one can use the full joint probability distribution from the posterior distribution of the SED-fitting to encode all these information from the inferred properties. However, in this work we use SFR estimate from the UV+IR calibration and the stellar masses which are determined from the SED-fitting. Therefore, we make an additional assumption that measured quantities for an individual galaxy are independent and follow a Gaussian distribution in log space. The following shows the conditional joint probability distribution of the measured properties given their true unknown values:

$$
\begin{aligned}
p(\hat{M}_i, \hat{\psi}_i | M_i, \psi_i) &= p(\hat{M}_i|M_i)p(\hat{\psi}_i|\psi_i) \\
p\left(\log(\frac{\hat{M}_i}{M_\odot})\middle|\log(\frac{M_i}{M_\odot})\right) &= \mathcal{N}(\log(\frac{M_i}{M_\odot}), \sigma^2_{M,i}) \\
p\left(\log(\frac{\hat{\psi}_i}{M_\odot/yr})\middle|\log(\frac{\psi_i}{M_\odot/yr})\right) &= \mathcal{N}(\log(\frac{\psi_i}{M_\odot/yr}), \sigma^2_{\psi,i})
\end{aligned}
$$

where the $\sigma_{M,i}$ and $\sigma_{\psi,i}$ are the stellar mass and SFR measurement uncertainties of the observed galaxy i in their log-space.

Using the above approximation and the fact that the distribution of the true SFRs given their stellar mass and galaxy type is also modeled as a Gaussian distribution we can marginalize over the true SFRs analytically.

Also, for the data with only upper limit (U), we marginalize over the possible values through the following:

$$
p\left(\log(\frac{\hat{\psi}_i}{M_\odot/yr}) \leq U \middle| M_i, \eta\right) = \int_{-\infty}^{U} \mathcal{N}\left(y|\mu_{i,\eta}, \sigma^2_{int,\eta} + \sigma^2_i\right) dy
$$

and subsequently replace these terms in the likelihood.

4.3.5 Prior specification and inference

To find the posterior distribution over the latent variables and the model parameters, a probabilistic programming language called `Stan` (Carpenter et al. 2017) is used. `Stan` uses a Markov chain method called Hamiltonian Monte Carlo to randomly draw a sequence from the posterior distribution of the latent variables and parameters (Neal et al. 2011). Hamiltonian Monte Carlo is very efficient in exploring the high dimensional geometry of the posterior exploiting the gradient information about the geometry of the posterior distribution using the gradient information of the model (see Betancourt 2017).

To make inferences about the model's parameters one have to assume prior distributions and to make sure that the prior distributions are relevant for the problem at hand, we make prior predictive check to make sure that the possible values of the parameters can cover the span of the observation through generating data and comparing it with what is observed. However, we make sure that the prior distributions are not restrictive/strongly informative. After the inference we compare the standard deviation of the marginal distribution of the posterior to prior and this ratio is at most 0.15, which means that the prior is at most weakly informative.

Table 4.1 describes the model parameters and the prior distribution used in the inference.

Also, since the model used in this work involves a mixture models, the posterior distribution over the parameters becomes multimodal since the parameters of each mixtures are fully exchangeable and the labeling of the different galaxy types becomes arbitrary. Also, since we are modeling the proportion of the quiescent to star-forming galaxies, we need to make sure that we are sampling from part of the posterior geometry that is consistent with this. To make sure that inference does not suffer from this inherent non-identifiablity of the mixture models, we transform the location parameters to be ordered through the `Stan` code, which resolves this issue in practice by limiting the parameter space.

Table 4.1: Prior Probability distributions for the model's parameters.

Quantity	**Prior Distribution**	**Definition**
π_{SB}	Beta$(1, 3)$	The proportion of the Star-burst
ξ_0	Normal(0, 2^2)	Constant term in log-odds of quiescent to star-forming
ξ_1	Normal(0, 2^2)	Slope of Stellar Mass in log-odds of quiescent to star-forming
ξ_2	Normal(0, 2^2)	Slope of Density Contrast in log-odds of quiescent to star-forming
ξ_3	Normal(0, 2^2)	Slope of Mass and Density interaction in log-odds of quiescent to star-forming
$\zeta_{0,SB}$	Normal(2, 3^2)$\Theta(\zeta_{0,SB} - \zeta_{0,SF} - \sigma_{int,SF} - \sigma_{int,SB})$	intercept in log(SFR) vs log($\frac{M_s}{10^{10} M_\odot}$) relation for Star-burst (linear relation)
$\zeta_{0,SF}$	Normal(1.5, 3^2)	intercept in log(SFR) vs log($\frac{M_s}{10^{10} M_\odot}$) relation for Star-forming (linear relation)
$\zeta_{0,Q}$	Normal(1, 3^2)$\Theta(\zeta_{0,SF} - \zeta_{0,Q} - \sigma_{int,SF} - \sigma_{int,Q})$	intercept in log(SFR) vs log($\frac{M_s}{10^{10} M_\odot}$) relation for Quiescent (linear relation)
$\zeta_{1,SB}$	Normal(1, 1^2)$\Theta(\zeta_{1,SB} - \zeta_{1,SF})$	Slope in log(SFR) vs log($\frac{M_s}{10^{10} M_\odot}$) relation for Star-burst (linear relation)
$\zeta_{1,SF}$	Normal(1, 1^2)	Slope in log(SFR) vs log($\frac{M_s}{10^{10} M_\odot}$) relation for Star-forming (linear relation)
$\zeta_{1,Q}$	Normal(1, 1^2)$\Theta(\zeta_{1,SF} - \zeta_{1,Q})$	Slope in log(SFR) vs log($\frac{M_s}{10^{10} M_\odot}$) relation for Quiescent (linear relation)
$\sigma_{int,SB}$	Half-Normal(0, 2^2)	Intrinsic scatter in SFR vs M relation for Star-burst.
$\sigma_{int,SF}$	Half-Normal(0, 2^2)	Intrinsic scatter in SFR vs M relation for Star-forming.
$\sigma_{int,Q}$	Half-Normal(0, 2^2)	Intrinsic scatter in SFR vs M relation for Quiescent.
α	Normal(-1, 2^2)	The shape parameter of Schechter Stellar Mass Function.
M_*	log-Normal(ln(5×10^{10}), 2^2)	The scale parameter of Schechter Stellar Mass Function.
μ_Δ	Normal(0, 2^2)	The location parameter of Density contrast distribution.
σ_Δ	Half-Normal(0, 2^2)	The scale parameter of Density contrast distribution.
$\alpha_\Delta*$	Normal(0, 5^2)	The shape parameter of Density contrast distribution.

To make sure that the model used in this work is well specified and the algorithm is able to infer the true parameters, we use few simulations drawn from the model prior and infer the posterior distribution over the parameters which are showed to provide a reasonable estimate of the parameters (i.e., simulation based calibration Cook et al. 2006; Talts et al. 2018). Given that the description of the single model for capturing all the variability of the galaxies in transition to quiescent galaxies is not adequate, we use it as a model to capture the galaxies that are not part of the star forming sequence. Given these consideration, we also use samples from predictive posterior to simulate the observed space and compare it directly to observed properties to ensure that resulting predictive posterior is consistent with the observations and can reasonably reproduce the structure seen in the observation.

4.4 Results

In this section, we discuss the results from the inference on the discussed probabilistic model, and Figure 4.1 shows the resulting posterior distribution of the parameters at highest redshift bin. The inference on the resulting model can be separated into three parts: 1. Inference on the parameters of the star-forming sequence using the mixture of three components. 2. Inference of the population parameters such as the stellar mass probability distribution. 2. Inference on the modeling parameters of the galaxy type probability given their stellar mass and environment.

4.4.1 Star-Formation Rate versus Stellar Mass

Using the samples from the posterior distribution of the model's parameters and latent variables, one can find the parameters of the model for the mean of the $\log(SFR)$, and intrinsic scatter around the mean relation by marginalizing over the rest of the parameters in the entire joint probability distribution.

The known overall evolution of the mean relations is consistent with that galaxies have higher SFR on average at higher redshift. The Intrinsic scatter around the SFR forming the main sequence are 0.28, 0.24, 0.23, 0.24, 0.18 dex from the lowest redshift bin to the highest. The intrinsic scatter around the quiescent and starburst are more diverse and have values as high as 0.44 dex for the starburst at the lowest redshift bin. However, upon further investigation, we find that these galaxies have much higher SFR than their SED fitting predictions. The rest of the intrinsic scatters about 0.2-0.3 dex. However, we note that the intrinsic scatter around the quiescent can be relied on to a small degree as these

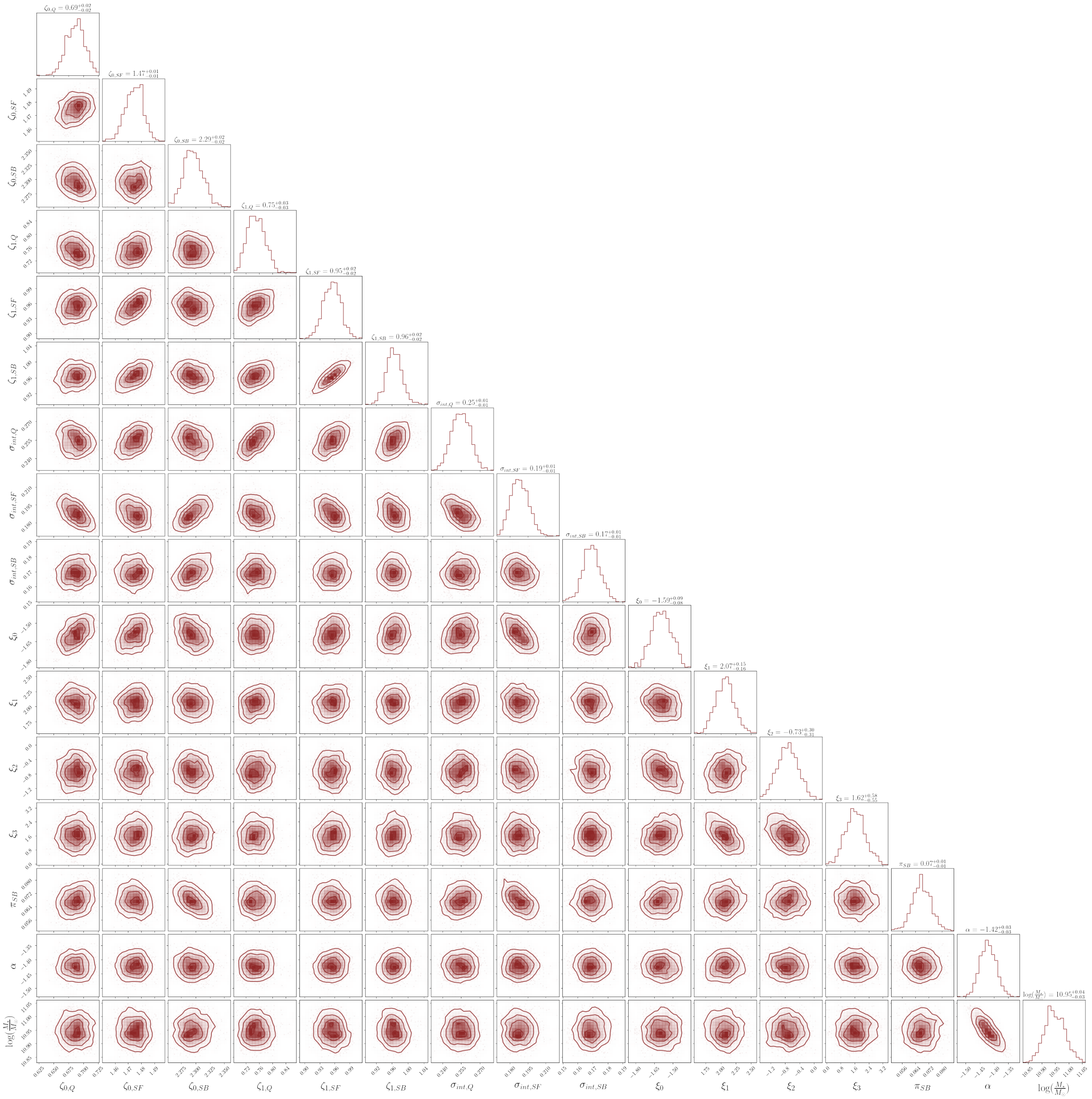


Figure 4.1: Posterior distribution of model parameters for the highest redshift bin.

galaxies can have SFRs so low that modeling them with a single line and scatter around it would be meaningless, and one should separate them as a different class.

Figure 4.2 shows the distribution of the data, and 30 posterior draws of the mixture model for the mean and two sigmas scatter. The galaxies are colored according to their posterior probability of belonging to each group. This shows an agreement with the overall mixture of these mean relations with intrinsic scatter compared to data and can describe the overall distribution reasonably well. We note that the galaxies are not assigned to each relation and have a probability of belonging to each of these mixtures and these uncertainties flow through the rest of the model in consistent way (i.e., especially near the boundary and for those with high SFR uncertainties).

4.4.2 Stellar Mass Functions

Using the inference results, one can easily translate the probability distribution over the masses to the stellar mass function using its definition. Figure 4.3 shows the evolution of the stellar mass distribution of the galaxies as a function of redshift. As the figure shows, we find a good agreement for the stellar mass function compared to the previous result in the literature. However, due to the lack of very massive galaxies in the high depth survey such as CANDELS, the massive end of the mass function does not have much support from the data as one would get from a wider field. We note that the stellar mass uncertainties are consistently considered here without any requirement for the binning in the stellar mass.

To calculate the stellar mass function for different galaxy types, we need to consider the probability of the galaxy type given its stellar mass and density contrast. However, since

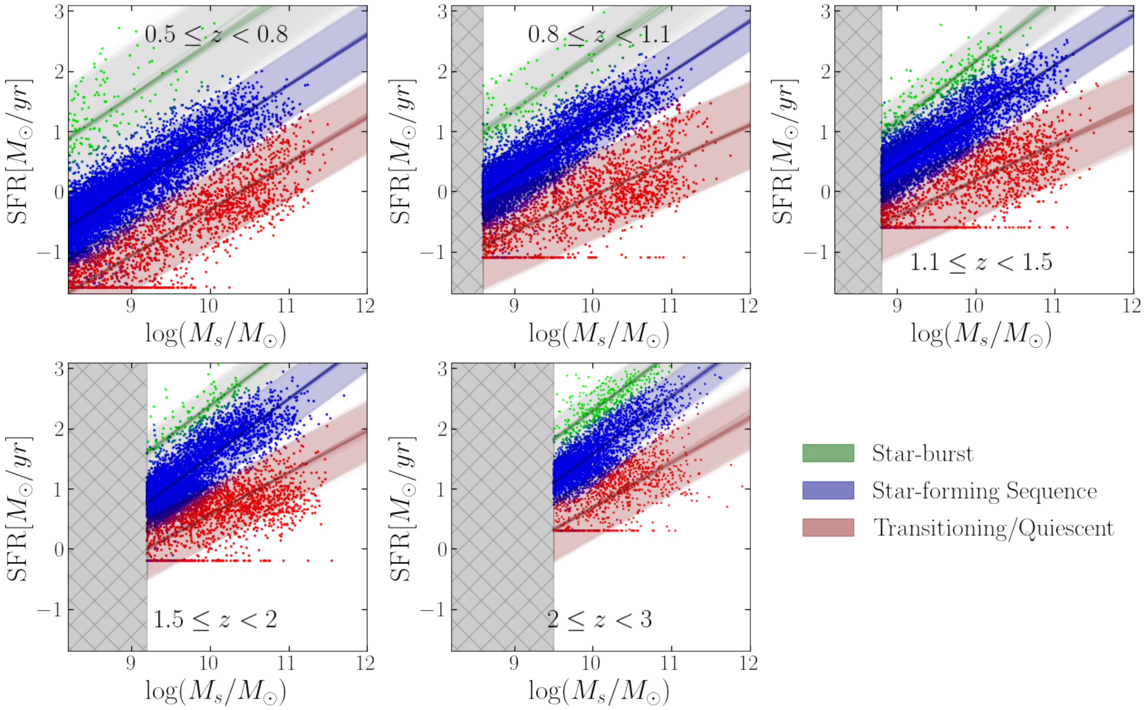


Figure 4.2: Star-formation versus Stellar Mass over cosmic time: The data points are colored with (Red, Green, Blue) proportional to their probability simplex (p_{Q}, p_{SB}, p_{SF}). Lines and opaque regions show random draws from the posterior distribution of the mean and 2-σ intrinsic scatter around it. The hatched grey area shows the mass completeness limit.

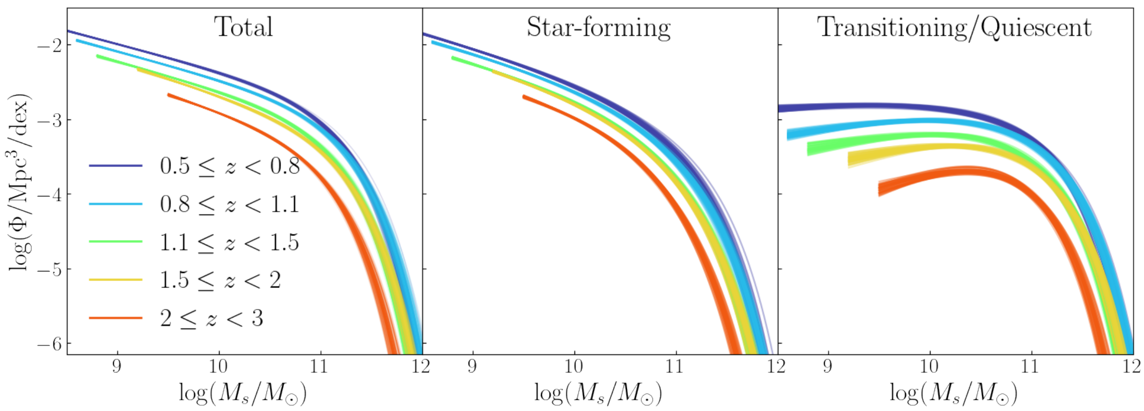


Figure 4.3: Stellar Mass function of the Galaxies: The Left panel shows the stellar-mass function calculated based on the probability distribution over the galaxies masses and the survey volume. The Middle/Right panel shows the stellar mass function of different galaxy types calculated using the probability of galaxy type given their stellar mass and density contrast and after integrating over the density contrast distribution.

we are interested in the overall shape of the mass functions averaged over their densities, we integrate over all values of the density contrast multiplied by the chance of the being in each class and the overall stellar mass probability. For doing that, we use the posterior draws, and after integrating the density contrast, we show the resulting posterior distribution of the galaxies as a function of their stellar mass after translating it to the stellar mass function.

The middle panel in Figure 4.3 shows the inferred stellar mass function of the star-forming galaxies that are on the star-forming sequence. The right panel shows the stellar mass function for the transitioning/Quiescent part of the SFR vs. stellar mass and their rapid build-up as we move toward lower redshift. It should be noted that this definition of the galaxy population is not necessarily equivalent to the population one might get using various color-color separation schemes or putting a hard threshold on their specific star-formation rates.

4.4.3 Quiescent Probability versus Stellar Mass and Local Density

The probability of a given galaxy to belong to a transitioning/quiescent population compared to the star-forming sequence, given their physical properties, can shed light on some of the processes involved in this transition. Here we describe the inferred probability that a given galaxy to have a certain type as a function of its stellar mass and environment traced by the density contrast. As the Figures 4.6 and 4.5 shows, we find that the probability of quiescent increases with the stellar mass at every environment probed here and across cosmic time. We also find that the galaxies with stellar masses higher than $\sim 10^{10.5} M_{\odot}$ have a higher chance of being in transition if they are in a dense region. These are all consistent with many previous studies (e.g., Peng et al. 2010; Darvish et al. 2016; Kawinwanichakij et al. 2017). For galaxies below this mass, this probability increase up to the redshift ~ 1.2 and after that starts to decrease gradually. This points to a possible reversal of transitioning probability with the density contrast at low masses, in the sense that low mass galaxies are less likely to be transitioning/quiescent if they are in the denser environment.

Figure 4.4 shows the evolution of the coefficient in the log odds of quiescent to star-forming relation and their marginal distribution, which encodes the behavior of the probability seen in the Figures 4.6, and 4.5. We find that the slope of the mass in this relation at $\log{(1+\Delta)} = 0$[5] increases gradually going to higher redshift, meaning that the effect of having a mass one dex higher changes the log-odds of finding a galaxy in the transitioning region compared to star-forming sequence from ~ 1 to ~ 2 from the redshift of 0.5 to 2.5. The slope of the density contrast at the mass of 10^{10} in this relation shows a

---

[5] when the interaction term is zero by construct

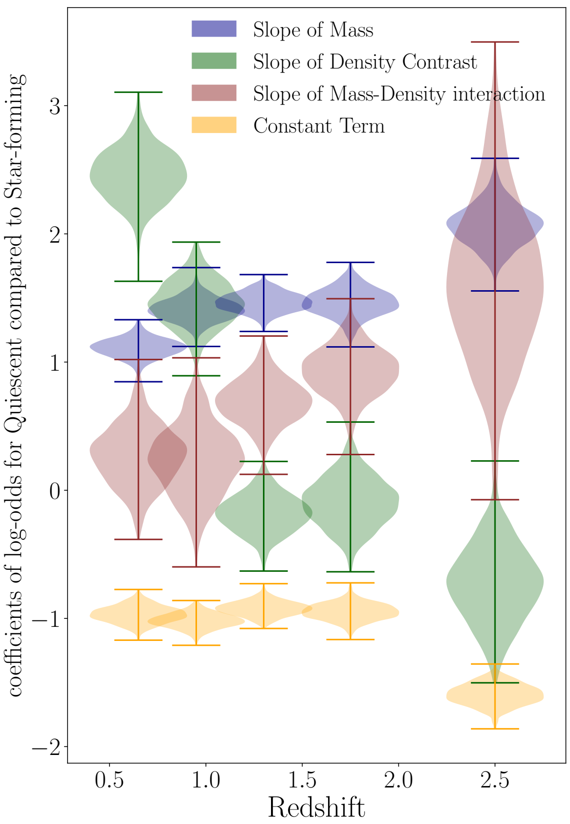


Figure 4.4: Coefficients of the relation describing the log-odds defined as the natural logarithm of the probability of being transitioning/quiescent to probability of being on the star-forming sequence as function of stellar mass and local density contrast or $\ln(\frac{\pi_Q}{\pi_{SF}})$. One should note that these coefficients have different physical units and should be compared to each other separately.

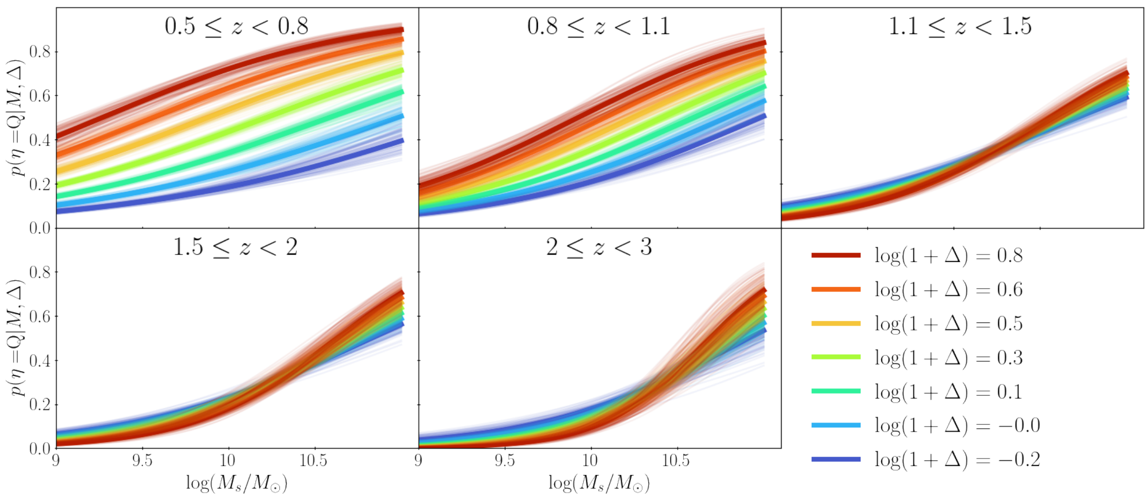


Figure 4.5: Quiescent Probability versus Stellar Mass

significant evolution from the redshift of 0.5 to 2.5[6]. We find that moving a galaxy with a mass of 10^{10} by one unit in the $\log(1+\Delta)$ changes the log-odds of being in transition from ~ 2.5 to ~ -1 from the redshift of 0.5 to 2.5. This shows that the effect of the environment alone in increasing this log-odds evolves rapidly from being very large at low redshift to a decreasing effect at the highest bin (although the marginal are consistent with being zero given its uncertainties).

The effect of the interaction term between the density contrast and stellar mass shows almost opposite trends and qualitatively follows the evolution of the mass. The interaction parameter changes the slope in quiescent probability as a function of mass for different environments and the slope for this probability as a function of environment for galaxies of different masses. In other words, a non-zero interaction shows that the slopes of the environment and mass in log-odds relation depend on each other. This shows that the

---

[6] One should note, however, that these slopes all have physical units inverse of what they are being multiplied to and the physical range of environment and stellar mass in the log space is not the same, i.e., one dex in log density covers almost all the available range of environment probed in this study

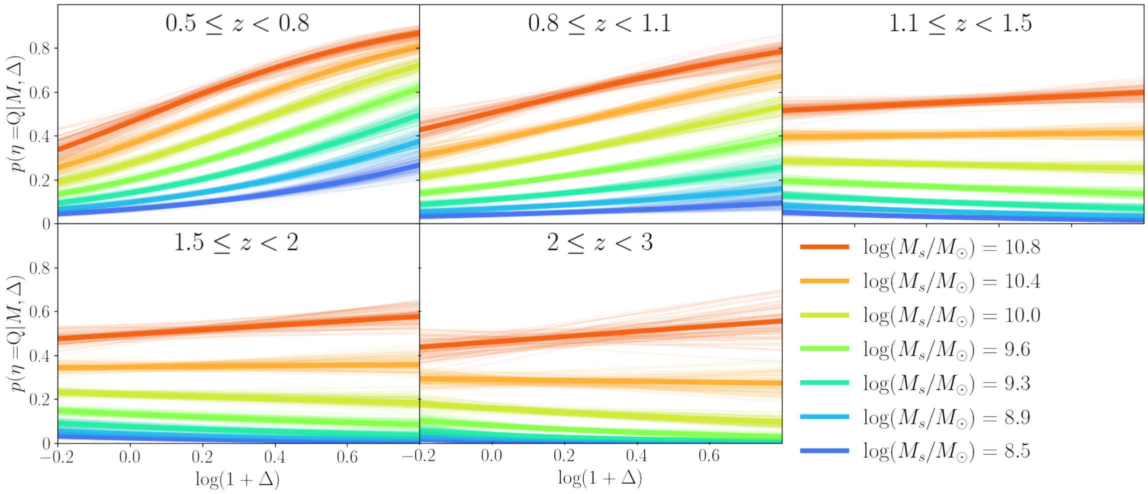


Figure 4.6: Quiescent Probability versus Density Contrast

effects of mass and environment are acting separately on how much they change the log-odds. However, this scenario changes as one moves to a higher redshift. At redshift higher than 1.2, the effect of the environment alone seems to have a small and negative contribution to the log-odds, but the interaction between the mass and environment becomes significantly more important. This means that the effect of the environment depends on the stellar mass (i.e., change of slope as a function of stellar mass), and higher stellar masses lead to increases in the effect of the density on the log-odds, and lower stellar masses decrease the effect of the environment. On the other hand, the effects of the stellar mass, traced by its slope in log-odds, depends on the environment and increases from under-dense to over-dense regions. The overall result could suggest that the increase in the log-odds of transition galaxies compared to star-forming might be related to differences in the underlying physical processes in the environmental quenching of the galaxies at a redshift of higher than ~ 1 in comparison to below it, however, only from higher abundances, we cannot pin down how many of these objects are in transition because of the environmental processes, how many were accreted onto denser environment as in transition galaxies [7], and how many are in transition because of the change in the efficiency of their internal mechanism at different environment.

This change in the importance of the interaction term that affects the slope of the environment is greater for more massive galaxies could imprint a particular behavior on the quantity called environmental quenching efficiency. This is the ratio of the excess of the fraction of quenched (in our case transitioning and quiescent) galaxies in the over-dense region compared to the under-dense region to the fraction of the star-forming galaxies in the under-dense region. Similarly, one can define a mass quenching efficiency as well. Here are the definitions based on our probabilistic model:

$$\epsilon_{env}(M) = \frac{p(\eta = Q|M, \Delta > \Delta_{75}) - p(\eta = Q|M, \Delta < \Delta_{25})}{1 - p(\eta = Q|M, \Delta < \Delta_{25})}$$

$$\epsilon_{mass}(M) = \frac{p(\eta = Q|M, \Delta < \Delta_{75}) - p(\eta = Q|M_{min}, \Delta < \Delta_{75})}{1 - p(\eta = Q|M_{min}, \Delta < \Delta_{75})}$$

We used our definition of the probability of being quiescent/transitioning given the stellar mass and density contrast. We use the average of this probability over the lower %25 quantiles of the density distribution (Δ_{25}) as the quiescent probability in the under-dense region[8], and the average over-densities higher than %75 quantiles (Δ_{75}) the quiescent probability in the over-dense region.

---

[7] which would imply that the observed fraction would be highly dependent on the fraction of these galaxies within the over-density, before accretion

[8] same definition for star-forming in under-dense regions

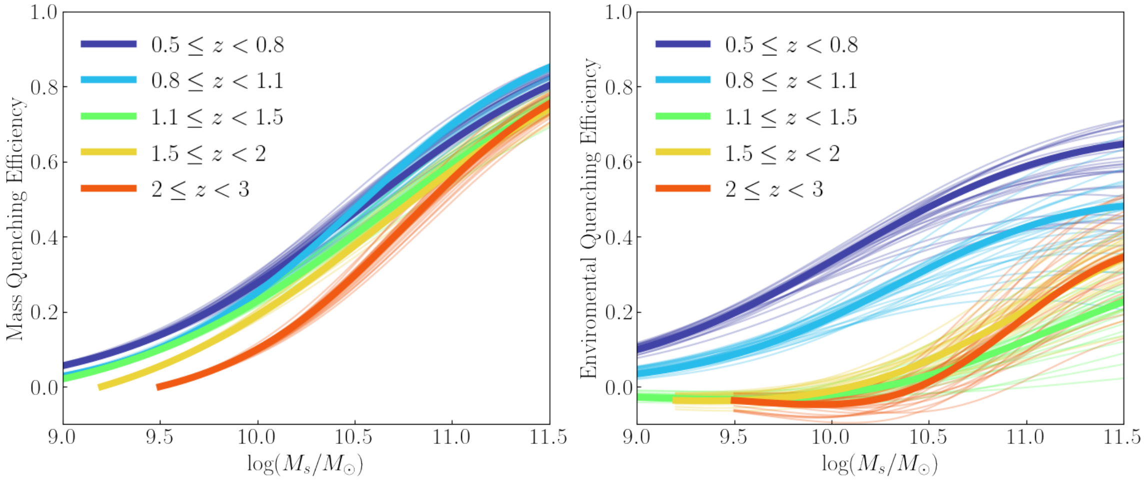


Figure 4.7: Quenching efficiency of Environment and Mass as function of Stellar Mass.

However, we only find a positive effect from the environment at high redshift on increasing the odds of a galaxy to be at quiescent/transition region for massive galaxies, and the less massive galaxies are found to get zero or negative effect from the environment. Given the limited dynamic range of galaxies' stellar mass and the significant uncertainties on the parameters, this can only be suggestive.

As the Figure 4.7 shows, we find that the environmental quenching efficiency increase as a function of redshift at lower masses and becomes more important than the mass quenching efficiency at lower masses, similar to what have been reported before (e.g., Darvish et al. 2016; Kawinwanichakij et al. 2017; Chartab et al. 2020). The uncertainties in this relation is quite large at high stellar mass but given that its evolution seems to indicate that the environmental quenching efficiency increases with stellar mass, and it is a steep relation at redshift higher than 1 but tends to flatten out towards the lower redshift. For example this efficiency increases from about 0 to $20-30\%$ when comparing galaxies of $10^{10}M_\odot$ to 10^{11} for redshift higher than 1.5 but for the lowest redshift bins this value decrease to only

$0 - 10\%$. The fact that the environmental quenching efficiency increases with stellar mass shows another way expressing the consistency with the "overconspumption" scenario (McGee et al. 2014), which were reported in previous studies as well (e.g., Kawinwanichakij et al. 2017; Chartab et al. 2020). In this scenario, galaxies that are accreted onto a larger halo lose their gas given the lack of fresh gas accretion and deplete their remaining gas through star-formation and its related outflows. This scenario predicts that since the galaxies with higher masses have higher star-formation rate, they tend to deplete their gas faster and start the transition to becoming quiescent faster.

However, Belli et al. 2017b have found that the quiescent galaxies at high redshift seem to go through two distinct quenching processes in terms of their time scale, and fast quenching is likely happening after violent gas-rich mergers that lead to their compaction and going through the blue nugget phase and turning to the red nugget (see Dekel et al. 2019). Another quenching channel is related to the depletion of their gas reservoir through star-formation, which is similar to the overconsumption scenario discussed here. Additionally, there seems to be growing evidence of such galaxies with a low amount of gas at redshift higher than ~ 1.5 (Bezanson et al. 2019; Whitaker et al. 2021), which is consistent with this picture.

We note that in this study, the measurements uncertainties have been treated consistently through the model; however, there are significant systematic uncertainties in the data that are not considered within the model, such as: 1. Measurement uncertainties in the photometric redshifts, 1. density contrasts are assumed to be exact and without any uncertainties, 3. The SFR and Mass measurements could have significant systematic uncertainties, and their associated estimation of uncertainties might not be accurate, which is especially true for the quiescent population where the UV+IR overestimates the SFR because of the older stellar population heating the dust and the subsequent contribution to longer wavelengths (e.g., Fumagalli et al. 2014). Also, the model used here is a very simplistic representation of reality and has some major missing parts, such as the independence assumption between the stellar mass distribution and density contrast. We note that the independence assumption was made after trying to model the stellar mass distribution parameters as a function of density contrast, which resulted in a dependence that was hardly constrained by the data. Also, we note that given these limitations of the model and the quality of the data, and the fact that we have a wide posterior distribution of the coefficient, particularly for the environment term and interaction between environment and stellar mass, the result presented here should be interpreted as only suggestive and accurate under these significant assumptions.

4.5 Discussion and Conclusion

In this study, we use the latest available catalogs of physical measurements in the CANDELS, including homogeneous photometric redshift, local environment measurements, and star-formation rate measurements from UV+IR, to build a consistent probabilistic model for the dependence of the star-formation activity of the galaxies on their local environment and stellar mass in the presence of the noisy and censored data. We simultaneously model the star-forming versus mass plane using a mixture of three components consisting of 1. star-burst, 2. star-forming, 3. transitioning/quiescent. We also model the probability that a given galaxy belongs to a quiescent or star-forming population based on its environment and stellar mass. In order to take into account the effect of noisy measurements in stellar mass, we simultaneously model the stellar mass distribution for the CANDELS galaxy population and find these relations between the "true" physical parameters of the galaxies by treating them as latent variables within the model and marginalizing over those consistently.

We find a general agreement between the probability of transitioning/quiescent galaxies (i.e., quiescent fraction) given their stellar mass and local environment to what has been found in the literature. Some of these are that at all redshifts and for all environments, an increase in the stellar mass increases the probability of being in the transitioning/quiescent population. We also find that out to a redshift of ~ 3 at masses larger than $\sim 10^{10} M_{\odot}$, higher density contrasts increase the chance of being in transition/quiescent population. At lower masses, the situation seems to be reversed around the redshift of $\sim 1 - 1.5$, and for galaxies with lower stellar masses, a denser environment lowers the chance of being a quiescent galaxy, but this is a small effect.

We find that the effects of the environment traced by its slope in the log-odds of the quiescent to star-forming have significant evolution, and the environment by itself is an essential factor to determine what fraction of the galaxies are in transition at all redshifts, but its importance depends on the stellar mass above redshift of ~ 1.2. At lower redshifts, the slope seems independent of the stellar mass, and galaxies of different stellar mass experience the same environmental factor in their log-odds. However, at high redshift, there is a positive correlation between the slope of the environment as a function of stellar mass that the effects of the environment are more significant for more massive galaxies. For the low-mass galaxies, this slope becomes negative in the sense that moving to a denser environment at a given stellar mass decreases the odds of being quiescent.

For the slope of the stellar mass in the model for the log-odds, we find that this slope is constant at lower redshift in the sense that different environments do not change the effect from the stellar mass in making them quiescent more likely to be quiescent. At higher redshift, this slope depends on the local environment such that the effect of the stellar mass increases at the denser environment, and in an over-dense environment, increasing the mass increases the odds of being in transition faster than under-dense.

The strong interaction between the effects from the environment and stellar mass at high redshifts, which makes more massive galaxies move off from the star-forming sequence more efficiently if located in a denser environment, can be interpreted as consistent with the "overconsumption" scenario McGee et al. 2014) for the quenching of the most massive galaxies. In this scenario, those galaxies that cannot get refueled through cosmological accretion either as they become a satellite of a larger halo or because their halo grows and is subsequently more susceptible to shock heating of the accreted gas. These galaxies deplete their remaining gas through star-formation and outflows. The effect of the environment in this setting depends on the galaxies' stellar mass and redshift as more massive galaxies tend to have a higher star-formation rate and this rate increases with redshift. This is consistent with the growing evidence of such a population of quiescent galaxies with low gas content at a redshift of above ~ 1.5 (e.g., Bezanson et al. 2019; Whitaker et al. 2021). Also, as we find that the effect of the mass in determining the log-odds of being quiescent compare to star-forming is larger in denser environment, it might suggest that the mass quenching mechanisms such as AGN feedback can work more efficiently at denser environment. And if the effect of environment is related to the amount of gas accretion, we speculate that increase in this effect for more massive galaxies at denser environment comes from the AGN that can keep the gas in the halo hot and increase the cooling time more efficiently.

However, one needs to consider other physical properties of galaxies, such as their morphology and how this changes in a different environment in order to rule out other scenarios such as mergers as primary mechanism. For example, what fraction of the quiescent galaxies go through a fast morphological transformation through mergers and subsequently stop their star-formation following the compaction and possibly a quasar phase, in comparison to less eventful evolution such as depleting their gas reservoir through star-formation and outflows. For example, Darvish et al. (2016) has argued that the increase in the efficiency of environmental quenching with stellar mass and increase in the mass quenching efficiency at denser environment can be explain by mergers, and is consistent with the results from this study. Also, Kawinwanichakij et al. (2017), has found that the quiescent galaxies in the denser environment tend to have similar morphology to the field quiescent, and since the overconsumption scenario does not change the morphology of a galaxy, another process must be invoked to change the morphology of the quiescent. On the other hand, there seems to be an indication from the spatially resolved kinematic studies of gravitationally lensed quiescent galaxies at redshift > 2, that they are rotationally supported (Toft et al. 2017; Newman et al. 2018b), which can separate the quenching mechanism from their morphological transformation at least in some galaxies.

Putting a better constraint on these effects requires a more extensive set of data. Understanding the complex interactions between different galaxies' properties requires detailed modeling of observed quantities, which requires enough high-quality data over a variety of physical properties. Future surveys allow us to build much better constraints on the simplistic observational model used here and consider the effects of other physical properties of galaxies and their interactions simultaneously by modeling their joint probability distribution while accounting for different types of noise in the measurements.

4.6 Future Extension

One way to extend this analysis and provide a better constraints on the parameters of the model, such as the interaction term in the log-odds would be to use all the data available by modeling the selection effects consistently. Here we discuss the model in such a scenario where we can extract extra information by modeling the selection effects in photometrically selected catalogs.

4.6.1 Extension of the Probabilistic Model

To describe the full probabilistic joint distribution, we start with the joint probability over all the galaxies within the population where some of them happened to be observed in our sample and some are missing (see Gelman 2004; Kelly et al. 2008; Mantz 2019; Little and Rubin 2019, chapter 15); assuming that the full distribution over the population can be separated into the observed and unobserved probability distributions as the following:

$$p(\{M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \Delta_i, I_i, \eta_i\}_{i=1}^{N}|\boldsymbol{\theta}, N) =$$

$$\binom{N}{N_{obs}} p(\{M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \Delta_i, I_i = 1, \eta_i\}_{i=1}^{N_{obs}}|\boldsymbol{\theta}) \times p(\{M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \Delta_i, I_i = 0, \eta_i\}_{i=N_{obs}+1}^{N}|\boldsymbol{\theta}, N)$$

where the $\{\hat{M}_i, \hat{\psi}_i, \Delta_i\}_{i=1}^{N_{obs}}$ is the set of *observed physical properties* of N_{obs} galaxies in the sample same as before but now we use the full sample instead of limiting to mass complete sample with $\hat{M}_i$, $\hat{\psi}_i, \Delta_i$ are the stellar mass, star-formation rate, and local density contrast of galaxy i respectively. I_i is the indicator binary variable that shows whether galaxy i has ended up in the sample (i.e,. $I_i = 1$). $\{\hat{M}_i, \hat{\psi}_i, \Delta_i\}_{i=N_{obs}+1}^{N}$ is the set of physical properties of $N - N_{obs}$ galaxies within the population that has not been observed and we have to marginalize over them (i.e., population has N galaxies and we observed N_{obs}). The binomial factor comes in since we have an exchangeable sequence of observed galaxies[9] selected through a random process from a larger population. This random process can depend on the physical properties of the galaxies and the survey detection efficiency. η_i is a categorical variable that indicates the galaxy type (i.e., star-forming, star-burst, and quiescent). $\boldsymbol{\theta}$ stands for all the parameters used in the model and $\{M_i, \psi_i\}$ are the *latent variables of the model for true physical properties* of the galaxies that are not directly observed and need to be evaluated within the probabilistic model.

Since, the latent variables and the unobserved galaxies are not known we need to marginalize over them in order to find the likelihood of the parameters based only on the actual observations.

---

[9] Different ways that a galaxy can be permuted within the population and end up in the sample.

First we describe the conditional independence of physical properties to specify our model as the following:

$$p(M,\psi,\hat{M},\hat{\psi},\Delta,I,\eta|\boldsymbol{\theta}) = p(I|\hat{M},\eta)p(\hat{M},\hat{\psi}|M,\psi)p(\psi|M,\eta,\boldsymbol{\xi})p(\eta|M,\Delta,\boldsymbol{\zeta})p(M|\boldsymbol{\alpha})p(\Delta|\boldsymbol{\mu})$$

where $p(I|\hat{M},\eta)$ is the probability that a galaxy with a given stellar mass and galaxy type has been observed in the sample or not and must be estimated through modeling the survey detection probability.

We define the likelihood function of the parameters for the observed galaxies as the following after marginalizing the categorical variable of galaxy type and assuming that the galaxies are independent and identically distributed according the joint distribution over their individual physical properties:

$$\mathcal{L}_{obs} = p(\{M_i,\psi_i,\hat{M}_i,\hat{\psi}_i,\Delta_i,I_i=1\}_{i=1}^{N_{obs}}|\boldsymbol{\theta}) = \prod_{i=1}^{N_{obs}}\sum_{\eta} p(M_i,\psi_i,\hat{M}_i,\hat{\psi}_i,\Delta_i,I_i=1,\eta_i=\eta|\boldsymbol{\theta})$$

To marginalize over the unobserved part of the joint probability distribution we need to sum over all the nuisance parameters which gives us the overall probability of not selecting a galaxy in the sample:

$$p(I=0|\boldsymbol{\theta}) = \sum_{\eta}\int p(M,\psi,\hat{M},\hat{\psi},\Delta,I=0,\eta|\boldsymbol{\theta})dMd\psi d\hat{M}d\hat{\psi}d\Delta$$

In order to make these integrals numerically tractable, we impose some major assumptions. First, we assume that the measurements uncertainties are independent and can be separated $p(\hat{M}, \hat{\psi}|M, \psi) = p(\hat{M}|M)p(\hat{\psi}|\psi)$. This allows applying the law of total probability over the $\hat{\psi}$ and ψ, which reduces the integral dimensions from five to three. Furthermore, we assume that only for the unobserved galaxies the distribution of the stellar mass measurement given the true stellar masses can be described as the delta function, which reduces the three dimensional integral into an integral over two dimensions which can be numerically estimated during the inference as a function of model parameters. The final probability of not been selected in the sample is:

$$p(I = 0|\boldsymbol{\theta}) = \sum_{\eta} \int p(I = 0|M, \eta)p(\eta|M, \Delta, \boldsymbol{\xi})p(M|\boldsymbol{\alpha})p(\Delta|\boldsymbol{\mu})dMd\Delta$$

Now we can describe the joint probability distribution over the population after marginalizing the physical properties of the unobserved galaxies as the following:

$$p(\{M_i, \psi_i, \hat{M}_i, \hat{\psi}_i, \Delta_i, I_i = 1, \eta_i\}_{i=1}^{N_{obs}}, \{I_i = 0\}_{i=N_{obs}+1}^{N}|\boldsymbol{\theta}, N) = \binom{N}{N_{obs}} p(I = 0|\boldsymbol{\theta})^{N-N_{obs}} \mathcal{L}_{obs}$$

In order to make the above model fully based on the observed only variables we make some assumption about the number of objects within the population and some prior over their parameters. In reality, this number is regulated by the cosmology and astrophysical processes involved and should be properly modeled. However, one can make an ignorance assumption about this number and try to infer this number within the model. One simple assumption is to model this number as a Poisson distributed random variable with rate λ.

These allows the marginalization over the possible number of galaxies within the population as the following (look at Mantz 2019 for full derivation):

$$\mathcal{L} = \sum_{N=N_{obs}}^{\infty} \binom{N}{N_{obs}} p(N) p(I=0|\boldsymbol{\theta})^{N-N_{obs}} \mathcal{L}_{obs} = \sum_{N=N_{obs}}^{\infty} \binom{N}{N_{obs}} \frac{e^{-\lambda}\lambda^N}{N!} p(I=0|\boldsymbol{\theta})^{N-N_{obs}} \mathcal{L}_{obs}$$

$$\propto e^{-\lambda p(I=1|\boldsymbol{\theta})} \lambda^{N_{obs}} \mathcal{L}_{obs}$$

In addition, assuming a gamma distribution with the shape and rate parameters α, β that is the conjugate prior for Poisson distribution yields an analytical solution for marginalizing the Poisson parameter. The resulting marginalization is shown as the following (see Mantz 2019 equation 8):

$$\mathcal{L}_f = \int_0^{\infty} d\lambda p(\lambda) \mathcal{L} \propto (p(I=1|\boldsymbol{\theta}) + \beta)^{-(N_{obs}+\alpha)} \mathcal{L}_{obs}$$

For fitting the model one can further assume the parameters of the gamma distribution are $\beta \to 0$, and $\alpha = 1/2$, which yields a Jeffreys prior (invariant under different parametrization by construction) over the Poisson parameter that is $p(\lambda) \propto 1/\sqrt{\lambda}$. The $\mathcal{L}_f$ is the final likelihood function that take into account the selection function as well as the likelihood of the parameters given the observations.

Also one can easily extend the stellar mass distribution to better capture the full shape of the stellar mass function using a double Schechter distribution. In order to build the double Schechter mass function, one can employ a mixture model of two Schechter distribution component as the following:

$$p(M|\boldsymbol{\alpha}) = \Upsilon p(M|\alpha_1, M_*) + (1 - \Upsilon)p(M|\alpha_2, M_*)$$

where the parameters are grouped as $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, M_*, \Upsilon\}$ which are the lower mass slope for two component, their shared exponential cut-off scale parameter, and overall probability of the first component respectively.

For extending this for all the available galaxies and not only those that are mass complete, we need to model that as a function of stellar mass, which can be done using either mock observational catalogs or having access to a deeper survey to build an empirical model. However, in order to make use of the very efficient Hamiltonian Monte Carlo sampler, one needs to make a accurate enough selection function that at least has first order differentiable function.

Chapter 5

Summary and Conclusions

The focus of this thesis is on the selections and study of the galaxies at different stages of their star-formation activity at high redshift. In chapters 2, and 3 we focused on selection of massive quiescent galaxies at redshifts higher than 3, starting from more traditional approach for their selections in 2 to the use of statistical learning approach in 3 that can provide a robust sample of these galaxies while being more computationally efficient. In chapter 4, we turned our focus on the galaxies below the redshifts of 3 and provide a probabilistic modeling framework that can simultaneously take into account sources of measurement uncertainties, without having to use a pre-defined notion of galaxy types based on their star-formation activity to extract information about the role of environment in shaping the evolution of galaxies, something that are not usually considered when studying these environmental effects.

5.1 Summary of Chapter 2

Using the CANDELS photometric catalogs for the *HST*/ACS and WFC3, we identified massive evolved galaxies at $3 < z < 4.5$, employing three different selection methods. We find the comoving number density of these objects to be $\sim 2 \times 10^{-5}$ and $8 \times 10^{-6} Mpc^{-3}$ after correction for completeness for two redshift bins centered at $z = 3.4, 4.7$. We quantify a measure of how much confidence we should have for each candidate galaxy from different selections and what are the conservative error estimates propagated into our selection. Then we compare the evolution of the corresponding number densities and their stellar mass density with numerical simulations, semi-analytical models, and previous observational estimates, which shows slight tension at higher redshifts as the models tend to underestimate the number and mass densities. However, a larger simulation volume is needed, since the estimates of these values in simulation at high redshift are subject to significant Poisson uncertainties, in addition to the need for a larger sample of these objects, which are going to be available through the upcoming surveys.

5.2 Summary of Chapter 3

This chapter investigates the ability of statistical learning techniques applied to simulated photometric catalogs to identify a sub-population of galaxies based on their multiwavelength photometric measurements. We use the simulated catalogs from the Santa Cruz semi-analytical model of galaxy formation created to match CANDELS GOODS-South observations. We further modified the catalog to match the observations in terms of their uncertainties in photometries and the fraction of the missing values at different bands and colors. We use an ensemble of random forest regressors for different bands using an iterative approach called Multiple Imputation by Chained Equations (MICE) to find an estimate of all the missing values, which we show to be very effective using the knowledge of the ground truth. We then use massive quiescent galaxies at high redshift as an example sub-population, and we find several decision boundaries varying in their complexity. We also find the selection based on the physical properties inferred from LEPHARE's Spectral Energy Distribution fitting to compare all these methods based on their ability to recover the actual population and their robustness to contaminants. We find that Bayesian Logistic Regression, Random Forest, Support Vector Machines, and Neural Network outperform SED fitting on the complete data with known redshifts. Given the computational cost of more robust approaches to SED-fitting for the extensive incoming multiwavelength surveys, the statistical learning techniques trained on the high-fidelity simulated catalogs offer great potential for selection and further studies (e.g., spectroscopic) of specific sub-populations of galaxies.

5.3 Summary of Chapter 4

Using the latest measurements of star-formation rates, photometric redshifts, and local density in the CANDELS, we provide a probabilistic modeling framework for extracting the role of galaxies local environment in dictating the galaxies type based on their star-formation activity, while taking into account the measurements uncertainties as well. We find that the odds of a galaxy being below the star-forming sequence versus on it, depends on their stellar mass out to the redshift of 3 and the effect of the stellar mass alone is higher at denser environments. This odds depends on the environment out to the redshift of 1 and at higher redshift this dependence become much smaller. However, this dependence show strong dependence on the stellar mass and more massive galaxies ($M > 10^{10.5} M_{\odot}$) are more likely to be quiescent of they move to a denser environment. Overall, we find that the effect of the stellar mass independent of the environment in changing this odds is strong at all redshift considered ($0.5 < z < 3$), but for the environment this effects is mostly mass independent at redshift. We also find that the mass quenching efficiency depends strongly on the stellar mass at all redshift, and we find small evolution of this efficiency with redshift particularly for higher stellar masses. We also find that the environmental quenching efficiency depends on the stellar mass at all stellar masses above the mass in which our sample is mass complete for redshift below ~ 1 and at lower masses this efficiency is higher than mass quenching efficiency. At redshift higher than ~ 1, however, the environmental quenching efficiency grows with increasing the stellar mass only for galaxies with $M > 10^{10.5} M_{\odot}$.

Bibliography

Abadi, M., Agarwal, A., Barham, P., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Agrawal, A., Verschueren, R., Diamond, S., and Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60.

Allingham, J. U. (2018). Unsupervised automatic dataset repair. Master's thesis, Computer Laboratory, University of Cambridge.

Alt, H. and Godau, M. (1995). Computing the frÉchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 05(01n02):75–91.

Amit, Y. and Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7):1545–1588.

Arnouts, S., Cristiani, S., Moscardini, L., et al. (1999). Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North. *MNRAS*, 310(2):540–556.

Arnouts, S., Le Floc'h, E., Chevallard, J., Johnson, B. D., Ilbert, O., Treyer, M., Aussel, H., Capak, P., Sanders, D. B., Scoville, N., McCracken, H. J., Milliard, B., Pozzetti, L., and Salvato, M. (2013). Encoding of the infrared excess in the NUVrK color diagram for star-forming galaxies. *A&A*, 558:A67.

Bach, F. R. and Jordan, M. I. (2005). Predictive low-rank decomposition for kernel methods. page 33–40.

Baldry, I. K., Balogh, M. L., Bower, R. G., Glazebrook, K., Nichol, R. C., Bamford, S. P., and Budavari, T. (2006). Galaxy bimodality versus stellar mass and environment. *MNRAS*, 373(2):469–483.

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., and Szalay, A. S. (2004). Quantifying the Bimodal Color-Magnitude Distribution of Galaxies. *ApJ*, 600(2):681–694.

Ball, N. M. and Brunner, R. J. (2010). Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19(07):1049–1106.

Ball, N. M., Loveday, J., Fukugita, M., Nakamura, O., Okamura, S., Brinkmann, J., and Brunner, R. J. (2004). Galaxy types in the Sloan Digital Sky Survey using supervised artificial neural networks. *MNRAS*, 348(3):1038–1046.

Balogh, M. L., Baldry, I. K., Nichol, R., Miller, C., Bower, R., and Glazebrook, K. (2004). The Bimodal Galaxy Color Distribution: Dependence on Luminosity and Environment. *ApJL*, 615(2):L101–L104.

Balogh, M. L., McGee, S. L., Mok, A., Muzzin, A., van der Burg, R. F. J., Bower, R. G., Finoguenov, A., Hoekstra, H., Lidman, C., Mulchaey, J. S., Noble, A., Parker, L. C., Tanaka, M., Wilman, D. J., Webb, T., Wilson, G., and Yee, H. K. C. (2016). Evidence for a change in the dominant satellite galaxy quenching mechanism at z = 1. *MNRAS*, 456(4):4364–4376.

Balogh, M. L., Navarro, J. F., and Morris, S. L. (2000). The Origin of Star Formation Gradients in Rich Galaxy Clusters. *ApJ*, 540(1):113–121.

Barkana, R. and Loeb, A. (1999). The Photoevaporation of Dwarf Galaxies during Reionization. *ApJ*, 523(1):54–65.

Baron, D. (2019). Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, page arXiv:1904.07248.

Barro, G., Pérez-González, P. G., Cava, A., Brammer, G., Pandya, V., Eliche Moral, C., Esquej, P., Domínguez-Sánchez, H., Alcalde Pampliega, B., Guo, Y., Koekemoer, A. M., Trump, J. R., Ashby, M. L. N., Cardiel, N., Castellano, M., Conselice, C. J., Dickinson, M. E., Dolch, T., Donley, J. L., Espino Briones, N., Faber, S. M., Fazio, G. G., Ferguson, H., Finkelstein, S., Fontana, A., Galametz, A., Gardner, J. P., Gawiser, E., Giavalisco, M., Grazian, A., Grogin, N. A., Hathi, N. P., Hemmati, S., Hernán-Caballero, A., Kocevski, D., Koo, D. C., Kodra, D., Lee, K.-S., Lin, L., Lucas, R. A., Mobasher, B., McGrath, E. J., Nandra, K., Nayyeri, H., Newman, J. A., Pforr, J., Peth, M., Rafelski, M., Rodríguez-Munoz, L., Salvato, M., Stefanon, M., van der Wel, A., Willner, S. P., Wiklind, T., and Wuyts, S. (2019). The CANDELS/SHARDS Multiwavelength Catalog in GOODS-N: Photometry, Photometric Redshifts, Stellar Masses, Emission-line Fluxes, and Star Formation Rates. *ApJS*, 243(2):22.

Barro, G., Trump, J. R., Koo, D. C., Dekel, A., Kassin, S. A., Kocevski, D. D., Faber, S. M., van der Wel, A., Guo, Y., Pérez-González, P. G., Toloba, E., Fang, J. J., Pacifici, C., Simons, R., Campbell, R. D., Ceverino, D., Finkelstein, S. L., Goodrich, B., Kassis, M., Koekemoer, A. M., Konidaris, N. P., Livermore, R. C., Lyke, J. E., Mobasher, B., Nayyeri, H., Peth, M., Primack, J. R., Rizzi, L., Somerville, R. S., Wirth, G. D., and

Zolotov, A. (2014). Keck-I MOSFIRE Spectroscopy of Compact Star-forming Galaxies at z >~2: High Velocity Dispersions in Progenitors of Compact Quiescent Galaxies. *ApJ*, 795(2):145.

Barshan, E., Ghodsi, A., Azimifar, Z., and Jahromi, M. Z. (2011). Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, 44(7):1357–1371.

Bazell, D. and Aha, D. W. (2001). Ensembles of Classifiers for Morphological Galaxy Classification. *ApJ*, 548(1):219–223.

Beck, R., Szapudi, I., Flewelling, H., Holmberg, C., Magnier, E., and Chambers, K. C. (2021). PS1-STRM: neural network source classification and photometric redshift catalogue for PS1 3π DR1. *MNRAS*, 500(2):1633–1644.

Behroozi, P., Wechsler, R. H., Hearin, A. P., and Conroy, C. (2019). UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10. *MNRAS*, 488(3):3143–3194.

Behroozi, P. S., Conroy, C., and Wechsler, R. H. (2010). A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for 0 < z < 4. *ApJ*, 717(1):379–403.

Behroozi, P. S., Marchesini, D., Wechsler, R. H., Muzzin, A., Papovich, C., and Stefanon, M. (2013a). Using Cumulative Number Densities to Compare Galaxies across Cosmic Time. *ApJL*, 777(1):L10.

Behroozi, P. S., Wechsler, R. H., and Conroy, C. (2013b). The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8. *ApJ*, 770(1):57.

Behroozi, P. S., Wechsler, R. H., and Conroy, C. (2013c). The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8. *ApJ*, 770(1):57.

Belli, S., Genzel, R., Förster Schreiber, N. M., Wisnioski, E., Wilman, D. J., Wuyts, S., Mendel, J. T., Beifiori, A., Bender, R., Brammer, G. B., Burkert, A., Chan, J., Davies, R. L., Davies, R., Fabricius, M., Fossati, M., Galametz, A., Lang, P., Lutz, D., Momcheva, I. G., Nelson, E. J., Saglia, R. P., Tacconi, L. J., Tadaki, K.-i., Übler, H., and van Dokkum, P. (2017a). KMOS^{3D} Reveals Low-level Star Formation Activity in Massive Quiescent Galaxies at 0.7 < z < 2.7. *ApJL*, 841(1):L6.

Belli, S., Newman, A. B., and Ellis, R. S. (2017b). MOSFIRE Spectroscopy of Quiescent Galaxies at 1.5 < z < 2.5. I. Evolution of Structural and Dynamical Properties. *ApJ*, 834(1):18.

Belli, S., Newman, A. B., and Ellis, R. S. (2019). MOSFIRE Spectroscopy of Quiescent Galaxies at 1.5 < z < 2.5. II. Star Formation Histories and Galaxy Quenching. *ApJ*, 874(1):17.

Belli, S., Newman, A. B., Ellis, R. S., and Konidaris, N. P. (2014). MOSFIRE Absorption Line Spectroscopy of z > 2 Quiescent Galaxies: Probing a Period of Rapid Size Growth. *ApJL*, 788(2):L29.

Benson, A. J., Bower, R. G., Frenk, C. S., Lacey, C. G., Baugh, C. M., and Cole, S. (2003). What Shapes the Luminosity Function of Galaxies? *ApJ*, 599(1):38–49.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis; 2nd ed.* Springer series in statistics. Springer, New York.

Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., and Nasi, E. (1994). Theoretical isochrones from models with new radiative opacities. *A&AS*, 106:275–302.

Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *A&AS*, 117:393–404.

Best, P. N., Kauffmann, G., Heckman, T. M., Brinchmann, J., Charlot, S., Ivezić, Ž., and White, S. D. M. (2005). The host galaxies of radio-loud active galactic nuclei: mass dependences, gas cooling and active galactic nuclei feedback. *MNRAS*, 362(1):25–40.

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv e-prints*, page arXiv:1701.02434.

Béthermin, M., Doré, O., and Lagache, G. (2012). Where stars form and live at high redshift: clues from the infrared. *A&A*, 537:L5.

Béthermin, M., Wu, H.-Y., Lagache, G., Davidzon, I., Ponthieu, N., Cousin, M., Wang, L., Doré, O., Daddi, E., and Lapi, A. (2017). The impact of clustering and angular resolution on far-infrared and millimeter continuum observations. *A&A*, 607:A89.

Bezanson, R., Spilker, J., Williams, C. C., Whitaker, K. E., Narayanan, D., Weiner, B., and Franx, M. (2019). Extremely Low Molecular Gas Content in a Compact, Quiescent Galaxy at z = 1.522. *ApJL*, 873(2):L19.

Binney, J. (1977). The physics of dissipational galaxy formation. *ApJ*, 215:483–491.

Birnboim, Y. and Dekel, A. (2003). Virial shocks in galactic haloes? *MNRAS*, 345(1):349–364.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* springer.

Bisigello, L., Caputi, K. I., Grogin, N., and Koekemoer, A. (2018). Analysis of the SFR-M* plane at z < 3: single fitting versus multi-Gaussian decomposition. *A&A*, 609:A82.

Blain, A. W., Smail, I., Ivison, R. J., Kneib, J. P., and Frayer, D. T. (2002). Submillimeter galaxies. *PhR*, 369(2):111–176.

Blanton, M. R., Eisenstein, D., Hogg, D. W., Schlegel, D. J., and Brinkmann, J. (2005). Relationship between Environment and the Broadband Optical Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 629(1):143–157.

Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. (1984). Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311(5986):517–525.

Bondi, H. (1952). On Spherically Symmetrical Accretion. *Monthly Notices of the Royal Astronomical Society*, 112(2):195–204.

Bondi, H. and Hoyle, F. (1944). On the Mechanism of Accretion by Stars. *Monthly Notices of the Royal Astronomical Society*, 104(5):273–282.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. page 144–152.

Bouché, N., Dekel, A., Genzel, R., Genel, S., Cresci, G., Förster Schreiber, N. M., Shapiro, K. L., Davies, R. I., and Tacconi, L. (2010). The Impact of Cold Gas Accretion Above a Mass Floor on Galaxy Scaling Relations. *ApJ*, 718(2):1001–1018.

Bouwens, R. J., Bradley, L., Zitrin, A., Coe, D., Franx, M., Zheng, W., Smit, R., Host, O., Postman, M., Moustakas, L., Labbé, I., Carrasco, M., Molino, A., Donahue, M., Kelson, D. D., Meneghetti, M., Benítez, N., Lemze, D., Umetsu, K., Broadhurst, T., Moustakas, J., Rosati, P., Jouvel, S., Bartelmann, M., Ford, H., Graves, G., Grillo, C., Infante, L., Jimenez-Teja, Y., Lahav, O., Maoz, D., Medezinski, E., Melchior, P., Merten, J., Nonino, M., Ogaz, S., and Seitz, S. (2014). A Census of Star-forming Galaxies in the Z ~9-10 Universe based on HST+Spitzer Observations over 19 Clash Clusters: Three Candidate Z ~9-10 Galaxies and Improved Constraints on the Star Formation Rate Density at Z ~9.2. *ApJ*, 795(2):126.

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Labbé, I., Trenti, M., van Dokkum, P., Franx, M., Stiavelli, M., Carollo, C. M., Magee, D., and Gonzalez, V. (2011). Ultraviolet Luminosity Functions from 132 z ~7 and z ~8 Lyman-break Galaxies in the Ultra-deep HUDF09 and Wide-area Early Release Science WFC3/IR Observations. *ApJ*, 737(2):90.

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., and Lacey, C. G. (2006). Breaking the hierarchy of galaxy formation. *MNRAS*, 370(2):645–655.

Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., Marchesini, D., Franx, M., Kriek, M., Labbé, I., Lee, K. S., Muzzin, A., Quadri, R. F., Rudnick, G., and Williams, R. (2011). The Number Density and Mass Density of Star-forming and Quiescent Galaxies at 0.4 <= z <= 2.2. *ApJ*, 739(1):24.

Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Brennan, R., Pandya, V., Somerville, R. S., Barro, G., Taylor, E. N., Wuyts, S., Bell, E. F., Dekel, A., Ferguson, H. C., McIntosh, D. H., Papovich, C., and Primack, J. (2015). Quenching and morphological transformation in semi-analytic models and CANDELS. *MNRAS*, 451(3):2933–2956.

Bruzual, G. and Charlot, S. (2003a). Stellar population synthesis at the resolution of 2003. *MNRAS*, 344(4):1000–1028.

Bruzual, G. and Charlot, S. (2003b). Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028.

Bullock, J. S., Kravtsov, A. V., and Weinberg, D. H. (2000). Reionization and the Abundance of Galactic Satellites. *ApJ*, 539(2):517–521.

Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., and Storchi-Bergmann, T. (2000). The Dust Content and Opacity of Actively Star-forming Galaxies. *ApJ*, 533:682–695.

Capak, P., Carilli, C. L., Lee, N., Aldcroft, T., Aussel, H., Schinnerer, E., Wilson, G. W., Yun, M. S., Blain, A., Giavalisco, M., Ilbert, O., Kartaltepe, J., Lee, K. S., McCracken, H., Mobasher, B., Salvato, M., Sasaki, S., Scott, K. S., Sheth, K., Shioya, Y., Thompson, D., Elvis, M., Sanders, D. B., Scoville, N. Z., and Tanaguchi, Y. (2008). Spectroscopic Confirmation of an Extreme Starburst at Redshift 4.547. *ApJL*, 681(2):L53.

Cappelluti, N., Comastri, A., Fontana, A., Zamorani, G., Amorin, R., Castellano, M., Merlin, E., Santini, P., Elbaz, D., Schreiber, C., Shu, X., Wang, T., Dunlop, J. S., Bourne, N., Bruce, V. A., Buitrago, F., Michałowski, M. J., Derriere, S., Ferguson, H. C., Faber, S. M., and Vito, F. (2016). Chandra Counterparts of CANDELS GOODS-S Sources. *ApJ*, 823(2):95.

Caputi, K. I., Dunlop, J. S., McLure, R. J., Huang, J. S., Fazio, G. G., Ashby, M. L. N., Castellano, M., Fontana, A., Cirasuolo, M., Almaini, O., Bell, E. F., Dickinson, M., Donley, J. L., Faber, S. M., Ferguson, H. C., Giavalisco, M., Grogin, N. A., Kocevski, D. D., Koekemoer, A. M., Koo, D. C., Lai, K., Newman, J. A., and Somerville, R. S. (2012). The Nature of Extremely Red H - [4.5] > 4 Galaxies Revealed with SEDS and CANDELS. *ApJL*, 750(1):L20.

Carliles, S., Budavári, T., Heinis, S., Priebe, C., and Szalay, A. S. (2010). Random Forests for Photometric Redshifts. *ApJ*, 712(1):511–515.

Carnall, A. C., Leja, J., Johnson, B. D., McLure, R. J., Dunlop, J. S., and Conroy, C. (2019a). How to Measure Galaxy Star Formation Histories. I. Parametric Models. *ApJ*, 873(1):44.

Carnall, A. C., McLure, R. J., Dunlop, J. S., Cullen, F., McLeod, D. J., Wild, V., Johnson, B. D., Appleby, S., Davé, R., Amorin, R., Bolzonella, M., Castellano, M., Cimatti, A., Cucciati, O., Marchi, F., Gargiulo, A., Garilli, B., Pentericci, L., Pozzetti, L., Schreiber, C., Talia, M., and Zamorani, G. (2019b). The VANDELS survey: the star-formation histories of massive quiescent galaxies at 1.0 < z < 1.3. *arXiv e-prints*, page arXiv:1903.11082.

Carnall, A. C., McLure, R. J., Dunlop, J. S., and Davé, R. (2018). Inferring the star formation histories of massive quiescent galaxies with BAGPIPES: evidence for multiple quenching mechanisms. *MNRAS*, 480(4):4379–4401.

Carnall, A. C., Walker, S., McLure, R. J., Dunlop, J. S., McLeod, D. J., Cullen, F., Wild, V., Amorin, R., Bolzonella, M., Castellano, M., Cimatti, A., Cucciati, O., Fontana, A., Gargiulo, A., Garilli, B., Jarvis, M. J., Pentericci, L., Pozzetti, L., Zamorani, G., Calabro, A., Hathi, N. P., and Koekemoer, A. M. (2020). Timing the earliest quenching events with a robust sample of massive quiescent galaxies at 2 < z < 5. *arXiv e-prints*, page arXiv:2001.11975.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32.

Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., and Blaizot, J. (2006). Modelling the galaxy bimodality: shutdown above a critical halo mass. *MNRAS*, 370(4):1651–1665.

Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. *PASP*, 115:763–795.

Chartab, N., Mobasher, B., Darvish, B., Finkelstein, S., Guo, Y., Kodra, D., Lee, K.-S., Newman, J. A., Pacifici, C., Papovich, C., Sattari, Z., Shahidi, A., Dickinson, M. E., Faber, S. M., Ferguson, H. C., Giavalisco, M., and Jafariyazani, M. (2020). Large-scale Structures in the CANDELS Fields: The Role of the Environment in Star Formation

Activity. *ApJ*, 890(1):7.

Cimatti, A., Daddi, E., Mignoli, M., Pozzetti, L., Renzini, A., Zamorani, G., Broadhurst, T., Fontana, A., Saracco, P., Poli, F., Cristiani, S., D'Odorico, S., Giallongo, E., Gilmozzi, R., and Menci, N. (2002). The K20 survey. I. Disentangling old and dusty star-forming galaxies in the ERO population. *A&A*, 381:L68–L72.

Cimatti, A., Daddi, E., Renzini, A., Cassata, P., Vanzella, E., Pozzetti, L., Cristiani, S., Fontana, A., Rodighiero, G., Mignoli, M., et al. (2004). Old galaxies in the young universe. *Nature*, 430(6996):184–187.

Civano, F., Marchesi, S., Comastri, A., Urry, M. C., Elvis, M., Cappelluti, N., Puccetti, S., Brusa, M., Zamorani, G., Hasinger, G., Aldcroft, T., Alexand er, D. M., Allevato, V., Brunner, H., Capak, P., Finoguenov, A., Fiore, F., Fruscione, A., Gilli, R., Glotfelty, K., Griffiths, R. E., Hao, H., Harrison, F. A., Jahnke, K., Kartaltepe, J., Karim, A., LaMassa, S. M., Lanzuisi, G., Miyaji, T., Ranalli, P., Salvato, M., Sargent, M., Scoville, N. J., Schawinski, K., Schinnerer, E., Silverman, J., Smolcic, V., Stern, D., Toft, S., Trakhtenbrot, B., Treister, E., and Vignali, C. (2016). The Chandra Cosmos Legacy Survey: Overview and Point Source Catalog. *ApJ*, 819(1):62.

Cole, S. (1991). Modeling Galaxy Formation in Evolving Dark Matter Halos. *ApJ*, 367:45.

Collister, A. A. and Lahav, O. (2004). ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP*, 116(818):345–351.

Conroy, C. and Gunn, J. E. (2010). The Propagation of Uncertainties in Stellar Population Synthesis Modeling. III. Model Calibration, Comparison, and Evaluation. *ApJ*, 712(2):833–857.

Conroy, C., Gunn, J. E., and White, M. (2009). The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies. *ApJ*, 699(1):486–506.

Cook, S. R., Gelman, A., and Rubin, D. B. (2006). Validation of software for bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. pages 273–297.

Crain, R. A., Schaye, J., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., and Trayford, J. W. (2015). The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. *MNRAS*, 450(2):1937–1961.

Cranmer, M. D., Galvez, R., Anderson, L., Spergel, D. N., and Ho, S. (2019). Modeling the Gaia Color-Magnitude Diagram with Bayesian Neural Flows to Constrain Distance Estimates. *arXiv e-prints*, page arXiv:1908.08045.

Cristianini, N., Shawe-Taylor, J., et al. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., and Yoshida, N. (2006). The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365(1):11–28.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314.

Daddi, E., Cimatti, A., Renzini, A., Vernet, J., Conselice, C., Pozzetti, L., Mignoli, M., Tozzi, P., Broadhurst, T., di Serego Alighieri, S., Fontana, A., Nonino, M., Rosati, P., and Zamorani, G. (2004). Near-Infrared Bright Galaxies at z~=2. Entering the Spheroid Formation Epoch? *ApJL*, 600(2):L127–L130.

Daddi, E., Dickinson, M., Morrison, G., Chary, R., Cimatti, A., Elbaz, D., Frayer, D., Renzini, A., Pope, A., Alexander, D. M., Bauer, F. E., Giavalisco, M., Huynh, M., Kurk, J., and Mignoli, M. (2007). Multiwavelength Study of Massive Galaxies at z~2. I. Star Formation and Galaxy Growth. *ApJ*, 670(1):156–172.

Daddi, E., Renzini, A., Pirzkal, N., Cimatti, A., Malhotra, S., Stiavelli, M., Xu, C., Pasquali, A., Rhoads, J. E., Brusa, M., di Serego Alighieri, S., Ferguson, H. C., Koekemoer, A. M., Moustakas, L. A., Panagia, N., and Windhorst, R. A. (2005). Passively Evolving Early-Type Galaxies at 1.4 <~z <~2.5 in the Hubble Ultra Deep Field. *ApJ*, 626(2):680–697.

Dahlen, T., Mobasher, B., Faber, S. M., Ferguson, H. C., Barro, G., Finkelstein, S. L., Finlator, K., Fontana, A., Gruetzbauch, R., Johnson, S., Pforr, J., Salvato, M., Wiklind, T., Wuyts, S., Acquaviva, V., Dickinson, M. E., Guo, Y., Huang, J., Huang, K.-H., Newman, J. A., Bell, E. F., Conselice, C. J., Galametz, A., Gawiser, E., Giavalisco, M., Grogin, N. A., Hathi, N., Kocevski, D., Koekemoer, A. M., Koo, D. C., Lee, K.-S., McGrath, E. J., Papovich, C., Peth, M., Ryan, R., Somerville, R., Weiner, B., and Wilson, G. (2013). A Critical Assessment of Photometric Redshift Methods: A CANDELS Investigation. *ApJ*, 775(2):93.

Damjanov, I., Abraham, R. G., Glazebrook, K., McCarthy, P. J., Caris, E., Carlberg, R. G., Chen, H.-W., Crampton, D., Green, A. W., Jørgensen, I., Juneau, S., Le Borgne, D., Marzke, R. O., Mentuch, E., Murowinski, R., Roth, K., Savaglio, S., and Yan, H. (2011). Red Nuggets at High Redshift: Structural Evolution of Quiescent Galaxies Over 10 Gyr of Cosmic History. *ApJL*, 739(2):L44.

Darvish, B., Martin, C., Gonçalves, T. S., Mobasher, B., Scoville, N. Z., and Sobral, D. (2018). Quenching or Bursting: The Role of Stellar Mass, Environment, and Specific Star Formation Rate to z1. *ApJ*, 853(2):155.

Darvish, B., Mobasher, B., Sobral, D., Rettura, A., Scoville, N., Faisst, A., and Capak, P. (2016). The Effects of the Local Environment and Stellar Mass on Galaxy Quenching to z $\sim$ 3. *ApJ*, 825(2):113.

Darvish, B., Sobral, D., Mobasher, B., Scoville, N. Z., Best, P., Sales, L. V., and Smail, I. (2014). Cosmic Web and Star Formation Activity in Galaxies at z ~1. *ApJ*, 796(1):51.

Davé, R., Finlator, K., and Oppenheimer, B. D. (2012). An analytic model for the evolution of the stellar, gas and metal content of galaxies. *MNRAS*, 421(1):98–107.

Davé, R., Oppenheimer, B. D., and Finlator, K. (2011). Galaxy evolution in cosmological simulations with outflows - I. Stellar masses and star formation rates. *MNRAS*, 415(1):11–31.

Davidzon, I., Laigle, C., Capak, P. L., Ilbert, O., Masters, D. C., Hemmati, S., Apostolakos, N., Coupon, J., de la Torre, S., Devriendt, J., Dubois, Y., Kashino, D., Paltani, S., and Pichon, C. (2019). HORIZON-AGN virtual observatory - 2. Template-free estimates of galaxy properties from colours. *MNRAS*, 489(4):4817–4835.

Dekel, A. and Birnboim, Y. (2006). Galaxy bimodality due to cold flows and shock heating. *MNRAS*, 368(1):2–20.

Dekel, A., Birnboim, Y., Engel, G., Freundlich, J., Goerdt, T., Mumcuoglu, M., Neistein, E., Pichon, C., Teyssier, R., and Zinger, E. (2009). Cold streams in early massive hot haloes as the main mode of galaxy formation. *Nature*, 457(7228):451–454.

Dekel, A., Lapiner, S., and Dubois, Y. (2019). Origin of the Golden Mass of Galaxies and Black Holes. *arXiv e-prints*, page arXiv:1904.08431.

Dekel, A. and Mandelker, N. (2014). An analytic solution for the minimal bathtub toy model: challenges in the star formation history of high-z galaxies. *MNRAS*, 444(3):2071–2084.

Dekel, A. and Silk, J. (1986). The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation. *ApJ*, 303:39.

Dekel, A. and Woo, J. (2003). Feedback and the fundamental line of low-luminosity low-surface-brightness/dwarf galaxies. *MNRAS*, 344(4):1131–1144.

D'Eugenio, C., Daddi, E., Gobat, R., Strazzullo, V., Lustig, P., Delvecchio, I., Jin, S., Cimatti, A., and Onodera, M. (2021). HST grism spectroscopy of z $\sim$ 3 massive quiescent galaxies. Approaching the metamorphosis. *A&A*, 653:A32.

Diamond, S. and Boyd, S. (2016). CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5.

Djorgovski, S. G., Mahabal, A. A., Donalek, C., Graham, M. J., Drake, A. J., Moghaddam, B., and Turmon, M. (2012). Flashes in a Star Stream: Automated Classification of Astronomical Transient Events. *arXiv e-prints*, page arXiv:1209.1681.

Do, C. B. and Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature Biotechnology*, 26(8):897–899.

Donnari, M., Pillepich, A., Nelson, D., Vogelsberger, M., Genel, S., Weinberger, R., Marinacci, F., Springel, V., and Hernquist, L. (2019). The star formation activity of IllustrisTNG galaxies: main sequence, UVJ diagram, quenched fractions, and systematics. *MNRAS*, 485(4):4817–4840.

Dressler, A. (1980). Galaxy morphology in rich clusters: implications for the formation and evolution of galaxies. *ApJ*, 236:351–365.

Drineas, P. and Mahoney, M. W. (2005). On the nyström method for approximating a gram matrix for improved kernel-based learning. *journal of machine learning research*, 6(Dec):2153–2175.

Duda, R. and Hart, P. (1973). *Pattern Classification and Scene Analysis.* Wiley.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

Eddington, A. S. (1913). On a Formula for Correcting Statistics for the Effects of a known Probable Error of Observation. *Monthly Notices of the Royal Astronomical Society*, 73(5):359–360.

Efstathiou, G. (1992). Suppressing the formation of dwarf galaxies via photoionization. *MNRAS*, 256(2):43P–47P.

Efstathiou, G., Ellis, R. S., and Peterson, B. A. (1988). Analysis of a complete galaxy redshift survey. II. The field-galaxy luminosity function. *MNRAS*, 232:431–461.

Elbaz, D., Daddi, E., Le Borgne, D., Dickinson, M., Alexander, D. M., Chary, R. R., Starck, J. L., Brandt, W. N., Kitzbichler, M., MacDonald, E., Nonino, M., Popesso, P., Stern, D., and Vanzella, E. (2007). The reversal of the star formation-density relation in the distant universe. *A&A*, 468(1):33–48.

Estrada-Carpenter, V., Papovich, C., Momcheva, I., Brammer, G., Long, J., Quadri, R. F., Bridge, J., Dickinson, M., Ferguson, H., Finkelstein, S., Giavalisco, M., Gosmeyer, C. M., Lotz, J., Salmon, B., Skelton, R. E., Trump, J. R., and Weiner, B. (2019). CLEAR. I. Ages and Metallicities of Quiescent Galaxies at 1.0 < z < 1.8 Derived from Deep Hubble Space Telescope Grism Data. *ApJ*, 870(2):133.

Evans, I. N., Primini, F. A., Glotfelty, K. J., Anderson, C. S., Bonaventura, N. R., Chen, J. C., Davis, J. E., Doe, S. M., Evans, J. D., Fabbiano, G., Galle, E. C., Gibbs, Danny G., I., Grier, J. D., Hain, R. M., Hall, D. M., Harbo, P. N., He, X. H., Houck, J. C., Karovska, M., Kashyap, V. L., Lauer, J., McCollough, M. L., McDowell, J. C., Miller, J. B., Mitschang, A. W., Morgan, D. L., Mossman, A. E., Nichols, J. S., Nowak, M. A., Plummer, D. A., Refsdal, B. L., Rots, A. H., Siemiginowska, A., Sundheim, B. A., Tibbetts, M. S., Van Stone, D. W., Winkelman, S. L., and Zografou, P. (2010). The Chandra Source Catalog. *ApJS*, 189(1):37–82.

Fabian, A. C. (2012). Observational Evidence of Active Galactic Nuclei Feedback. *ARA&A*, 50:455–489.

Fadely, R., Hogg, D. W., and Willman, B. (2012). Star-Galaxy Classification in Multi-band Optical Imaging. *ApJ*, 760(1):15.

Falkendal, T., De Breuck, C., Lehnert, M. D., Drouart, G., Vernet, J., Emonts, B., Lee, M., Nesvadba, N. P. H., Seymour, N., Béthermin, M., Kolwa, S., Gullberg, B., and Wylezalek, D. (2019). Massive galaxies on the road to quenching: ALMA observations of powerful high redshift radio galaxies. *A&A*, 621:A27.

Fall, S. M. and Efstathiou, G. (1980). Formation and rotation of disc galaxies with haloes. *MNRAS*, 193:189–206.

Fang, J. J., Faber, S. M., Koo, D. C., Rodríguez-Puebla, A., Guo, Y., Barro, G., Behroozi, P., Brammer, G., Chen, Z., Dekel, A., Ferguson, H. C., Gawiser, E., Giavalisco, M., Kartaltepe, J., Kocevski, D. D., Koekemoer, A. M., McGrath, E. J., McIntosh, D., Newman, J. A., Pacifici, C., Pandya, V., Pérez-González, P. G., Primack, J. R., Salmon, B., Trump, J. R., Weiner, B., Willner, S. P., Acquaviva, V., Dahlen, T., Finkelstein, S. L., Finlator, K., Fontana, A., Galametz, A., Grogin, N. A., Gruetzbauch, R., Johnson, S., Mobasher, B., Papovich, C. J., Pforr, J., Salvato, M., Santini, P., van der Wel, A., Wiklind, T., and Wuyts, S. (2018). Demographics of Star-forming Galaxies since z 2.5. I. The UVJ Diagram in CANDELS. *ApJ*, 858(2):100.

Feldmann, R. and Mayer, L. (2015). The Argo simulation - I. Quenching of massive galaxies at high redshift as a result of cosmological starvation. *MNRAS*, 446(2):1939–1956.

Feroz, F. and Hobson, M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses.

MNRAS, 384(2):449–463.

Feroz, F., Hobson, M. P., and Bridges, M. (2009). MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4):1601–1614.

Firth, A. E., Lahav, O., and Somerville, R. S. (2003). Estimating photometric redshifts with artificial neural networks. *MNRAS*, 339(4):1195–1202.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188.

Fontana, A., Santini, P., Grazian, A., Pentericci, L., Fiore, F., Castellano, M., Giallongo, E., Menci, N., Salimbeni, S., Cristiani, S., Nonino, M., and Vanzella, E. (2009). The fraction of quiescent massive galaxies in the early Universe. *A&A*, 501(1):15–20.

Forrest, B., Annunziatella, M., Wilson, G., Marchesini, D., Muzzin, A., Cooper, M. C., Marsan, Z. C., McConachie, I., Chan, J. C. C., Gomez, P., Kado-Fong, E., Barbera, F. L., Labbé, I., Lange-Vagle, D., Nantais, J., Nonino, M., Peña, T., Saracco, P., Stefanon, M., and van der Burg, R. F. J. (2020a). An Extremely Massive Quiescent Galaxy at z = 3.493: Evidence of Insufficiently Rapid Quenching Mechanisms in Theoretical Models. *ApJL*, 890(1):L1.

Forrest, B., Annunziatella, M., Wilson, G., Marchesini, D., Muzzin, A., Cooper, M. C., Marsan, Z. C., McConachie, I., Chan, J. C. C., Gomez, P., Kado-Fong, E., L Barbera, F., Labbé, I., Lange-Vagle, D., Nantais, J., Nonino, M., Peña, T., Saracco, P., Stefanon, M., and van der Burg, R. F. J. (2020b). An Extremely Massive Quiescent Galaxy at z = 3.493: Evidence of Insufficiently Rapid Quenching Mechanisms in Theoretical Models. *ApJL*, 890(1):L1.

Forrest, B., Marsan, Z. C., Annunziatella, M., Wilson, G., Muzzin, A., Marchesini, D., Cooper, M. C., Chan, J. C. C., McConachie, I., Gomez, P., Kado-Fong, E., Barbera, F. L., Lange-Vagle, D., Nantais, J., Nonino, M., Saracco, P., Stefanon, M., and van der Burg, R. F. J. (2020c). The Massive Ancient Galaxies at z > 3 NEar-infrared (MAGAZ3NE) Survey: Confirmation of Extremely Rapid Star Formation and Quenching Timescales for Massive Galaxies in the Early Universe. *ApJ*, 903(1):47.

Förster Schreiber, N. M., Übler, H., Davies, R. L., Genzel, R., Wisnioski, E., Belli, S., Shimizu, T., Lutz, D., Fossati, M., Herrera-Camus, R., Mendel, J. T., Tacconi, L. J., Wilman, D., Beifiori, A., Brammer, G. B., Burkert, A., Carollo, C. M., Davies, R. I., Eisenhauer, F., Fabricius, M., Lilly, S. J., Momcheva, I., Naab, T., Nelson, E. J., Price, S. H., Renzini, A., Saglia, R., Sternberg, A., van Dokkum, P., and Wuyts, S. (2019). The KMOS^{3D} Survey: Demographics and Properties of Galactic Outflows at z = 0.6-2.7. *ApJ*, 875(1):21.

Franx, M., Labbé, I., Rudnick, G., van Dokkum, P. G., Daddi, E., Förster Schreiber, N. M., Moorwood, A., Rix, H.-W., Röttgering, H., van der Wel, A., van der Werf, P., and van Starkenburg, L. (2003). A Significant Population of Red, Near-Infrared-selected High-Redshift Galaxies. *ApJL*, 587(2):L79–L82.

Frenk, C. S. and White, S. D. M. (2012). Dark matter and cosmic structure. *Annalen der Physik*, 524(9-10):507–534.

Fumagalli, M., Franx, M., van Dokkum, P., Whitaker, K. E., Skelton, R. E., Brammer, G., Nelson, E., Maseda, M., Momcheva, I., Kriek, M., Labbé, I., Lundgren, B., and Rix, H.-W. (2016). Ages of Massive Galaxies at 0.5 > z > 2.0 from 3D-HST Rest-frame Optical Spectroscopy. *ApJ*, 822(1):1.

Fumagalli, M., Labbé, I., Patel, S. G., Franx, M., van Dokkum, P., Brammer, G., da Cunha, E., Förster Schreiber, N. M., Kriek, M., Quadri, R., Rix, H.-W., Wake, D., Whitaker, K. E., Lundgren, B., Marchesini, D., Maseda, M., Momcheva, I., Nelson, E., Pacifici, C., and Skelton, R. E. (2014). How Dead are Dead Galaxies? Mid-infrared Fluxes of Quiescent Galaxies at Redshift $0.3 < z < 2.5$: Implications for Star Formation Rates and Dust Heating. *ApJ*, 796(1):35.

Galametz, M., Kennicutt, R. C., Calzetti, D., Aniano, G., Draine, B. T., Boquien, M., Brandl, B., Croxall, K. V., Dale, D. A., Engelbracht, C. W., Gordon, K. D., Groves, B., Hao, C.-N., Helou, G., Hinz, J. L., Hunt, L. K., Johnson, B. D., Li, Y., Murphy, E., Roussel, H., Sandstrom, K., Skibba, R. A., and Tabatabaei, F. S. (2013). Calibration of the total infrared luminosity of nearby galaxies from Spitzer and Herschel bands. *MNRAS*, 431:1956–1986.

Gelman, A. (2004). Parameterization and bayesian modeling. *Journal of the American Statistical Association*, 99(466):537–545.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2(4):1360 – 1383.

Gelman, A. and Raghunathan, T. (2001). Using conditional distributions for missing-data imputation. *Statistical Science*, 15.

Ghosh, J., Li, Y., and Mitra, R. (2018). On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression. *Bayesian Analysis*, 13(2):359 – 383.

Girelli, G., Bolzonella, M., and Cimatti, A. (2019). Massive and old quiescent galaxies at high redshift. *arXiv e-prints*, page arXiv:1910.07544.

Glazebrook, K., Schreiber, C., Labbé, I., Nanayakkara, T., Kacprzak, G. G., Oesch, P. A., Papovich, C., Spitler, L. R., Straatman, C. M. S., Tran, K.-V. H., and Yuan, T. (2017). A massive, quiescent galaxy at a redshift of 3.717. *Nature*, 544(7648):71–74.

Gnedin, N. Y. and Kravtsov, A. V. (2006). Fossils of Reionization in the Local Group. *ApJ*, 645(2):1054–1061.

Gobat, R., Strazzullo, V., Daddi, E., Onodera, M., Renzini, A., Béthermin, M., Dickinson, M., Carollo, M., and Cimatti, A. (2012). The Early Early Type: Discovery of a Passive Galaxy at z $_{spec}$ ~3. *ApJL*, 759(2):L44.

Gott, J. R., I. and Rees, M. J. (1975). A theory of galaxy formation and clustering. *A&A*, 45:365–376.

Granato, G. L., De Zotti, G., Silva, L., Bressan, A., and Danese, L. (2004). A Physical Model for the Coevolution of QSOs and Their Spheroidal Hosts. *ApJ*, 600(2):580–594.

Grazian, A., Salimbeni, S., Pentericci, L., Fontana, A., Nonino, M., Vanzella, E., Cristiani, S., de Santis, C., Gallozzi, S., Giallongo, E., and Santini, P. (2007). A comparison of LBGs, DRGs, and BzK galaxies: their contribution to the stellar mass density in the GOODS-MUSIC sample. *A&A*, 465(2):393–404.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005). Measuring statistical dependence with hilbert-schmidt norms. pages 63–77.

Grogin, N. A., Kocevski, D. D., Faber, S. M., Ferguson, H. C., Koekemoer, A. M., Riess, A. G., Acquaviva, V., Alexander, D. M., Almaini, O., Ashby, M. L. N., Barden, M., Bell, E. F., Bournaud, F., Brown, T. M., Caputi, K. I., Casertano, S., Cassata, P., Castellano, M., Challis, P., Chary, R.-R., Cheung, E., Cirasuolo, M., Conselice, C. J., Roshan Cooray, A., Croton, D. J., Daddi, E., Dahlen, T., Davé, R., de Mello, D. F., Dekel, A., Dickinson, M., Dolch, T., Donley, J. L., Dunlop, J. S., Dutton, A. A., Elbaz, D., Fazio, G. G., Filippenko, A. V., Finkelstein, S. L., Fontana, A., Gardner, J. P., Garnavich, P. M., Gawiser, E., Giavalisco, M., Grazian, A., Guo, Y., Hathi, N. P., Häussler, B., Hopkins, P. F., Huang, J.-S., Huang, K.-H., Jha, S. W., Kartaltepe, J. S., Kirshner, R. P., Koo, D. C., Lai, K., Lee, K.-S., Li, W., Lotz, J. M., Lucas, R. A., Madau, P., McCarthy, P. J., McGrath, E. J., McIntosh, D. H., McLure, R. J., Mobasher, B., Moustakas, L. A., Mozena, M., Nandra, K., Newman, J. A., Niemi, S.-M., Noeske, K. G., Papovich, C. J., Pentericci, L., Pope, A., Primack, J. R., Rajan, A., Ravindranath, S., Reddy, N. A., Renzini, A., Rix, H.-W., Robaina, A. R., Rodney, S. A., Rosario, D. J., Rosati, P., Salimbeni, S., Scarlata, C., Siana, B., Simard, L., Smidt, J., Somerville, R. S., Spinrad, H., Straughn, A. N., Strolger, L.-G., Telford, O., Teplitz, H. I., Trump, J. R., van der Wel, A., Villforth, C., Wechsler, R. H., Weiner, B. J., Wiklind, T., Wild, V., Wilson, G., Wuyts, S., Yan, H.-J., and Yun, M. S. (2011). CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey. *ApJS*, 197(2):35.

Guo, Q., White, S., Li, C., and Boylan-Kolchin, M. (2010). How do galaxies populate dark matter haloes? *MNRAS*, 404(3):1111–1120.

Guo, Y., Ferguson, H. C., Giavalisco, M., Barro, G., Willner, S. P., Ashby, M. L. N., Dahlen, T., Donley, J. L., Faber, S. M., Fontana, A., Galametz, A., Grazian, A., Huang, K.-H., Kocevski, D. D., Koekemoer, A. M., Koo, D. C., McGrath, E. J., Peth, M., Salvato, M., Wuyts, S., Castellano, M., Cooray, A. R., Dickinson, M. E., Dunlop, J. S., Fazio, G. G., Gardner, J. P., Gawiser, E., Grogin, N. A., Hathi, N. P., Hsu, L.-T., Lee, K.-S., Lucas, R. A., Mobasher, B., Nandra, K., Newman, J. A., and van der Wel, A. (2013). CANDELS Multi-wavelength Catalogs: Source Detection and Photometry in the GOODS-South Field. *ApJS*, 207:24.

Haehnelt, M. G., Natarajan, P., and Rees, M. J. (1998). High-redshift galaxies, their active nuclei and central black holes. *MNRAS*, 300(3):817–827.

Hahn, C., Starkenburg, T. K., Choi, E., Davé, R., Dickey, C. M., Geha, M. C., Genel, S., Hayward, C. C., Maller, A. H., Mandyam, N., Pandya, V., Popping, G., Rafieferantsoa, M., Somerville, R. S., and Tinker, J. L. (2019). IQ-Collaboratory 1.1: The Star-forming Sequence of Simulated Central Galaxies. *ApJ*, 872(2):160.

Haines, C. P., Pereira, M. J., Smith, G. P., Egami, E., Sanderson, A. J. R., Babul, A., Finoguenov, A., Merluzzi, P., Busarello, G., Rawle, T. D., and Okabe, N. (2013). LoCuSS: The Steady Decline and Slow Quenching of Star Formation in Cluster Galaxies over the Last Four Billion Years. *ApJ*, 775(2):126.

Harrison, C. M., Costa, T., Tadhunter, C. N., Flütsch, A., Kakkad, D., Perna, M., and Vietri, G. (2018). AGN outflows and feedback twenty years on. *Nature Astronomy*, 2:198–205.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer series in statistics. Springer.

Hemmati, S., Capak, P., Masters, D., Davidzon, I., Dorè, O., Kruk, J., Mobasher, B., Rhodes, J., Scolnic, D., and Stern, D. (2019a). Photometric Redshift Calibration Requirements for WFIRST Weak-lensing Cosmology: Predictions from CANDELS. *ApJ*, 877(2):117.

Hemmati, S., Capak, P., Pourrahmani, M., Nayyeri, H., Stern, D., Mobasher, B., Darvish, B., Davidzon, I., Ilbert, O., Masters, D., and Shahidi, A. (2019b). Bringing Manifold Learning and Dimensionality Reduction to SED Fitters. *ApJL*, 881(1):L14.

Hemmati, S., Capak, P., Pourrahmani, M., Nayyeri, H., Stern, D., Mobasher, B., Darvish, B., Davidzon, I., Ilbert, O., Masters, D., and Shahidi, A. (2019c). Bringing Manifold Learning and Dimensionality Reduction to SED Fitters. *ApJL*, 881(1):L14.

Henrion, M., Mortlock, D. J., Hand, D. J., and Gandy, A. (2011). A Bayesian approach to star-galaxy classification. *MNRAS*, 412(4):2286–2302.

Hinton, G. E. and Roweis, S. T. (2003). Stochastic neighbor embedding. pages 857–864.

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171–1220.

Hogg, D. W., Eilers, A.-C., and Rix, H.-W. (2019). Spectrophotometric Parallaxes with Linear Models: Accurate Distances for Luminous Red-giant Stars. *AJ*, 158(4):147.

Hopkins, P. F., Cox, T. J., Younger, J. D., and Hernquist, L. (2009a). How do Disks Survive Mergers? *ApJ*, 691(2):1168–1201.

Hopkins, P. F., Kereš, D., Oñorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., and Bullock, J. S. (2014). Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation. *MNRAS*, 445(1):581–603.

Hopkins, P. F., Somerville, R. S., Cox, T. J., Hernquist, L., Jogee, S., Kereš, D., Ma, C.-P., Robertson, B., and Stewart, K. (2009b). The effects of gas on morphological transformation in mergers: implications for bulge and disc demographics. *MNRAS*, 397(2):802–814.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257.

Hornik, K., Stinchcombe, M., and White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5):551 – 560.

Hoyle, F. and Lyttleton, R. A. (1939). The effect of interstellar matter on climatic variation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(3):405–415.

Ilbert, O., Arnouts, S., McCracken, H. J., Bolzonella, M., Bertin, E., Le Fèvre, O., Mellier, Y., Zamorani, G., Pellò, R., Iovino, A., Tresse, L., Le Brun, V., Bottini, D., Garilli, B., Maccagni, D., Picat, J. P., Scaramella, R., Scodeggio, M., Vettolani, G., Zanichelli, A., Adami, C., Bardelli, S., Cappi, A., Charlot, S., Ciliegi, P., Contini, T., Cucciati, O., Foucaud, S., Franzetti, P., Gavignaud, I., Guzzo, L., Marano, B., Marinoni, C., Mazure, A., Meneux, B., Merighi, R., Paltani, S., Pollo, A., Pozzetti, L., Radovich, M., Zucca, E., Bondi, M., Bongiorno, A., Busarello, G., de La Torre, S., Gregorini, L., Lamareille, F., Mathez, G., Merluzzi, P., Ripepi, V., Rizzo, D., and Vergani, D. (2006). Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *A&A*, 457(3):841–856.

Ilbert, O., Capak, P., Salvato, M., Aussel, H., McCracken, H. J., Sanders, D. B., Scoville, N., Kartaltepe, J., Arnouts, S., Le Floc'h, E., Mobasher, B., Taniguchi, Y., Lamareille, F., Leauthaud, A., Sasaki, S., Thompson, D., Zamojski, M., Zamorani, G., Bardelli, S., Bolzonella, M., Bongiorno, A., Brusa, M., Caputi, K. I., Carollo, C. M., Contini, T., Cook, R., Coppa, G., Cucciati, O., de la Torre, S., de Ravel, L., Franzetti, P., Garilli, B., Hasinger, G., Iovino, A., Kampczyk, P., Kneib, J. P., Knobel, C., Kovac, K., Le Borgne, J. F., Le Brun, V., Le Fèvre, O., Lilly, S., Looper, D., Maier, C., Mainieri, V., Mellier, Y., Mignoli, M., Murayama, T., Pellò, R., Peng, Y., Pérez-Montero, E., Renzini, A., Ricciardelli, E., Schiminovich, D., Scodeggio, M., Shioya, Y., Silverman, J., Surace, J., Tanaka, M., Tasca, L., Tresse, L., Vergani, D., and Zucca, E. (2009). Cosmos Photometric Redshifts with 30-Bands for 2-deg^2. *ApJ*, 690(2):1236–1249.

Inoue, A. K., Shimizu, I., Iwata, I., and Tanaka, M. (2014). An updated analytic model for attenuation by the intergalactic medium. *MNRAS*, 442(2):1805–1820.

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., Auza, N., Axelrod, T. S., Bard, D. J., Barr, J. D., Barrau, A., Bartlett, J. G., Bauer, A. E., Bauman, B. J., Baumont, S., Bechtol, E., Bechtol, K., Becker, A. C., Becla, J., Beldica, C., Bellavia, S., Bianco, F. B., Biswas, R., Blanc, G., Blazek, J., Blandford, R. D., Bloom, J. S., Bogart, J., Bond, T. W., Booth, M. T., Borgland, A. W., Borne, K., Bosch, J. F., Boutigny, D., Brackett, C. A., Bradshaw, A., Brandt, W. N., Brown, M. E., Bullock, J. S., Burchat, P., Burke, D. L., Cagnoli, G., Calabrese, D., Callahan, S., Callen, A. L., Carlin, J. L., Carlson, E. L., Chandrasekharan, S., Charles-Emerson, G., Chesley, S., Cheu, E. C., Chiang, H.-F., Chiang, J., Chirino, C., Chow, D., Ciardi, D. R., Claver, C. F., Cohen-Tanugi, J., Cockrum, J. J., Coles, R., Connolly, A. J., Cook, K. H., Cooray, A., Covey, K. R., Cribbs, C., Cui, W., Cutri, R., Daly, P. N., Daniel, S. F., Daruich, F., Daubard, G., Daues, G., Dawson, W., Delgado, F., Dellapenna, A., de Peyster, R., de Val-Borro, M., Digel, S. W., Doherty, P., Dubois, R., Dubois-Felsmann, G. P., Durech, J., Economou, F., Eifler, T., Eracleous, M., Emmons, B. L., Fausti Neto, A., Ferguson, H., Figueroa, E., Fisher-Levine, M., Focke, W., Foss, M. D., Frank, J., Freemon, M. D., Gangler, E., Gawiser, E., Geary, J. C., Gee, P., Geha, M., Gessner, C. J. B., Gibson, R. R., Gilmore, D. K., Glanzman, T., Glick, W., Goldina, T., Goldstein, D. A., Goodenow, I., Graham, M. L., Gressler, W. J., Gris, P., Guy, L. P., Guyonnet, A., Haller, G., Harris, R., Hascall, P. A., Haupt, J., Hernandez, F., Herrmann, S., Hileman, E., Hoblitt, J., Hodgson, J. A., Hogan, C., Howard, J. D., Huang, D., Huffer, M. E., Ingraham, P., Innes, W. R., Jacoby, S. H., Jain, B., Jammes, F., Jee, M. J., Jenness, T., Jernigan, G., Jevremović, D., Johns, K., Johnson, A. S., Johnson, M. W. G., Jones, R. L., Juramy-Gilles, C., Jurić, M., Kalirai, J. S., Kallivayalil, N. J., Kalmbach, B., Kantor, J. P., Karst, P., Kasliwal, M. M., Kelly, H., Kessler, R., Kinnison, V., Kirkby, D., Knox, L., Kotov, I. V., Krabbendam, V. L., Krughoff, K. S., Kubánek, P., Kuczewski, J., Kulkarni, S., Ku, J., Kurita, N. R., Lage, C. S., Lambert, R., Lange, T., Langton, J. B., Le Guillou, L., Levine, D., Liang, M., Lim, K.-T., Lintott, C. J., Long, K. E., Lopez, M., Lotz, P. J., Lupton, R. H., Lust, N. B., MacArthur, L. A., Mahabal, A., Mandelbaum, R., Markiewicz, T. W., Marsh, D. S., Marshall, P. J., Marshall, S., May, M., McKercher, R., McQueen, M., Meyers, J., Migliore, M., Miller, M., Mills, D. J., Miraval, C., Moeyens, J., Moolekamp, F. E., Monet, D. G., Moniez, M., Monkewitz, S., Montgomery, C., Morrison, C. B., Mueller, F., Muller, G. P., Muñoz Arancibia, F., Neill, D. R., Newbry, S. P., Nief, J.-Y., Nomerotski, A., Nordby, M., O'Connor, P., Oliver, J., Olivier, S. S., Olsen, K., O'Mullane, W., Ortiz, S., Osier, S., Owen, R. E., Pain, R., Palecek, P. E., Parejko, J. K., Parsons, J. B., Pease, N. M., Peterson, J. M., Peterson, J. R., Petravick, D. L., Libby Petrick, M. E., Petry, C. E., Pierfederici, F., Pietrowicz, S., Pike, R., Pinto, P. A., Plante, R., Plate, S., Plutchak, J. P., Price, P. A., Prouza, M., Radeka, V., Rajagopal, J., Rasmussen, A. P., Regnault, N., Reil, K. A., Reiss, D. J., Reuter, M. A., Ridgway, S. T., Riot, V. J., Ritz, S., Robinson, S., Roby, W., Roodman, A., Rosing, W., Roucelle, C., Rumore, M. R., Russo, S., Saha, A., Sassolas, B., Schalk, T. L., Schellart, P., Schindler, R. H., Schmidt, S., Schneider, D. P., Schneider, M. D., Schoening, W., Schumacher, G., Schwamb, M. E., Sebag, J., Selvy, B., Sembroski, G. H., Seppala, L. G., Serio, A., Serrano, E., Shaw, R. A., Shipsey, I., Sick, J., Silvestri, N., Slater, C. T., Smith, J. A., Smith, R. C., Sobhani, S., Soldahl, C., Storrie-Lombardi, L., Stover, E., Strauss, M. A., Street, R. A., Stubbs, C. W., Sullivan, I. S., Sweeney, D., Swinbank, J. D., Szalay, A., Takacs, P., Tether, S. A., Thaler, J. J., Thayer, J. G., Thomas, S., Thornton, A. J., Thukral, V., Tice, J., Trilling, D. E., Turri, M., Van Berg, R., Vanden Berk, D., Vetter, K., Virieux, F., Vucina, T., Wahl, W., Walkowicz, L., Walsh, B., Walter, C. W., Wang, D. L., Wang, S.-Y., Warner, M., Wiecha, O., Willman, B., Winters, S. E., Wittman, D., Wolff, S. C., Wood-Vasey, W. M., Wu, X., Xin, B., Yoachim, P., and Zhan, H. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111.

Iyer, K. and Gawiser, E. (2017). Reconstruction of Galaxy Star Formation Histories through SED Fitting:The Dense Basis Approach. *ApJ*, 838(2):127.

Iyer, K. G., Gawiser, E., Faber, S. M., Ferguson, H. C., Kartaltepe, J., Koekemoer, A. M., Pacifici, C., and Somerville, R. S. (2019). Nonparametric Star Formation History Reconstruction with Gaussian Processes. I. Counting Major Episodes of Star Formation. *ApJ*, 879(2):116.

Jeffreys, H. (1938). The correction of frequencies for a known standard error of observation. *MNRAS*, 98:190.

Jeffreys, H. (1961). *Theory of Probability.* Oxford, Oxford, England, third edition.

Johnson, B. D., Leja, J., Conroy, C., and Speagle, J. S. (2021). Stellar Population Inference with Prospector. *ApJS*, 254(2):22.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules.

Journal of the American Statistical Association, 91(435):1343–1370.

Katz, N., Hernquist, L., and Weinberg, D. H. (1992). Galaxies and Gas in a Cold Dark Matter Universe. *ApJL*, 399:L109.

Kauffmann, G., White, S. D. M., and Guiderdoni, B. (1993). The formation and evolution of galaxies within merging dark matter haloes. *MNRAS*, 264:201–218.

Kauffmann, G., White, S. D. M., Heckman, T. M., Ménard, B., Brinchmann, J., Charlot, S., Tremonti, C., and Brinkmann, J. (2004). The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. *MNRAS*, 353(3):713–731.

Kawinwanichakij, L., Papovich, C., Quadri, R. F., Glazebrook, K., Kacprzak, G. G., Allen, R. J., Bell, E. F., Croton, D. J., Dekel, A., Ferguson, H. C., Forrest, B., Grogin, N. A., Guo, Y., Kocevski, D. D., Koekemoer, A. M., Labbé, I., Lucas, R. A., Nanayakkara, T., Spitler, L. R., Straatman, C. M. S., Tran, K.-V. H., Tomczak, A., and van Dokkum, P. (2017). Effect of Local Environment and Stellar Mass on Galaxy Quenching and Morphology at $0.5 < z < 2.0$. *ApJ*, 847(2):134.

Kelly, B. C., Fan, X., and Vestergaard, M. (2008). A Flexible Method of Estimating Luminosity Functions. *ApJ*, 682(2):874–895.

Kennicutt, Robert C., J. (1989). The Star Formation Law in Galactic Disks. *ApJ*, 344:685.

Kennicutt, Jr., R. C. (1998). Star Formation in Galaxies Along the Hubble Sequence. *ARA&A*, 36:189–232.

Kereš, D., Katz, N., Weinberg, D. H., and Davé, R. (2005). How do galaxies get their gas? *MNRAS*, 363(1):2–28.

Kim, E. J. and Brunner, R. J. (2017). Star-galaxy classification using deep convolutional neural networks. *MNRAS*, 464(4):4463–4475.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kiviluoto, K. (1996). Topology preservation in self-organizing maps. In *Proceedings of International Conference on Neural Networks (ICNN'96)*, volume 1, pages 294–299 vol.1.

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., and Heß, S. (2016). MultiDark simulations: the story of dark matter halo concentrations and density profiles. *MNRAS*, 457(4):4340–4359.

Klypin, A. A., Trujillo-Gomez, S., and Primack, J. (2011). Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. *ApJ*, 740(2):102.

Kodra, D. (2019). *The Galaxy Morphology-Density Relation at High Redshift with Candels.* PhD thesis, University of Pittsburgh.

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., Grogin, N. A., Kocevski, D. D., Koo, D. C., Lai, K., Lotz, J. M., Lucas, R. A., McGrath, E. J., Ogaz, S., Rajan, A., Riess, A. G., Rodney, S. A., Strolger, L., Casertano, S., Castellano, M., Dahlen, T., Dickinson, M., Dolch, T., Fontana, A., Giavalisco, M., Grazian, A., Guo, Y., Hathi, N. P., Huang, K.-H., van der Wel, A., Yan, H.-J., Acquaviva, V., Alexander, D. M., Almaini, O., Ashby, M. L. N., Barden, M., Bell, E. F., Bournaud, F., Brown, T. M., Caputi, K. I., Cassata, P., Challis, P. J., Chary, R.-R., Cheung, E., Cirasuolo, M., Conselice, C. J., Roshan Cooray, A., Croton, D. J., Daddi, E., Davé, R., de Mello, D. F., de Ravel, L., Dekel, A., Donley, J. L., Dunlop, J. S., Dutton, A. A., Elbaz, D., Fazio, G. G., Filippenko, A. V., Finkelstein, S. L., Frazer, C., Gardner, J. P., Garnavich, P. M., Gawiser, E., Gruetzbauch, R., Hartley, W. G., Häussler, B., Herrington, J., Hopkins, P. F., Huang, J.-S., Jha, S. W., Johnson, A., Kartaltepe, J. S., Khostovan, A. A., Kirshner, R. P., Lani, C., Lee, K.-S., Li, W., Madau, P., McCarthy, P. J., McIntosh, D. H., McLure, R. J., McPartland, C., Mobasher, B., Moreira, H., Mortlock, A., Moustakas, L. A., Mozena, M., Nandra, K., Newman, J. A., Nielsen, J. L., Niemi, S., Noeske, K. G., Papovich, C. J., Pentericci, L., Pope, A., Primack, J. R., Ravindranath, S., Reddy, N. A., Renzini, A., Rix, H.-W., Robaina, A. R., Rosario, D. J., Rosati, P., Salimbeni, S., Scarlata, C., Siana, B., Simard, L., Smidt, J., Snyder, D., Somerville, R. S., Spinrad, H., Straughn, A. N., Telford, O., Teplitz, H. I., Trump, J. R., Vargas, C., Villforth, C., Wagner, C. R., Wand ro, P., Wechsler, R. H., Weiner, B. J., Wiklind, T., Wild, V., Wilson, G., Wuyts, S., and Yun, M. S. (2011). CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics. *ApJS*, 197(2):36.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4):261–283.

Kovács, A. and Szapudi, I. (2015). Star-galaxy separation strategies for WISE-2MASS all-sky infrared galaxy catalogues. *MNRAS*, 448(2):1305–1313.

Labbé, I., Huang, J., Franx, M., Rudnick, G., Barmby, P., Daddi, E., van Dokkum, P. G., Fazio, G. G., Schreiber, N. M. F., Moorwood, A. F. M., Rix, H.-W., Röttgering, H., Trujillo, I., and van der Werf, P. (2005). IRAC Mid-Infrared Imaging of the Hubble Deep Field-South: Star Formation Histories and Stellar Masses of Red Galaxies at z>2. *ApJL*, 624(2):L81–L84.

Lacey, C. and Cole, S. (1993). Merger rates in hierarchical models of galaxy formation. *MNRAS*, 262(3):627–649.

Lahav, O., Naim, A., Buta, R. J., Corwin, H. G., de Vaucouleurs, G., Dressler, A., Huchra, J. P., van den Bergh, S., Raychaudhury, S., Sodre, L., J., and Storrie-Lombardi, M. C. (1995). Galaxies, Human Eyes, and Artificial Neural Networks. *Science*, 267(5199):859–862.

Laidler, V. G., Papovich, C., Grogin, N. A., Idzi, R., Dickinson, M., Ferguson, H. C., Hilbert, B., Clubb, K., and Ravindranath, S. (2007). TFIT: A Photometry Package Using Prior Information for Mixed-Resolution Data Sets. *PASP*, 119(861):1325–1344.

Laird, E. S., Nandra, K., Georgakakis, A., Aird, J. A., Barmby, P., Conselice, C. J., Coil, A. L., Davis, M., Faber, S. M., Fazio, G. G., Guhathakurta, P., Koo, D. C., Sarajedini, V., and Willmer, C. N. A. (2009). AEGIS-X: the Chandra Deep Survey of the Extended Groth Strip. *ApJS*, 180(1):102–116.

Lane, K. P., Almaini, O., Foucaud, S., Simpson, C., Smail, I., McLure, R. J., Conselice, C. J., Cirasuolo, M., Page, M. J., Dunlop, J. S., Hirst, P., Watson, M. G., and Sekiguchi, K. (2007). The colour selection of distant galaxies in the UKIDSS Ultra Deep Survey Early Data Release. *MNRAS*, 379(1):L25–L29.

Larson, R. B. (1974). Effects of supernovae on the early evolution of galaxies. *MNRAS*, 169:229–246.

Larson, R. B., Tinsley, B. M., and Caldwell, C. N. (1980). The evolution of disk galaxies and the origin of S0 galaxies. *ApJ*, 237:692–707.

Lee, N., Sanders, D. B., Casey, C. M., Toft, S., Scoville, N. Z., Hung, C.-L., Le Floc'h, E., Ilbert, O., Zahid, H. J., Aussel, H., Capak, P., Kartaltepe, J. S., Kewley, L. J., Li, Y., Schawinski, K., Sheth, K., and Xiao, Q. (2015). A Turnover in the Galaxy Main Sequence of Star Formation at M $_*$ ~10^{10} M $_\odot$ for Redshifts z $<$ 1.3. *ApJ*, 801(2):80.

Leistedt, B. and Hogg, D. W. (2017a). Data-driven, Interpretable Photometric Redshifts Trained on Heterogeneous and Unrepresentative Data. *ApJ*, 838(1):5.

Leistedt, B. and Hogg, D. W. (2017b). Hierarchical Probabilistic Inference of the Color-Magnitude Diagram and Shrinkage of Stellar Distance Uncertainties. *AJ*, 154(6):222.

Leistedt, B., Hogg, D. W., Wechsler, R. H., and DeRose, J. (2019). Hierarchical Modeling and Statistical Calibration for Photometric Redshifts. *ApJ*, 881(1):80.

Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., and Speagle, J. S. (2019). How to Measure Galaxy Star Formation Histories. II. Nonparametric Models. *ApJ*, 876(1):3.

Leja, J., Speagle, J. S., Ting, Y.-S., Johnson, B. D., Conroy, C., Whitaker, K. E., Nelson, E. J., van Dokkum, P., and Franx, M. (2021). A New Census of the 0.2< z <3.0 Universe, Part II: The Star-Forming Sequence. *arXiv e-prints*, page arXiv:2110.04314.

Leslie, S. K., Schinnerer, E., Liu, D., Magnelli, B., Algera, H., Karim, A., Davidzon, I., Gozaliasl, G., Jiménez-Andrade, E. F., Lang, P., Sargent, M. T., Novak, M., Groves, B., Smolčić, V., Zamorani, G., Vaccari, M., Battisti, A., Vardoulaki, E., Peng, Y., and Kartaltepe, J. (2020). The VLA-COSMOS 3 GHz Large Project: Evolution of Specific Star Formation Rates out to z $\sim$ 5. *ApJ*, 899(1):58.

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., and Peng, Y. (2013). Gas Regulation of Galaxies: The Evolution of the Cosmic Specific Star Formation Rate, the Metallicity-Mass-Star-formation Rate Relation, and the Stellar Content of Halos. *ApJ*, 772:119.

Lin, L., Jian, H.-Y., Foucaud, S., Norberg, P., Bower, R. G., Cole, S., Arnalte-Mur, P., Chen, C.-W., Coupon, J., Hsieh, B.-C., Heinis, S., Phleps, S., Chen, W.-P., Lee, C.-H., Burgett, W., Chambers, K. C., Denneau, L., Draper, P., Flewelling, H., Hodapp, K. W., Huber, M. E., Kaiser, N., Kudritzki, R. P., Magnier, E. A., Metcalfe, N., Price, P. A., Tonry, J. L., Wainscoat, R. J., and Waters, C. (2014). The Pan-STARRS1 Medium-Deep Survey: The Role of Galaxy Group Environment in the Star Formation Rate versus Stellar Mass Relation and Quiescent Fraction out to z ~0.8. *ApJ*, 782(1):33.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, 389(3):1179–1189.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Loh, W.-Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1):14–23.

Loh, W.-Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, 82(3):329–348.

Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605.

Madau, P. (1995). Radiative Transfer in a Clumpy Universe: The Colors of High-Redshift Galaxies. *ApJ*, 441:18.

Malmquist, K. G. (1922). On some relations in stellar statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52.

Malmquist, K. G. (1925). A contribution to the problem of determining the distribution in space of the stars. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 106:1–12.

Man, A. W. S., Lehnert, M. D., Vernet, J. D. R., De Breuck, C., and Falkendal, T. (2019). Quenching by gas compression and consumption. A case study of a massive radio galaxy at z = 2.57. *A&A*, 624:A81.

Mancini, C., Matute, I., Cimatti, A., Daddi, E., Dickinson, M., Rodighiero, G., Bolzonella, M., and Pozzetti, L. (2009). Searching for massive galaxies at z $\geq$ 3.5 in GOODS-North. *A&A*, 500(2):705–723.

Mantz, A. B. (2019). Coping with selection effects: a Primer on regression with truncated data. *MNRAS*, 485(4):4863–4872.

Marchesini, D., Whitaker, K. E., Brammer, G., van Dokkum, P. G., Labbé, I., Muzzin, A., Quadri, R. F., Kriek, M., Lee, K.-S., Rudnick, G., Franx, M., Illingworth, G. D., and Wake, D. (2010). The Most Massive Galaxies at 3.0 <= z < 4.0 in the Newfirm Medium-band Survey: Properties and Improved Constraints on the Stellar Mass Function. *ApJ*, 725(1):1277–1295.

Marinacci, F., Vogelsberger, M., Pakmor, R., Torrey, P., Springel, V., Hernquist, L., Nelson, D., Weinberger, R., Pillepich, A., Naiman, J., and Genel, S. (2018). First results from the IllustrisTNG simulations: radio haloes and magnetic fields. *MNRAS*, 480(4):5113–5139.

Martin, D. C., Gonçalves, T. S., Darvish, B., Seibert, M., and Schiminovich, D. (2017). Quenching or Bursting: Star Formation Acceleration—A New Methodology for Tracing Galaxy Evolution. *ApJ*, 842(1):20.

Masini, A., Civano, F., Comastri, A., Fornasini, F., Ballantyne, D. R., Lansbury, G. B., Treister, E., Alexand er, D. M., Boorman, P. G., Brandt, W. N., Farrah, D., Gandhi, P., Harrison, F. A., Hickox, R. C., Kocevski, D. D., Lanz, L., Marchesi, S., Puccetti, S., Ricci, C., Saez, C., Stern, D., and Zappacosta, L. (2018). The NuSTAR Extragalactic Surveys: Source Catalog and the Compton-thick Fraction in the UDS Field. *ApJS*, 235(1):17.

Masters, D. C., Stern, D. K., Cohen, J. G., Capak, P. L., Stanford, S. A., Hernitschek, N., Galametz, A., Davidzon, I., Rhodes, J. D., Sand ers, D., Mobasher, B., Castander, F., Pruett, K., and Fotopoulou, S. (2019). The Complete Calibration of the Color-Redshift Relation (C3R2) Survey: Analysis and Data Release 2. *ApJ*, 877(2):81.

McAlpine, S., Helly, J. C., Schaller, M., Trayford, J. W., Qu, Y., Furlong, M., Bower, R. G., Crain, R. A., Schaye, J., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Jenkins, A., Rosas-Guevara, Y., White, S. D. M., Baes, M., Camps, P., and Lemson, G. (2016). The EAGLE simulations of galaxy formation: Public release of halo and galaxy catalogues. *Astronomy and Computing*, 15:72–89.

McCracken, H. J., Milvang-Jensen, B., Dunlop, J., Franx, M., Fynbo, J. P. U., Le Fèvre, O., Holt, J., Caputi, K. I., Goranova, Y., Buitrago, F., Emerson, J. P., Freudling, W., Hudelot, P., López-Sanjuan, C., Magnard, F., Mellier, Y., Møller, P., Nilsson, K. K., Sutherland, W., Tasca, L., and Zabl, J. (2012). UltraVISTA: a new ultra-deep near-infrared survey in COSMOS. *A&A*, 544:A156.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

McGee, S. L., Bower, R. G., and Balogh, M. L. (2014). Overconsumption, outflows and the quenching of satellite galaxies. *MNRAS*, 442:L105–L109.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*.

McInnes, L., Healy, J., Saul, N., and Grossberger, L. (2018). Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861.

Ménard, B., Scranton, R., Schmidt, S., Morrison, C., Jeong, D., Budavari, T., and Rahman, M. (2013). Clustering-based redshift estimation: method and application to data. *arXiv e-prints*, page arXiv:1303.4722.

Merlin, E., Fontana, A., Castellano, M., Santini, P., Torelli, M., Boutsia, K., Wang, T., Grazian, A., Pentericci, L., Schreiber, C., Ciesla, L., McLure, R., Derriere, S., Dunlop, J. S., and Elbaz, D. (2018). Chasing passive galaxies in the early Universe: a critical analysis in CANDELS GOODS-South. *MNRAS*, 473(2):2098–2123.

Merlin, E., Fortuni, F., Torelli, M., Santini, P., Castellano, M., Fontana, A., Grazian, A., Pentericci, L., Pilo, S., and Schmidt, K. B. (2019). Red & Dead CANDELS: massive passive galaxies at the dawn of the Universe. *MNRAS*, page 2241.

Minsky, M. and Papert, S. A. (1969). *Perceptrons: An introduction to computational geometry.* MIT press.

Mobasher, B., Dahlen, T., Ferguson, H. C., Acquaviva, V., Barro, G., Finkelstein, S. L., Fontana, A., Gruetzbauch, R., Johnson, S., Lu, Y., Papovich, C. J., Pforr, J., Salvato, M., Somerville, R. S., Wiklind, T., Wuyts, S., Ashby, M. L. N., Bell, E., Conselice, C. J., Dickinson, M. E., Faber, S. r. M., Fazio, G., Finlator, K., Galametz, A., Gawiser, E., Giavalisco, M., Grazian, A., Grogin, N. A., Guo, Y., Hathi, N., Kocevski, D., Koekemoer, A. M., Koo, D. C., Newman, J. A., Reddy, N., Santini, P., and Wechsler, R. H. (2015). A Critical Assessment of Stellar Mass Measurement Methods. *ApJ*, 808(1):101.

Mobasher, B., Dickinson, M., Ferguson, H. C., Giavalisco, M., Wiklind, T., Stark, D., Ellis, R. S., Fall, S. M., Grogin, N. A., Moustakas, L. A., Panagia, N., Sosey, M., Stiavelli, M., Bergeron, E., Casertano, S., Ingraham, P., Koekemoer, A., Labbé, I., Livio, M., Rodgers, B., Scarlata, C., Vernet, J., Renzini, A., Rosati, P., Kuntschner, H., Kümmel, M., Walsh, J. R., Chary, R., Eisenhardt, P., Pirzkal, N., and Stern, D. (2005). Evidence for a Massive Poststarburst Galaxy at z~6.5. *ApJ*, 635(2):832–844.

Moster, B. P., Naab, T., and White, S. D. M. (2013). Galactic star formation and accretion histories from matching galaxies to dark matter haloes. *MNRAS*, 428(4):3121–3138.

Moster, B. P., Somerville, R. S., Maulbetsch, C., van den Bosch, F. C., Macciò, A. V., Naab, T., and Oser, L. (2010). Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. *ApJ*, 710(2):903–923.

Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery*, 2(4):345–389.

Muzzin, A., Marchesini, D., Stefanon, M., Franx, M., McCracken, H. J., Milvang-Jensen, B., Dunlop, J. S., Fynbo, J. P. U., Brammer, G., Labbé, I., and van Dokkum, P. G. (2013). The Evolution of the Stellar Mass Functions of Star-forming and Quiescent Galaxies to z = 4 from the COSMOS/UltraVISTA Survey. *ApJ*, 777:18.

Muzzin, A., Marchesini, D., van Dokkum, P. G., Labbé, I., Kriek, M., and Franx, M. (2009). A Near-Infrared Spectroscopic Survey of K-Selected Galaxies at z~2.3: Comparison of Stellar Population Synthesis Codes and Constraints from the Rest-Frame NIR. *ApJ*, 701(2):1839–1864.

Muzzin, A., Wilson, G., Yee, H. K. C., Gilbank, D., Hoekstra, H., Demarco, R., Balogh, M., van Dokkum, P., Franx, M., Ellingson, E., Hicks, A., Nantais, J., Noble, A., Lacy, M., Lidman, C., Rettura, A., Surace, J., and Webb, T. (2012). The Gemini Cluster Astrophysics Spectroscopic Survey (GCLASS): The Role of Environment and Self-regulation in Galaxy Evolution at z ~1. *ApJ*, 746(2):188.

Naab, T. and Ostriker, J. P. (2017). Theoretical challenges in galaxy formation. *Annual Review of Astronomy and Astrophysics*, 55(1):59–109.

Naiman, J. P., Pillepich, A., Springel, V., Ramirez-Ruiz, E., Torrey, P., Vogelsberger, M., Pakmor, R., Nelson, D., Marinacci, F., Hernquist, L., Weinberger, R., and Genel, S. (2018). First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. *MNRAS*, 477(1):1206–1224.

Nandra, K., Laird, E. S., Aird, J. A., Salvato, M., Georgakakis, A., Barro, G., Perez-Gonzalez, P. G., Barmby, P., Chary, R. R., Coil, A., Cooper, M. C., Davis, M., Dickinson, M., Faber, S. M., Fazio, G. G., Guhathakurta, P., Gwyn, S., Hsu, L. T., Huang, J. S., Ivison, R. J., Koo, D. C., Newman, J. A., Rangel, C., Yamada, T., and Willmer, C. (2015). AEGIS-X: Deep Chandra Imaging of the Central Groth Strip. *ApJS*, 220(1):10.

Navarro, J. F. and Benz, W. (1991). Dynamics of Cooling Gas in Galactic Dark Halos. *ApJ*, 380:320.

Nayyeri, H., Hemmati, S., Mobasher, B., Ferguson, H. C., Cooray, A., Barro, G., Faber, S. M., Dickinson, M., Koekemoer, A. M., Peth, M., Salvato, M., Ashby, M. L. N., Darvish, B., Donley, J., Durbin, M., Finkelstein, S., Fontana, A., Grogin, N. A., Gruetzbauch, R., Huang, K., Khostovan, A. A., Kocevski, D., Kodra, D., Lee, B., Newman, J., Pacifici, C., Pforr, J., Stefanon, M., Wiklind, T., Willner, S. P., Wuyts, S., Castellano, M., Conselice, C., Dolch, T., Dunlop, J. S., Galametz, A., Hathi, N. P., Lucas, R. A., and Yan, H. (2017). CANDELS Multi-wavelength Catalogs: Source Identification and Photometry in the CANDELS COSMOS Survey Field. *ApJS*, 228:7.

Nayyeri, H., Mobasher, B., Hemmati, S., De Barros, S., Ferguson, H. C., Wiklind, T., Dahlen, T., Dickinson, M., Giavalisco, M., Fontana, A., Ashby, M., Barro, G., Guo, Y., Hathi, N. P., Kassin, S., Koekemoer, A., Willner, S., Dunlop, J. S., Paris, D., and Targett, T. A. (2014). A Study of Massive and Evolved Galaxies at High Redshift. *ApJ*, 794(1):68.

Neal, R. M. et al. (2011). Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2.

Nelson, D., Pillepich, A., Springel, V., Weinberger, R., Hernquist, L., Pakmor, R., Genel, S., Torrey, P., Vogelsberger, M., Kauffmann, G., Marinacci, F., and Naiman, J. (2018). First results from the IllustrisTNG simulations: the galaxy colour bimodality. *MNRAS*, 475(1):624–647.

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B., and Hernquist, L. (2019). The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2.

Nemtsov, A., Averbuch, A., and Schclar, A. (2013). Matrix Compression using the Nystroöm Method. *arXiv e-prints*, page arXiv:1305.0203.

Newman, A. B., Belli, S., and Ellis, R. S. (2015). Discovery of a Strongly Lensed Massive Quiescent Galaxy at z = 2.636: Spatially Resolved Spectroscopy and Indications of Rotation. *ApJL*, 813(1):L7.

Newman, A. B., Belli, S., Ellis, R. S., and Patel, S. G. (2018a). Resolving Quiescent Galaxies at z 2. I. Search for Gravitationally Lensed Sources and Characterization of Their Structure, Stellar Populations, and Line Emission. *ApJ*, 862(2):125.

Newman, A. B., Belli, S., Ellis, R. S., and Patel, S. G. (2018b). Resolving Quiescent Galaxies at z 2. I. Search for Gravitationally Lensed Sources and Characterization of Their Structure, Stellar Populations, and Line Emission. *ApJ*, 862(2):125.

Newman, A. B., Ellis, R. S., Treu, T., and Bundy, K. (2010). Keck Spectroscopy of z > 1 Field Spheroidals: Dynamical Constraints on the Growth Rate of Red "Nuggets". *ApJL*, 717(2):L103–L107.

Noeske, K. G., Weiner, B. J., Faber, S. M., Papovich, C., Koo, D. C., Somerville, R. S., Bundy, K., Conselice, C. J., Newman, J. A., Schiminovich, D., Le Floc'h, E., Coil, A. L., Rieke, G. H., Lotz, J. M., Primack, J. R., Barmby, P., Cooper, M. C., Davis, M., Ellis, R. S., Fazio, G. G., Guhathakurta, P., Huang, J., Kassin, S. A., Martin, D. C., Phillips, A. C., Rich, R. M., Small, T. A., Willmer, C. N. A., and Wilson, G. (2007). Star Formation in AEGIS Field Galaxies since $z=1.1$: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies. *ApJL*, 660:L43–L46.

Odewahn, S. C., Stockwell, E. B., Pennington, R. L., Humphreys, R. M., and Zumach, W. A. (1992). Automated Star/Galaxy Discrimination With Neural Networks. *AJ*, 103:318.

Odewahn, S. C., Windhorst, R. A., Driver, S. P., and Keel, W. C. (1996). Automated Morphological Classification in Deep Hubble Space Telescope UBVI Fields: Rapidly and Passively Evolving Faint Galaxy Populations. *ApJL*, 472:L13.

Oesch, P. A., Brammer, G., van Dokkum, P. G., Illingworth, G. D., Bouwens, R. J., Labbé, I., Franx, M., Momcheva, I., Ashby, M. L. N., Fazio, G. G., Gonzalez, V., Holden, B., Magee, D., Skelton, R. E., Smit, R., Spitler, L. R., Trenti, M., and Willner, S. P. (2016). A Remarkably Luminous Galaxy at $z=11.1$ Measured with Hubble Space Telescope Grism Spectroscopy. *ApJ*, 819(2):129.

Oke, J. B. and Gunn, J. E. (1983). Secondary standard stars for absolute spectrophotometry. *ApJ*, 266:713–717.

Old, L. J., Balogh, M. L., van der Burg, R. F. J., Biviano, A., Yee, H. K. C., Pintos-Castro, I., Webb, K., Muzzin, A., Rudnick, G., Vulcani, B., Poggianti, B., Cooper, M., Zaritsky, D., Cerulo, P., Wilson, G., Chan, J. C. C., Lidman, C., McGee, S., Demarco, R., Forrest, B., De Lucia, G., Gilbank, D., Kukstas, E., McCarthy, I. G., Jablonka, P., Nantais, J., Noble, A., Reeves, A. M. M., and Shipley, H. (2020). The GOGREEN survey: the environmental dependence of the star-forming galaxy main sequence at $1.0 < z < 1.5$. *MNRAS*, 493(4):5987–6000.

Pacifici, C., Kassin, S. A., Weiner, B. J., Holden, B., Gardner, J. P., Faber, S. M., Ferguson, H. C., Koo, D. C., Primack, J. R., Bell, E. F., Dekel, A., Gawiser, E., Giavalisco, M., Rafelski, M., Simons, R. C., Barro, G., Croton, D. J., Davé, R., Fontana, A., Grogin, N. A., Koekemoer, A. M., Lee, S.-K., Salmon, B., Somerville, R., and Behroozi, P. (2016).

The Evolution of Star Formation Histories of Quiescent Galaxies. *ApJ*, 832(1):79.

Papovich, C., Dickinson, M., and Ferguson, H. C. (2001). The Stellar Populations and Evolution of Lyman Break Galaxies. *ApJ*, 559(2):620–653.

Patel, S. G., Kelson, D. D., Holden, B. P., Franx, M., and Illingworth, G. D. (2011). The Star-formation-rate-Density Relation at $0.6 < z < 0.9$ and the Role of Star-forming Galaxies. *ApJ*, 735(1):53.

Pearson, W. J., Wang, L., Hurley, P. D., Małek, K., Buat, V., Burgarella, D., Farrah, D., Oliver, S. J., Smith, D. J. B., and van der Tak, F. F. S. (2018). Main sequence of star forming galaxies beyond the Herschel confusion limit. *A&A*, 615:A146.

Peebles, P. J. E. (1970). Structure of the Coma Cluster of Galaxies. *AJ*, 75:13.

Peng, Y.-j., Lilly, S. J., Kovač, K., Bolzonella, M., Pozzetti, L., Renzini, A., Zamorani, G., Ilbert, O., Knobel, C., Iovino, A., Maier, C., Cucciati, O., Tasca, L., Carollo, C. M., Silverman, J., Kampczyk, P., de Ravel, L., Sanders, D., Scoville, N., Contini, T., Mainieri, V., Scodeggio, M., Kneib, J.-P., Le Fèvre, O., Bardelli, S., Bongiorno, A., Caputi, K., Coppa, G., de la Torre, S., Franzetti, P., Garilli, B., Lamareille, F., Le Borgne, J.-F., Le Brun, V., Mignoli, M., Perez Montero, E., Pello, R., Ricciardelli, E., Tanaka, M., Tresse, L., Vergani, D., Welikala, N., Zucca, E., Oesch, P., Abbas, U., Barnes, L., Bordoloi, R., Bottini, D., Cappi, A., Cassata, P., Cimatti, A., Fumana, M., Hasinger, G., Koekemoer, A., Leauthaud, A., Maccagni, D., Marinoni, C., McCracken, H., Memeo, P., Meneux, B., Nair, P., Porciani, C., Presotto, V., and Scaramella, R. (2010). Mass and Environment as Drivers of Galaxy Evolution in SDSS and zCOSMOS and the Origin of the Schechter Function. *ApJ*, 721(1):193–221.

Pillepich, A., Nelson, D., Hernquist, L., Springel, V., Pakmor, R., Torrey, P., Weinberger, R., Genel, S., Naiman, J. P., Marinacci, F., and Vogelsberger, M. (2018). First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. *MNRAS*, 475(1):648–675.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., Battaner, E., Battye, R., Benabed, K., Benoît, A., Benoit-Lévy, A., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bonaldi, A., Bonavera, L., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Catalano, A., Challinor, A., Chamballu, A., Chary, R. R., Chiang, H. C., Chluba, J., Christensen, P. R., Church, S., Clements, D. L., Colombi, S., Colombo, L. P. L., Combet, C., Coulais, A., Crill, B. P., Curto, A., Cuttaia, F., Danese, L., Davies, R. D., Davis, R. J., de Bernardis, P., de Rosa, A., de Zotti, G., Delabrouille, J., Désert, F. X., Di Valentino, E., Dickinson, C., Diego, J. M., Dolag, K., Dole, H., Donzelli, S., Doré, O., Douspis, M., Ducout, A., Dunkley, J., Dupac, X., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen,

H. K., Farhang, M., Fergusson, J., Finelli, F., Forni, O., Frailis, M., Fraisse, A. A., Franceschi, E., Frejsel, A., Galeotta, S., Galli, S., Ganga, K., Gauthier, C., Gerbino, M., Ghosh, T., Giard, M., Giraud-Héraud, Y., Giusarma, E., Gjerløw, E., González-Nuevo, J., Górski, K. M., Gratton, S., Gregorio, A., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Hansen, F. K., Hanson, D., Harrison, D. L., Helou, G., Henrot-Versillé, S., Hernández-Monteagudo, C., Herranz, D., Hildebrand t, S. R., Hivon, E., Hobson, M., Holmes, W. A., Hornstrup, A., Hovest, W., Huang, Z., Huffenberger, K. M., Hurier, G., Jaffe, A. H., Jaffe, T. R., Jones, W. C., Juvela, M., Keihänen, E., Keskitalo, R., Kisner, T. S., Kneissl, R., Knoche, J., Knox, L., Kunz, M., Kurki-Suonio, H., Lagache, G., Lähteenmäki, A., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Leahy, J. P., Leonardi, R., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Linden-Vørnle, M., López-Caniego, M., Lubin, P. M., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marchini, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Masi, S., Matarrese, S., McGehee, P., Meinhold, P. R., Melchiorri, A., Melin, J. B., Mendes, L., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M. A., Moneti, A., Montier, L., Morgante, G., Mortlock, D., Moss, A., Munshi, D., Murphy, J. A., Naselsky, P., Nati, F., Natoli, P., Netterfield, C. B., Nørgaard-Nielsen, H. U., Noviello, F., Novikov, D., Novikov, I., Oxborrow, C. A., Paci, F., Pagano, L., Pajot, F., Paladini, R., Paoletti, D., Partridge, B., Pasian, F., Patanchon, G., Pearson, T. J., Perdereau, O., Perotto, L., Perrotta, F., Pettorino, V., Piacentini, F., Piat, M., Pierpaoli, E., Pietrobon, D., Plaszczynski, S., Pointecouteau, E., Polenta, G., Popa, L., Pratt, G. W., Prézeau, G., Prunet, S., Puget, J. L., Rachen, J. P., Reach, W. T., Rebolo, R., Reinecke, M., Remazeilles, M., Renault, C., Renzi, A., Ristorcelli, I., Rocha, G., Rosset, C., Rossetti, M., Roudier, G., Rouillé d'Orfeuil, B., Rowan-Robinson, M., Rubiño-Martín, J. A., Rusholme, B., Said, N., Salvatelli, V., Salvati, L., Sandri, M., Santos, D., Savelainen, M., Savini, G., Scott, D., Seiffert, M. D., Serra, P., Shellard, E. P. S., Spencer, L. D., Spinelli, M., Stolyarov, V., Stompor, R., Sudiwala, R., Sunyaev, R., Sutton, D., Suur-Uski, A. S., Sygnet, J. F., Tauber, J. A., Terenzi, L., Toffolatti, L., Tomasi, M., Tristram, M., Trombetti, T., Tucci, M., Tuovinen, J., Türler, M., Umana, G., Valenziano, L., Valiviita, J., Van Tent, F., Vielva, P., Villa, F., Wade, L. A., Wandelt, B. D., Wehus, I. K., White, M., White, S. D. M., Wilkinson, A., Yvon, D., Zacchei, A., and Zonca, A. (2016). Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J. P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., Bouchet, F. R., Boulanger, F., Bucher, M., Burigana, C., Butler, R. C., Calabrese, E., Cardoso, J. F., Carron, J., Challinor, A., Chiang, H. C., Chluba, J., Colombo, L. P. L., Combet, C., Contreras, D., Crill, B. P., Cuttaia, F., de Bernardis, P., de Zotti, G., Delabrouille, J., Delouis, J. M., Di Valentino, E., Diego, J. M., Doré, O., Douspis, M., Ducout, A., Dupac, X., Dusini, S., Efstathiou, G., Elsner, F., Enßlin, T. A., Eriksen, H. K., Fantaye, Y., Farhang, M., Fergusson, J., Fernandez-Cobos, R., Finelli, F., Forastieri, F., Frailis, M., Fraisse, A. A., Franceschi, E., Frolov, A., Galeotta, S., Galli, S., Ganga, K., Génova-Santos, R. T., Gerbino, M., Ghosh, T., González-Nuevo, J., Górski, K. M., Gratton, S., Gruppuso, A., Gudmundsson, J. E., Hamann, J., Handley, W., Hansen, F. K., Herranz, D., Hildebrandt, S. R., Hivon, E., Huang, Z., Jaffe, A. H., Jones, W. C., Karakci, A., Keihänen, E., Keskitalo, R., Kiiveri, K., Kim, J., Kisner, T. S., Knox, L., Krachmalnicoff, N., Kunz, M., Kurki-Suonio, H., Lagache, G., Lamarre, J. M., Lasenby, A., Lattanzi, M., Lawrence, C. R., Le Jeune, M., Lemos, P., Lesgourgues, J., Levrier, F., Lewis, A., Liguori, M., Lilje, P. B., Lilley, M., Lindholm, V., López-Caniego, M., Lubin, P. M., Ma, Y. Z., Macías-Pérez, J. F., Maggio, G., Maino, D., Mandolesi, N., Mangilli, A., Marcos-Caballero, A., Maris, M., Martin, P. G., Martinelli, M., Martínez-González, E., Matarrese, S., Mauri, N., McEwen, J. D., Meinhold, P. R., Melchiorri, A., Mennella, A., Migliaccio, M., Millea, M., Mitra, S., Miville-Deschênes, M. A., Molinari, D., Montier, L., Morgante, G., Moss, A., Natoli, P., Nørgaard-Nielsen, H. U., Pagano, L., Paoletti, D., Partridge, B., Patanchon, G., Peiris, H. V., Perrotta, F., Pettorino, V., Piacentini, F., Polastri, L., Polenta, G., Puget, J. L., Rachen, J. P., Reinecke, M., Remazeilles, M., Renzi, A., Rocha, G., Rosset, C., Roudier, G., Rubiño-Martín, J. A., Ruiz-Granados, B., Salvati, L., Sandri, M., Savelainen, M., Scott, D., Shellard, E. P. S., Sirignano, C., Sirri, G., Spencer, L. D., Sunyaev, R., Suur-Uski, A. S., Tauber, J. A., Tavagnacco, D., Tenti, M., Toffolatti, L., Tomasi, M., Trombetti, T., Valenziano, L., Valiviita, J., Van Tent, B., Vibert, L., Vielva, P., Villa, F., Vittorio, N., Wand elt, B. D., Wehus, I. K., White, M., White, S. D. M., Zacchei, A., and Zonca, A. (2018). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, page arXiv:1807.06209.

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. pages 61–74.

Poličar, P. G., Stražar, M., and Zupan, B. (2019). opentsne: a modular python library for t-sne dimensionality reduction and embedding. *bioRxiv*.

Popping, G., Somerville, R. S., and Trager, S. C. (2014). Evolution of the atomic and molecular gas content of galaxies. *Monthly Notices of the Royal Astronomical Society*, 442(3):2398–2418.

Pozzetti, L., Bolzonella, M., Zucca, E., Zamorani, G., Lilly, S., Renzini, A., Moresco, M., Mignoli, M., Cassata, P., Tasca, L., Lamareille, F., Maier, C., Meneux, B., Halliday, C., Oesch, P., Vergani, D., Caputi, K., Kovač, K., Cimatti, A., Cucciati, O., Iovino, A., Peng, Y., Carollo, M., Contini, T., Kneib, J. P., Le Févre, O., Mainieri, V., Scodeggio, M., Bardelli, S., Bongiorno, A., Coppa, G., de la Torre, S., de Ravel, L., Franzetti, P., Garilli, B., Kampczyk, P., Knobel, C., Le Borgne, J. F., Le Brun, V., Pellò, R., Perez Montero, E., Ricciardelli, E., Silverman, J. D., Tanaka, M., Tresse, L., Abbas, U., Bottini, D., Cappi, A., Guzzo, L., Koekemoer, A. M., Leauthaud, A., Maccagni, D., Marinoni, C., McCracken, H. J., Memeo, P., Porciani, C., Scaramella, R., Scarlata, C., and Scoville, N. (2010). zCOSMOS - 10k-bright spectroscopic sample. The bimodality in the galaxy stellar mass function: exploring its evolution with redshift. *A&A*, 523:A13.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by

Self-Similar Gravitational Condensation. *ApJ*, 187:425–438.

Quinn, T., Katz, N., and Efstathiou, G. (1996). Photoionization and the formation of dwarf galaxies. *MNRAS*, 278(4):L49–L54.

Reddick, R. M., Wechsler, R. H., Tinker, J. L., and Behroozi, P. S. (2013). The Connection between Galaxies and Dark Matter Structures in the Local Universe. *ApJ*, 771(1):30.

Reddy, N. A., Erb, D. K., Steidel, C. C., Shapley, A. E., Adelberger, K. L., and Pettini, M. (2005). A Census of Optical and Near-Infrared Selected Star-forming and Passively Evolving Galaxies at Redshift z ~2. *ApJ*, 633(2):748–767.

Reddy, N. A., Pettini, M., Steidel, C. C., Shapley, A. E., Erb, D. K., and Law, D. R. (2012). The Characteristic Star Formation Histories of Galaxies at Redshifts z ~2-7. *ApJ*, 754(1):25.

Rees, M. J. (1986). Lyman absorption lines in quasar spectra - Evidence for gravitationally-confined gas in dark minihaloes. *MNRAS*, 218:25P–30P.

Rees, M. J. and Ostriker, J. P. (1977). Cooling, dynamics and fragmentation of massive gas clouds: clues to the masses and radii of galaxies and clusters. *MNRAS*, 179:541–559.

Renzini, A. and Peng, Y.-j. (2015). An Objective Definition for the Main Sequence of Star-forming Galaxies. *ApJL*, 801(2):L29.

Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., Labbe, I., Smit, R., Illingworth, G. D., van Dokkum, P., Holden, B., Gonzalez, V., Stefanon, M., Holwerda, B., and Wilkins, S. (2016). z 7 Galaxies with Red Spitzer/IRAC [3.6]-[4.5] Colors in the Full CANDELS Data Set: The Brightest-Known Galaxies at z ~7-9 and a Probable Spectroscopic Confirmation at z = 7.48. *ApJ*, 823(2):143.

Robertson, B., Cox, T. J., Hernquist, L., Franx, M., Hopkins, P. F., Martini, P., and Springel, V. (2006). The Fundamental Scaling Relations of Elliptical Galaxies. *ApJ*, 641(1):21–40.

Roche, N. D., Almaini, O., Dunlop, J., Ivison, R. J., and Willott, C. J. (2002). The clustering, number counts and morphology of extremely red (R-K > 5) galaxies to K<= 21. *MNRAS*, 337(4):1282–1298.

Rodighiero, G., Cimatti, A., Franceschini, A., Brusa, M., Fritz, J., and Bolzonella, M. (2007). Unveiling the oldest and most massive galaxies at very high redshift. *A&A*, 470(1):21–37.

Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., Sijacki, D., Pillepich, A., Sales, L. V., Torrey, P., Snyder, G., Nelson, D., Springel, V., Ma, C.-P., and Hernquist, L. (2015). The merger rate of galaxies in the Illustris simulation: a comparison with observations and semi-empirical models. *MNRAS*, 449(1):49–64.

Rodríguez-Puebla, A., Behroozi, P., Primack, J., Klypin, A., Lee, C., and Hellinger, D. (2016). Halo and subhalo demographics with Planck cosmological parameters: Bolshoi-Planck and MultiDark-Planck simulations. *MNRAS*, 462(1):893–916.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

Rudnick, G., Rix, H.-W., Franx, M., Labbé, I., Blanton, M., Daddi, E., Förster Schreiber, N. M., Moorwood, A., Röttgering, H., Trujillo, I., van der Wel, A., van der Werf, P., van Dokkum, P. G., and van Starkenburg, L. (2003). The Rest-Frame Optical Luminosity Density, Color, and Stellar Mass Density of the Universe from z = 0 to z = 3. *ApJ*, 599(2):847–864.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Salim, S., Rich, R. M., Charlot, S., Brinchmann, J., Johnson, B. D., Schiminovich, D., Seibert, M., Mallery, R., Heckman, T. M., Forster, K., Friedman, P. G., Martin, D. C., Morrissey, P., Neff, S. G., Small, T., Wyder, T. K., Bianchi, L., Donas, J., Lee, Y.-W., Madore, B. F., Milliard, B., Szalay, A. S., Welsh, B. Y., and Yi, S. K. (2007). UV Star Formation Rates in the Local Universe. *ApJS*, 173(2):267–292.

Sandage, A., Tammann, G. A., and Yahil, A. (1979). The velocity field of bright nearby galaxies. I. The variation of mean absolute magnitude with redshift for galaxies in a magnitude-limited sample. *ApJ*, 232:352–364.

Santini, P., Castellano, M., Merlin, E., Fontana, A., Fortuni, F., Kodra, D., Magnelli, B., Menci, N., Calabrò, A., Lovell, C. C., Pentericci, L., Testa, V., and Wilkins, S. M. (2020). The emergence of passive galaxies in the early Universe. *arXiv e-prints*, page arXiv:2011.10584.

Santini, P., Ferguson, H. C., Fontana, A., Mobasher, B., Barro, G., Castellano, M., Finkelstein, S. L., Grazian, A., Hsu, L. T., Lee, B., Lee, S. K., Pforr, J., Salvato, M., Wiklind, T., Wuyts, S., Almaini, O., Cooper, M. C., Galametz, A., Weiner, B., Amorin, R., Boutsia, K., Conselice, C. J., Dahlen, T., Dickinson, M. E., Giavalisco, M., Grogin, N. A., Guo, Y., Hathi, N. P., Kocevski, D., Koekemoer, A. M., Kurczynski, P., Merlin, E., Mortlock, A., Newman, J. A., Paris, D., Pentericci, L., Simons, R., and Willner, S. P. (2015). Stellar Masses from the CANDELS Survey: The GOODS-South and UDS Fields. *ApJ*, 801(2):97.

Santini, P., Fontana, A., Grazian, A., Salimbeni, S., Fiore, F., Fontanot, F., Boutsia, K., Castellano, M., Cristiani, S., de Santis, C., Gallozzi, S., Giallongo, E., Menci, N., Nonino, M., Paris, D., Pentericci, L., and Vanzella, E. (2009). Star formation and mass assembly in high redshift galaxies. *A&A*, 504(3):751–767.

Santini, P., Merlin, E., Fontana, A., Magnelli, B., Paris, D., Castellano, M., Grazian, A., Pentericci, L., Pilo, S., and Torelli, M. (2019). Passive galaxies in the early Universe: ALMA confirmation of z $\sim$ 3-5 candidates in the CANDELS GOODS-South field. *MNRAS*, 486(1):560–569.

Saracco, P., Marchesini, D., La Barbera, F., Gargiulo, A., Annunziatella, M., Forrest, B., Lange Vagle, D. J., Marsan, Z. C., Muzzin, A., Stefanon, M., and Wilson, G. (2020). The Rapid Buildup of Massive Early-type Galaxies: Supersolar Metallicity, High Velocity Dispersion, and Young Age for an Early-type Galaxy at z = 3.35. *ApJ*, 905(1):40.

Sargent, M. T., Béthermin, M., Daddi, E., and Elbaz, D. (2012). The Contribution of Starbursts and Normal Galaxies to Infrared Luminosity Functions at z < 2. *ApJL*, 747(2):L31.

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., Sawala, T., Thomas, P. A., and Trayford, J. (2015). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *MNRAS*, 446(1):521–554.

Schechter, P. (1976). An analytic expression for the luminosity function for galaxies. *ApJ*, 203:297–306.

Schmidt, M. (1959). The Rate of Star Formation. *ApJ*, 129:243.

Schmidt, M. (1968). Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *ApJ*, 151:393.

Schreiber, C., Glazebrook, K., Nanayakkara, T., Kacprzak, G. G., Labbé, I., Oesch, P., Yuan, T., Tran, K. V., Papovich, C., Spitler, L., and Straatman, C. (2018a). Near infrared spectroscopy and star-formation histories of 3 $\leq$ z $\leq$ 4 quiescent galaxies. *A&A*, 618:A85.

Schreiber, C., Glazebrook, K., Nanayakkara, T., Kacprzak, G. G., Labbé, I., Oesch, P., Yuan, T., Tran, K. V., Papovich, C., Spitler, L., and Straatman, C. (2018b). Near infrared spectroscopy and star-formation histories of 3 $\leq$ z $\leq$ 4 quiescent galaxies. *A&A*, 618:A85.

Schreiber, C., Labbé, I., Glazebrook, K., Bekiaris, G., Papovich, C., Costa, T., Elbaz, D., Kacprzak, G. G., Nanayakkara, T., Oesch, P., Pannella, M., Spitler, L., Straatman, C., Tran, K. V., and Wang, T. (2018c). Jekyll & Hyde: quiescence and extreme obscuration in a pair of massive galaxies 1.5 Gyr after the Big Bang. *A&A*, 611:A22.

Schreiber, C., Pannella, M., Elbaz, D., Béthermin, M., Inami, H., Dickinson, M., Magnelli, B., Wang, T., Aussel, H., Daddi, E., Juneau, S., Shu, X., Sargent, M. T., Buat, V., Faber, S. M., Ferguson, H. C., Giavalisco, M., Koekemoer, A. M., Magdis, G., Morrison, G. E., Papovich, C., Santini, P., and Scott, D. (2015). The Herschel view of the dominant mode of galaxy growth from z = 4 to the present day. *A&A*, 575:A74.

Schreiber, C., Pannella, M., Leiton, R., Elbaz, D., Wang, T., Okumura, K., and Labbé, I. (2017). The ALMA Redshift 4 Survey (AR4S). I. The massive end of the z = 4 main sequence of galaxies. *A&A*, 599:A134.

Sevilla-Noarbe, I., Hoyle, B., Marchã, M. J., Soumagnac, M. T., Bechtol, K., Drlica-Wagner, A., Abdalla, F., Aleksić, J., Avestruz, C., Balbinot, E., Banerji, M., Bertin, E., Bonnett, C., Brunner, R., Carrasco-Kind, M., Choi, A., Giannantonio, T., Kim, E., Lahav, O., Moraes, B., Nord, B., Ross, A. J., Rykoff, E. S., Santiago, B., Sheldon, E., Wei, K., Wester, W., Yanny, B., Abbott, T., Allam, S., Brooks, D., Carnero-Rosell, A., Carretero, J., Cunha, C., da Costa, L., Davis, C., de Vicente, J., Desai, S., Doel, P., Fernandez, E., Flaugher, B., Frieman, J., Garcia-Bellido, J., Gaztanaga, E., Gruen, D., Gruendl, R., Gschwend, J., Gutierrez, G., Hollowood, D. L., Honscheid, K., James, D., Jeltema, T., Kirk, D., Krause, E., Kuehn, K., Li, T. S., Lima, M., Maia, M. A. G., March, M., McMahon, R. G., Menanteau, F., Miquel, R., Ogando, R. L. C., Plazas, A. A., Sanchez, E., Scarpine, V., Schindler, R., Schubnell, M., Smith, M., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., Tucker, D. L., Walker, A. R., and DES Collaboration (2018). Star-galaxy classification in the Dark Energy Survey Y1 data set. *MNRAS*, 481(4):5451–5469.

Shahidi, A., Mobasher, B., Nayyeri, H., Hemmati, S., Wiklind, T., Chartab, N., Dickinson, M., Finkelstein, S. L., Pacifici, C., Papovich, C., Ferguson, H. C., Fontana, A., Giavalisco, M., Koekemoer, A., Newman, J., Sattari, Z., and Somerville, R. (2020). Selection of Massive Evolved Galaxies at $3 \leq z \leq 4.5$ in the CANDELS Fields. *ApJ*, 897(1):44.

Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., and Pettini, M. (2001). The Rest-Frame Optical Properties of z~=3 Galaxies. *ApJ*, 562(1):95–123.

Shaviv, N. J. and Dekel, A. (2003). Photo-evaporation by thermal winds in dwarf galaxies. *arXiv e-prints*, pages astro–ph/0305527.

Silk, J. (1977). On the fragmentation of cosmic gas clouds. I. The formation of galaxies and the first generation of stars. *ApJ*, 211:638–648.

Silk, J. and Rees, M. J. (1998). Quasars and galaxy formation. *A&A*, 331:L1–L4.

Simet, M., Chartab Soltani, N., Lu, Y., and Mobasher, B. (2019). Comparison of Observed

Galaxy Properties with Semianalytic Model Predictions using Machine Learning. *arXiv e-prints*, page arXiv:1905.08996.

Siudek, M., Małek, K., Scodeggio, M., Garilli, B., Pollo, A., Haines, C. P., Fritz, A., Bolzonella, M., de la Torre, S., Granett, B. R., Guzzo, L., Abbas, U., Adami, C., Bottini, D., Cappi, A., Cucciati, O., De Lucia, G., Davidzon, I., Franzetti, P., Iovino, A., Krywult, J., Le Brun, V., Le Fèvre, O., Maccagni, D., Marchetti, A., Marulli, F., Polletta, M., Tasca, L. A. M., Tojeiro, R., Vergani, D., Zanichelli, A., Arnouts, S., Bel, J., Branchini, E., Ilbert, O., Gargiulo, A., Moscardini, L., Takeuchi, T. T., and Zamorani, G. (2017). The VIMOS Public Extragalactic Redshift Survey (VIPERS). Star formation history of passive red galaxies. *A&A*, 597:A107.

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833–859.

Smolčić, V., Karim, A., Miettinen, O., Novak, M., Magnelli, B., Riechers, D. A., Schinnerer, E., Capak, P., Bondi, M., Ciliegi, P., Aravena, M., Bertoldi, F., Bourke, S., Banfield, J., Carilli, C. L., Civano, F., Ilbert, O., Intema, H. T., Le Fèvre, O., Finoguenov, A., Hallinan, G., Klöckner, H. R., Koekemoer, A., Laigle, C., Masters, D., McCracken, H. J., Mooley, K., Murphy, E., Navarette, F., Salvato, M., Sargent, M., Sheth, K., Toft, S., and Zamorani, G. (2015). Physical properties of z > 4 submillimeter galaxies in the COSMOS field. *A&A*, 576:A127.

Somerville, R. S. and Davé, R. (2015). Physical Models of Galaxy Formation in a Cosmological Framework. *ARA&A*, 53:51–113.

Somerville, R. S., Gilmore, R. C., Primack, J. R., and Domínguez, A. (2012). Galaxy properties from the ultraviolet to the far-infrared: Λ cold dark matter models confront observations. *MNRAS*, 423(3):1992–2015.

Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., and Hernquist, L. (2008). A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *MNRAS*, 391(2):481–506.

Somerville, R. S. and Kolatt, T. S. (1999). How to plant a merger tree. *MNRAS*, 305(1):1–14.

Somerville, R. S., Olsen, C., Yung, L. Y. A., Pacifici, C., Ferguson, H. C., Behroozi, P., Osborne, S., Wechsler, R. H., Pandya, V., Faber, S. M., Primack, J. R., and Dekel, A. (2021). Mock Lightcones and Theory Friendly Catalogs for the CANDELS Survey. *MNRAS*.

Somerville, R. S., Popping, G., and Trager, S. C. (2015). Star formation in semi-analytic galaxy formation models with multiphase gas. *MNRAS*, 453(4):4337–4367.

Somerville, R. S. and Primack, J. R. (1999). Semi-analytic modelling of galaxy formation: the local Universe. *MNRAS*, 310(4):1087–1110.

Somerville, R. S., Primack, J. R., and Faber, S. (2001). The nature of high-redshift galaxies. *Monthly Notices of the Royal Astronomical Society*, 320(4):504–528.

Speagle, J. S., Steinhardt, C. L., Capak, P. L., and Silverman, J. D. (2014). A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6. *ApJS*, 214(2):15.

Spitler, L. R., Straatman, C. M. S., Labbé, I., Glazebrook, K., Tran, K.-V. H., Kacprzak, G. G., Quadri, R. F., Papovich, C., Persson, S. E., van Dokkum, P., Allen, R., Kawinwanichakij, L., Kelson, D. D., McCarthy, P. J., Mehrtens, N., Monson, A. J., Nanayakkara, T., Rees, G., Tilvi, V., and Tomczak, A. R. (2014). Exploring the z = 3-4 Massive Galaxy Population with ZFOURGE: The Prevalence of Dusty and Quiescent Galaxies. *ApJL*, 787(2):L36.

Springel, V., Pakmor, R., Pillepich, A., Weinberger, R., Nelson, D., Hernquist, L., Vogelsberger, M., Genel, S., Torrey, P., Marinacci, F., and Naiman, J. (2018). First results from the IllustrisTNG simulations: matter and galaxy clustering. *MNRAS*, 475(1):676–698.

Städler, N., Stekhoven, D. J., and Bühlmann, P. (2014). Pattern alternating maximization algorithm for missing data in high-dimensional problems. *J. Mach. Learn. Res.*, 15(1):1903–1928.

Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., and Bunker, A. (2010). Keck spectroscopy of faint 3 < z < 7 Lyman break galaxies - I. New constraints on cosmic reionization from the luminosity and redshift-dependent fraction of Lyman α emission. *MNRAS*, 408(3):1628–1648.

Stefanon, M., Yan, H., Mobasher, B., Barro, G., Donley, J. L., Fontana, A., Hemmati, S., Koekemoer, A. M., Lee, B., Lee, S.-K., Nayyeri, H., Peth, M., Pforr, J., Salvato, M., Wiklind, T., Wuyts, S., Ashby, M. L. N., Castellano, M., Conselice, C. J., Cooper, M. C., Cooray, A. R., Dolch, T., Ferguson, H., Galametz, A., Giavalisco, M., Guo, Y., Willner, S. P., Dickinson, M. E., Faber, S. M., Fazio, G. G., Gardner, J. P., Gawiser, E., Grazian, A., Grogin, N. A., Kocevski, D., Koo, D. C., Lee, K.-S., Lucas, R. A., McGrath, E. J., Nandra, K., Newman, J. A., and van der Wel, A. (2017). CANDELS Multi-wavelength Catalogs: Source Identification and Photometry in the CANDELS Extended Groth Strip. *ApJS*, 229(2):32.

Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., and Giavalisco, M. (2003). Lyman Break Galaxies at Redshift z ~3: Survey Description and Full Data Set. *ApJ*, 592(2):728–754.

Steidel, C. C. and Hamilton, D. (1993). Deep Imaging of High Redshift QSO Fields Below the Lyman Limit. II. Number Counts and Colors of Field Galaxies. *AJ*, 105:2017.

Steidel, C. C., Pettini, M., and Hamilton, D. (1995). Lyman Imaging of High-Redshift Galaxies.III.New Observations of Four QSO Fields. *AJ*, 110:2519.

Steinhardt, C. L., Capak, P., Masters, D., and Speagle, J. S. (2016). The Impossibly Early Galaxy Problem. *ApJ*, 824(1):21.

Steinhardt, C. L., Weaver, J. R., Maxfield, J., Davidzon, I., Faisst, A. L., Masters, D., Schemel, M., and Toft, S. (2020). A Method to Distinguish Quiescent and Dusty Star-forming Galaxies with Machine Learning. *ApJ*, 891(2):136.

Stekhoven, D. J. and Bühlmann, P. (2011). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.

Storrie-Lombardi, M. C., Lahav, O., Sodre, L., J., and Storrie-Lombardi, L. J. (1992). Morphological Classification of Galaxies by Artificial Neural Networks. *MNRAS*, 259:8P.

Straatman, C. M. S., Labbé, I., Spitler, L. R., Allen, R., Altieri, B., Brammer, G. B., Dickinson, M., van Dokkum, P., Inami, H., Glazebrook, K., Kacprzak, G. G., Kawinwanichakij, L., Kelson, D. D., McCarthy, P. J., Mehrtens, N., Monson, A., Murphy, D., Papovich, C., Persson, S. E., Quadri, R., Rees, G., Tomczak, A., Tran, K.-V. H., and Tilvi, V. (2014). A Substantial Population of Massive Quiescent Galaxies at z ~4 from ZFOURGE. *ApJL*, 783(1):L14.

Swinbank, A. M., Simpson, J. M., Smail, I., et al. (2013). An ALMA survey of sub-millimetre Galaxies in the Extended Chandra Deep Field South: the far-infrared properties of SMGs. *Monthly Notices of the Royal Astronomical Society*, 438(2):1267–1287.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2018). Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *arXiv e-prints*, page arXiv:1804.06788.

Tanaka, M., Valentino, F., Toft, S., Onodera, M., Shimakawa, R., Ceverino, D., Faisst, A. L., Gallazzi, A., Gómez-Guijarro, C., Kubo, M., Magdis, G. E., Steinhardt, C. L., Stockmann, M., Yabe, K., and Zabl, J. (2019). Stellar Velocity Dispersion of a Massive Quenching Galaxy at z = 4.01. *ApJL*, 885(2):L34.

Taylor, E. N., Franx, M., van Dokkum, P. G., Bell, E. F., Brammer, G. B., Rudnick, G., Wuyts, S., Gawiser, E., Lira, P., Urry, C. M., and Rix, H.-W. (2009). The Rise of Massive Red Galaxies: The Color-Magnitude and Color-Stellar Mass Diagrams for z $_{phot}$ 2 from the Multiwavelength Survey by Yale-Chile. *ApJ*, 694(2):1171–1199.

The EAGLE team (2017). The EAGLE simulations of galaxy formation: Public release of particle data. *arXiv e-prints*, page arXiv:1706.09899.

Thomas, D., Maraston, C., Bender, R., and Mendes de Oliveira, C. (2005). The Epochs of Early-Type Galaxy Formation as a Function of Environment. *ApJ*, 621(2):673–694.

Thoul, A. A. and Weinberg, D. H. (1995). Hydrodynamic Simulations of Galaxy Formation. I. Dissipation and the Maximum Mass of Galaxies. *ApJ*, 442:480.

Tin Kam Ho (1995). Random decision forests. 1:278–282 vol.1.

Tin Kam Ho (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844.

Toft, S., Smolčić, V., Magnelli, B., Karim, A., Zirm, A., Michalowski, M., Capak, P., Sheth, K., Schawinski, K., Krogager, J. K., Wuyts, S., Sanders, D., Man, A. W. S., Lutz, D., Staguhn, J., Berta, S., Mccracken, H., Krpan, J., and Riechers, D. (2014). Submillimeter Galaxies as Progenitors of Compact Quiescent Galaxies. *ApJ*, 782(2):68.

Toft, S., Zabl, J., Richard, J., Gallazzi, A., Zibetti, S., Prescott, M., Grillo, C., Man, A. W. S., Lee, N. Y., Gómez-Guijarro, C., Stockmann, M., Magdis, G., and Steinhardt, C. L. (2017). A massive, dead disk galaxy in the early Universe. *Nature*, 546(7659):510–513.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays . *Bioinformatics*, 17(6):520–525.

Trujillo, I., Conselice, C. J., Bundy, K., Cooper, M. C., Eisenhardt, P., and Ellis, R. S. (2007). Strong size evolution of the most massive galaxies since z $\sim$ 2. *MNRAS*, 382(1):109–120.

Trujillo, I., Feulner, G., Goranova, Y., Hopp, U., Longhetti, M., Saracco, P., Bender, R., Braito, V., Della Ceca, R., Drory, N., Mannucci, F., and Severgnini, P. (2006). Extremely compact massive galaxies at z ~1.4. *MNRAS*, 373(1):L36–L40.

Trujillo-Gomez, S., Klypin, A., Primack, J., and Romanowsky, A. J. (2011). Galaxies in ΛCDM with Halo Abundance Matching: Luminosity-Velocity Relation, Baryonic Mass-Velocity Relation, Velocity Function, and Clustering. *ApJ*, 742(1):16.

Vale, A. and Ostriker, J. P. (2004). Linking halo mass to galaxy luminosity. *MNRAS*, 353(1):189–200.

Valentino, F., Tanaka, M., Davidzon, I., Toft, S., Gómez-Guijarro, C., Stockmann, M., Onodera, M., Brammer, G., Ceverino, D., Faisst, A. L., Gallazzi, A., Hayward, C. C., Ilbert, O., Kubo, M., Magdis, G. E., Selsing, J., Shimakawa, R., Sparre, M., Steinhardt, C., Yabe, K., and Zabl, J. (2020). Quiescent Galaxies 1.5 Billion Years after the Big Bang and Their Progenitors. *ApJ*, 889(2):93.

van Albada, G. B. (1960). Formation and evolution of clusters of galaxies (Errata: 15 330). *BAN*, 15:165.

van Albada, G. B. (1961). Evolution of clusters of galaxies under gravitational forces. *AJ*, 66:590.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67.

Van Den Bosch, F. C., Aquino, D., Yang, X., Mo, H. J., Pasquali, A., McIntosh, D. H., Weinmann, S. M., and Kang, X. (2008). The importance of satellite quenching for the build-up of the red sequence of present-day galaxies. *Monthly Notices of the Royal Astronomical Society*, 387(1):79–91.

van der Burg, R. F. J., Rudnick, G., Balogh, M. L., Muzzin, A., Lidman, C., Old, L. J., Shipley, H., Gilbank, D., McGee, S., Biviano, A., Cerulo, P., Chan, J. C. C., Cooper, M., De Lucia, G., Demarco, R., Forrest, B., Gwyn, S., Jablonka, P., Kukstas, E., Marchesini, D., Nantais, J., Noble, A., Pintos-Castro, I., Poggianti, B., Reeves, A. M. M., Stefanon, M., Vulcani, B., Webb, K., Wilson, G., Yee, H., and Zaritsky, D. (2020). The GOGREEN Survey: A deep stellar mass function of cluster galaxies at $1.0 < z < 1.4$ and the complex nature of satellite quenching. *A&A*, 638:A112.

van der Wel, A., Franx, M., van Dokkum, P. G., Skelton, R. E., Momcheva, I. G., Whitaker, K. E., Brammer, G. B., Bell, E. F., Rix, H. W., Wuyts, S., Ferguson, H. C., Holden, B. P., Barro, G., Koekemoer, A. M., Chang, Y.-Y., McGrath, E. J., Häussler, B., Dekel, A., Behroozi, P., Fumagalli, M., Leja, J., Lundgren, B. F., Maseda, M. V., Nelson, E. J., Wake, D. A., Patel, S. G., Labbé, I., Faber, S. M., Grogin, N. A., and Kocevski, D. D. (2014). 3D-HST+CANDELS: The Evolution of the Galaxy Size-Mass Distribution since $z = 3$. *ApJ*, 788(1):28.

van Dokkum, P. G., Förster Schreiber, N. M., Franx, M., Daddi, E., Illingworth, G. D., Labbé, I., Moorwood, A., Rix, H.-W., Röttgering, H., Rudnick, G., van der Wel, A., van der Werf, P., and van Starkenburg, L. (2003). Spectroscopic Confirmation of a Substantial Population of Luminous Red Galaxies at Redshifts z>~2. *ApJL*, 587(2):L83–L87.

van Dokkum, P. G., Franx, M., Kriek, M., Holden, B., Illingworth, G. D., Magee, D., Bouwens, R., Marchesini, D., Quadri, R., Rudnick, G., Taylor, E. N., and Toft, S. (2008). Confirmation of the Remarkable Compactness of Massive Quiescent Galaxies at z ~2.3: Early-Type Galaxies Did not Form in a Simple Monolithic Collapse. *ApJL*, 677(1):L5.

Vanzella, E., Cristiani, S., Fontana, A., Nonino, M., Arnouts, S., Giallongo, E., Grazian, A., Fasano, G., Popesso, P., Saracco, P., and Zaggia, S. (2004). Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS. *A&A*, 423:761–776.

Vapnik, V. (1992). Principles of risk minimization for learning theory. In *Advances in neural information processing systems*, pages 831–838.

Villmann, T. (1999). - topology preservation in self-organizing maps. In Oja, E. and Kaski, S., editors, *Kohonen Maps*, pages 279 – 292. Elsevier Science B.V., Amsterdam.

Wattenberg, M., Viégas, F., and Johnson, I. (2016). How to use t-sne effectively. *Distill*.

Wechsler, R. H. and Tinker, J. L. (2018). The connection between galaxies and their dark matter halos. *Annual Review of Astronomy and Astrophysics*, 56(1):435–487.

Weigel, A. K., Schawinski, K., and Bruderer, C. (2016). Stellar mass functions: methods, systematics and results for the local Universe. *MNRAS*, 459(2):2150–2187.

Whitaker, K. E., Franx, M., Leja, J., van Dokkum, P. G., Henry, A., Skelton, R. E., Fumagalli, M., Momcheva, I. G., Brammer, G. B., Labbé, I., Nelson, E. J., and Rigby, J. R. (2014a). Constraining the Low-mass Slope of the Star Formation Sequence at $0.5 < z < 2.5$. *ApJ*, 795(2):104.

Whitaker, K. E., Franx, M., Leja, J., van Dokkum, P. G., Henry, A., Skelton, R. E., Fumagalli, M., Momcheva, I. G., Brammer, G. B., Labbé, I., Nelson, E. J., and Rigby, J. R. (2014b). Constraining the Low-mass Slope of the Star Formation Sequence at 0.5 z 2.5. *ApJ*, 795:104.

Whitaker, K. E., Kriek, M., van Dokkum, P. G., Bezanson, R., Brammer, G., Franx, M., and Labbé, I. (2012a). A Large Population of Massive Compact Post-starburst Galaxies at z > 1: Implications for the Size Evolution and Quenching Mechanism of Quiescent Galaxies. *ApJ*, 745(2):179.

Whitaker, K. E., van Dokkum, P. G., Brammer, G., and Franx, M. (2012b). The Star Formation Mass Sequence Out to $z = 2.5$. *ApJL*, 754:L29.

Whitaker, K. E., van Dokkum, P. G., Brammer, G., Momcheva, I. G., Skelton, R., Franx, M., Kriek, M., Labbé, I., Fumagalli, M., Lundgren, B. F., Nelson, E. J., Patel, S. G., and Rix, H.-W. (2013). Quiescent Galaxies in the 3D-HST Survey: Spectroscopic Confirmation of a Large Number of Galaxies with Relatively Old Stellar Populations at z ~2. *ApJL*, 770(2):L39.

Whitaker, K. E., Williams, C. C., Mowla, L., Spilker, J. S., Toft, S., Narayanan, D., Pope, A., Magdis, G. E., van Dokkum, P. G., Akhshik, M., Bezanson, R., Brammer, G. B., Leja, J., Man, A., Nelson, E. J., Richard, J., Pacifici, C., Sharon, K., and Valentino, F. (2021). Quenching of star formation from a lack of inflowing gas to galaxies. *Nature*, 597(7877):485–488.

White, S. D. M. (1976). The dynamics of rich clusters of galaxies. *MNRAS*, 177:717–733.

White, S. D. M. and Frenk, C. S. (1991). Galaxy Formation through Hierarchical Clustering. *ApJ*, 379:52.

White, S. D. M. and Rees, M. J. (1978). Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358.

Wijesinghe, D. B., Hopkins, A. M., Brough, S., Taylor, E. N., Norberg, P., Bauer, A., Brown, M. J. I., Cameron, E., Conselice, C. J., Croom, S., Driver, S., Grootes, M. W., Jones, D. H., Kelvin, L., Loveday, J., Pimbblet, K. A., Popescu, C. C., Prescott, M., Sharp, R., Baldry, I., Sadler, E. M., Liske, J., Robotham, A. S. G., Bamford, S., Bland-Hawthorn, J., Gunawardhana, M., Meyer, M., Parkinson, H., Drinkwater, M. J., Peacock, J., and Tuffs, R. (2012). Galaxy And Mass Assembly (GAMA): galaxy environments and star formation rate variations. *MNRAS*, 423(4):3679–3691.

Wiklind, T., Dickinson, M., Ferguson, H. C., Giavalisco, M., Mobasher, B., Grogin, N. A., and Panagia, N. (2008). A Population of Massive and Evolved Galaxies at z gtrsim 5. *ApJ*, 676(2):781–806.

Wiklind, T., Mobasher, B., Dickinson, M., Ferguson, H., Giavalisco, M., Grogin, N., and Panagia, N. (2007). Massive and Evolved Galaxies at z >= 5. In Combes, F. and Palouš, J., editors, *Galaxy Evolution across the Hubble Time*, volume 235 of *IAU Symposium*, pages 368–372.

Wild, V., Almaini, O., Dunlop, J., Simpson, C., Rowlands, K., Bowler, R., Maltby, D., and McLure, R. (2016). The evolution of post-starburst galaxies from z=2 to 0.5. *MNRAS*, 463(1):832–844.

Wild, V., Taj Aldeen, L., Carnall, A., Maltby, D., Almaini, O., Werle, A., Wilkinson, A., Rowlands, K., Bolzonella, M., Castellano, M., Garguilo, A., McLure, R., Pentericci, L., and Pozzetti, L. (2020). The star formation histories of z~1 post-starburst galaxies. *arXiv e-prints*, page arXiv:2001.09154.

Williams, C. K. I. and Seeger, M. (2001). Using the nyström method to speed up kernel machines. pages 682–688.

Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., and Labbé, I. (2009). Detection of Quiescent Galaxies in a Bicolor Sequence from Z = 0-2. *ApJ*, 691(2):1879–1895.

Wolf, C., Meisenheimer, K., Rix, H. W., Borch, A., Dye, S., and Kleinheinrich, M. (2003). The COMBO-17 survey: Evolution of the galaxy luminosity function from 25 000 galaxies with 0.2< z <1.2. *A&A*, 401:73–98.

Wuyts, S., Förster Schreiber, N. M., Lutz, D., Nordon, R., Berta, S., Altieri, B., Andreani, P., Aussel, H., Bongiovanni, A., Cepa, J., Cimatti, A., Daddi, E., Elbaz, D., Genzel, R., Koekemoer, A. M., Magnelli, B., Maiolino, R., McGrath, E. J., Pérez García, A., Poglitsch, A., Popesso, P., Pozzi, F., Sanchez-Portal, M., Sturm, E., Tacconi, L., and Valtchanov, I. (2011). On Star Formation Rates and Star Formation Histories of Galaxies Out to z ˜ 3. *ApJ*, 738:106.

Wuyts, S., Franx, M., Cox, T. J., Hernquist, L., Hopkins, P. F., Robertson, B. E., and van Dokkum, P. G. (2009). Recovering Stellar Population Properties and Redshifts from Broadband Photometry of Simulated Galaxies: Lessons for SED Modeling. *ApJ*, 696(1):348–369.

Wuyts, S., Labbé, I., Förster Schreiber, N. M., Franx, M., Rudnick, G., Brammer, G. B., and van Dokkum, P. G. (2008). FIREWORKS U_{38}-to-24 μm Photometry of the GOODS Chandra Deep Field-South: Multiwavelength Catalog and Total Infrared Properties of Distant K_s-selected Galaxies. *ApJ*, 682(2):985–1003.

Wuyts, S., Labbé, I., Franx, M., Rudnick, G., van Dokkum, P. G., Fazio, G. G., Förster Schreiber, N. M., Huang, J., Moorwood, A. F. M., Rix, H.-W., Röttgering, H., and van der Werf, P. (2007). What Do We Learn from IRAC Observations of Galaxies at 2 < z < 3.5? *ApJ*, 655(1):51–65.

Xue, Y. Q., Luo, B., Brandt, W. N., Alexander, D. M., Bauer, F. E., Lehmer, B. D., and Yang, G. (2016). The 2 Ms Chandra Deep Field-North Survey and the 250 ks Extended Chandra Deep Field-South Survey: Improved Point-source Catalogs. *ApJS*, 224(2):15.

Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Popping, G., and Davé, R. (2019a). Semi-analytic forecasts for JWST - I. UV luminosity functions at z = 4-10. *MNRAS*, 483(3):2983–3006.

Yung, L. Y. A., Somerville, R. S., Popping, G., Finkelstein, S. L., Ferguson, H. C., and Davé, R. (2019b). Semi-analytic forecasts for JWST - II. Physical properties and scaling relations for galaxies at z = 4-10. *MNRAS*, 490(2):2855–2879.

Zhang, H. (2004). The optimality of naive bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2004*, 2.