Rijksuniversiteit Groningen

Bachelor Thesis

The Estimation of Kinematic Parameters Through Photometric Analysis

Using TNG50 Simulated Galaxies for a Linear Regression Model

Abstract

This thesis aims to create a method as to estimate kinematic properties of galaxies, like the fraction of stars within a galaxy that is rotating in bulk rotation, through using photometric parameters, forfeiting spectroscopic analysis. To achieve this 1160 simulated galaxies from the IllustrisTNG TNG50 volume were used, which were displayed on a cutout with a resolution of 1600 by 1600 pixels. From these the data catalogues from TNG50, visual classification through manual inspection and morphological parameters estimation were used to compare these parameters to each other. Within these parameters it was found that visual classification has a higher prevalence of red spiral galaxies than through kinematic classification. Furthermore the Gini coefficient and M_{20} statistic did not seem to distinguish between visually classified galaxies. From these parameters a linear regression model was created as to determine the disk fraction f_d through photometric input parameters, resulting with a score of determination of 0.4444, mean squared error of 0.02250 and mean absolute deviation of 0.1242.

Author:
J.H. Bonhof
S4045742

Supervisors:
L. Wang
A. La Marca

Abstract

Contents

1 Introduction

One of the most iconic objects in the night sky are the innumerable galaxies found within the universe. Ranging from opaque spheroids to flat disk-like galaxies with long winding arms, these superstructures of space form a core part of astronomy. Galaxies themselves are composed out of many different objects, together forming large scale structures that help shape the galaxy into a visually distinct shape. Within them can be found countless amounts of stars, dust and gas, which all add up to the total mass of the galaxy. Two distinct shapes that can be seen are early- and late-type galaxies, displayed in figure 1 and 2 respectively. There is the third group in which galaxies can be placed, that being lenticular, but for this thesis the groups will be limited to early- and late-type.

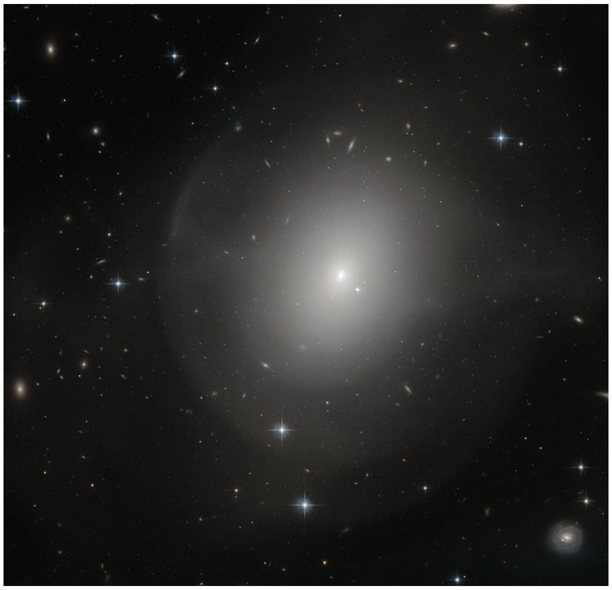

Figure 1: Early-type galaxy NGC 2865, acquired from the Hubble Space Telescope by ESA/Hubble and NASA.

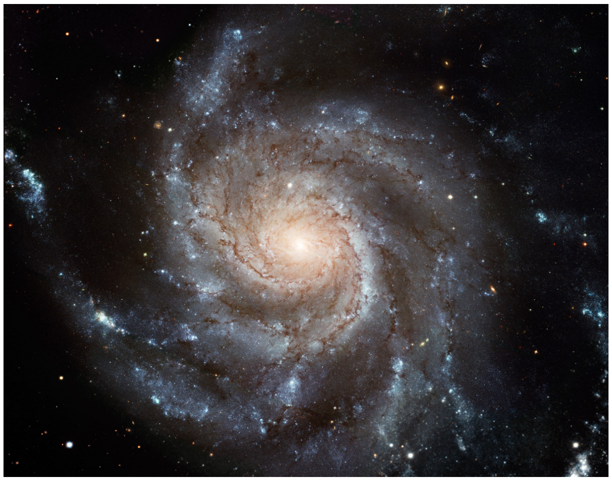

Figure 2: Late-type galaxy M101, acquired from the Hubble Space Telescope by ESA and NASA.

There are multiple properties of galaxies that can be both observed and calculated. These properties can be internal, influenced by the stellar population and gas distribution, while other external properties rely on the large scale structure of the galaxy. These properties can be divided into three different categories to help us distinguish galaxies from each other in more depth. The first category is the morphology of the galaxy, that being the study of separating galaxies based on their shape. This separation can be further simplified to the distinction between early-type galaxies that have a diffuse spheroid-like shape and galaxies that have spiral-like arms (also often referred to as late-type galaxies) (Lintott et al. 2008). These separation are for example used in weak gravitational lensing studies, where the strength of said lensing computed to the mass of the galaxy combined with the classification of the galaxy helps further the understanding of galaxy evolution over time (Mandelbaum et al. 2006). Early-type galaxies in relation to the universe's timescale within morphological context will be referred to as elliptical galaxies for the sake of simplicity and late-type galaxies as spiral galaxies.

The second category that is commonly used is through kinematic classification. Within this specific field galaxies are divided between being supported mainly through either rotation or dispersion.

Lastly there is the third category, colours or visual data. The morphological classification of galaxies has strong correlation to colours, where spiral galaxies tend to be more 'blue' whereas elliptical galaxies tend to be more 'red'.

1.1 Galaxy Development

When it comes to galaxies over a span of time, there is a good reason as to why early-type galaxies are called as such. During the early stages of the universe these galaxies formed more often (van der Wel et al. 2005), where nowadays they are a smaller percentage of the total universe count, with them totalling to approximately 12% of the total galaxies within our local supercluster (Loveday 1996). Compared to the more recent spiral galaxies, these elliptical galaxies have a far lower star formation rate, with their stellar population consisting largely out of red stars with a long lifespan. Within this it can be observed as in figure 3 that these galaxies have a far greater star formation rate in the past, showing that they were more actively developing galaxies back then than current day (Naab et al. 2007).

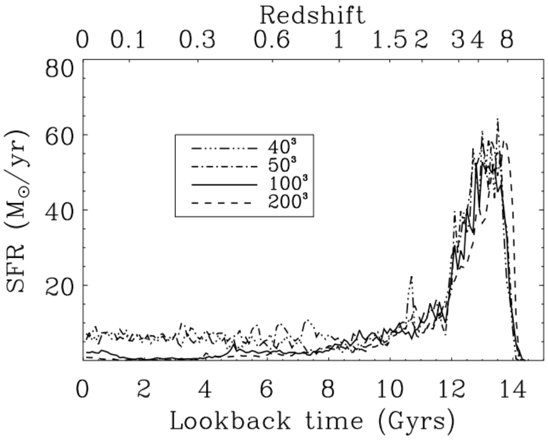

Figure 3: Star formation rate of a simulated elliptical galaxy over time in the past (Naab et al. 2007)

When it comes to spiral galaxies, which can also be referred to as late-type galaxies, one major indicator for why they are more recent is from the increased star formation rate and young blue stars that are commonly observed within their systems. In the research paper of Melvin et al. 2014, it was found that through numerous different observation methods, late-type galaxies (with as main focus the barred late-type galaxies) form a smaller fraction of the total galaxy population when looking at greater redshifts. This can be further seen in figure 4. These spiral galaxies can subsequently be divided between non-barred and barred galaxies, where barred galaxies are found to be late creations when the galaxies are sufficiently gravitationally stable and experience friction from surrounding dark matter (Kraljic, Bournaud, and Martig 2012). These differences between the two galaxies can be further observed within the morphology of the galaxies, which allows for a clear divide of the different galaxy groups.

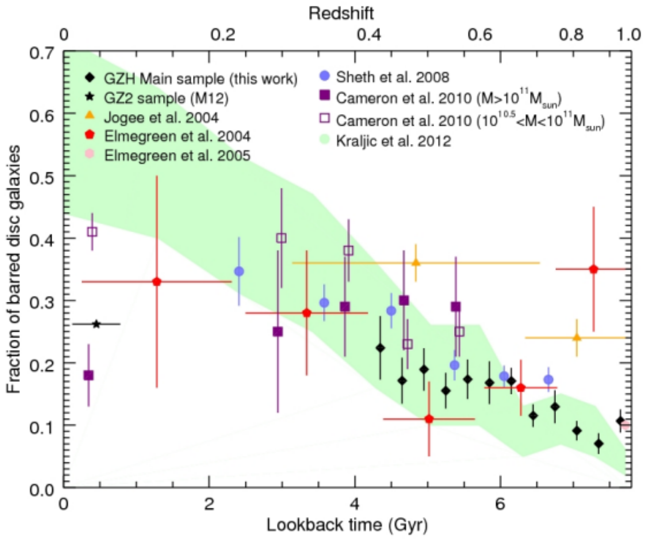

Figure 4: Fraction of barred disk galaxies plotted over time in the past, comparing different results from papers. In green a simulated plot of bar fraction is displayed. (Melvin et al. 2014)

1.2 Morphology

Morphology dates back to 1926, where Hubble published his own classification of the morphology of different galaxies (Hubble 1926). These morphologies were divided across 3 groups: spiral (which itself is divided between normal and barred galaxies), lenticular and elliptical. While being relatively rudimentary, Hubble was the first to classify galaxies through visual images of them, and the creation of the tuning fork classification that followed is still used to this day. Originally Hubble hypothesised that elliptical galaxies were early-stage galaxies that later on settled into a shape of spiral galaxies, but this turned out to be incorrect as both types of galaxies are created independent of each other (Mo, van den Bosch, and White 2010). There is however an apparent gradient to the morphology of the galaxies, with early-type galaxies ranging from elliptical to lenticular galaxies. Furthermore the 'arms' of the spiral galaxies can range from nearly imperceptible to incredibly distinct. Even more so, spiral galaxies can have a visible 'bar' in the centre and therefore belong to a distinct branch from non-bar galaxies which creates the tuning fork. In figure 5 these 3 different groups are displayed with the tuning fork division of it.

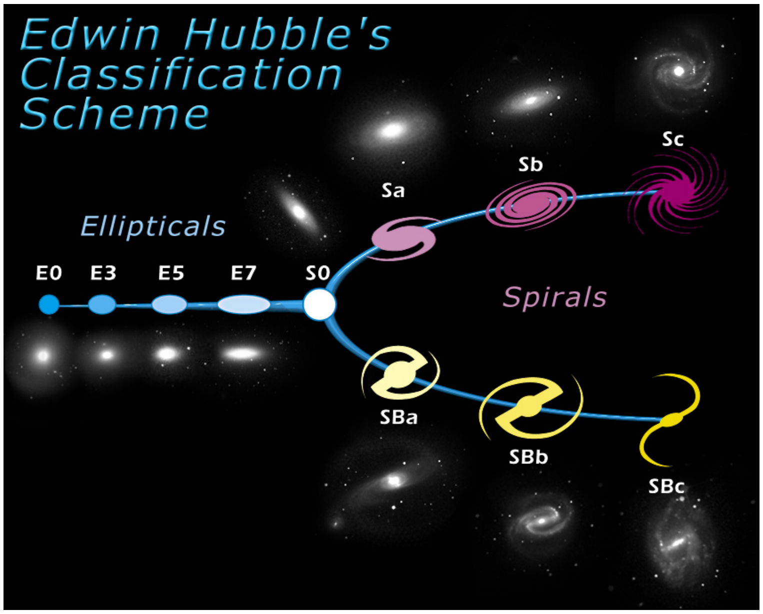

Figure 5: The Hubble Tuning Fork. Note the difference between barred (SB galaxies) and non-barred (S galaxies) visually.

For the sake of consistency in this project, a binary between spheroid galaxies and spiral galaxies is set up, merging barred and non-barred galaxies to disk (or spiral) galaxies. This allows for a clearer comparison between visual and kinematic classification of the galaxies. There are of course galaxies that do not fit any typical classification like elliptical and spiral galaxies. These galaxies often occur due to external masses disrupting the structure of the galaxy, or they even have these masses merge with the main body. These are classified under mergers and irregular galaxies and are only commonly detected through their atypical shape.

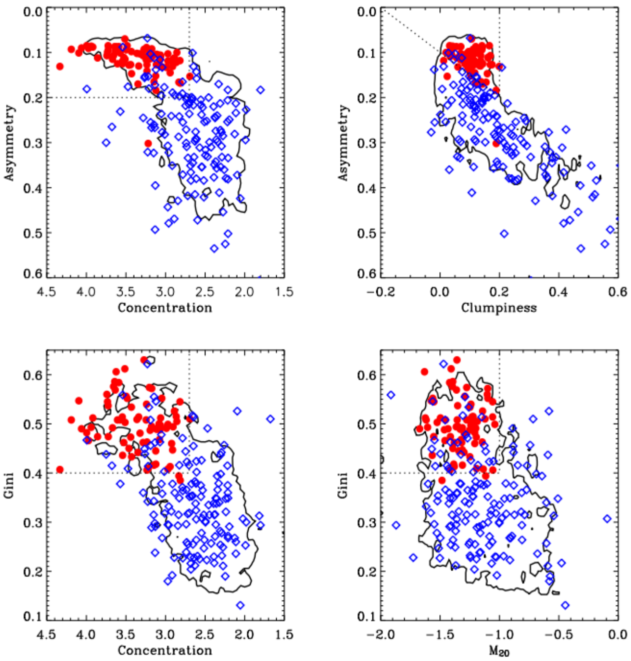

Figure 6: Visually classified galaxies plotted with different morphological parameters. Red data points represent early-type galaxies and blue data points late-type galaxies, classified on being disk dominated or spheroid dominated within observed images. The contour within the plots resembles the area where within 90% of the observed galaxies are located. (Cassata et al. 2007)

Beyond the visual shape of the galaxies, there are also morphological parameters that can be determined from the image data. These parameters are a central part in studies on the morphology of galaxies, as certain parameters like the Concentration, Asymmetry and Smoothness (often also called clumpiness as a high index indicates a galaxy with a high prevalence of small-scale structures) are used to distinguish between early-type and late-type galaxies. In the research of Cassata et al. 2007, these parameters are use of different galaxies based on observations within the Cosmic Evolution Survey (COSMOS)(Scoville et al. 2007). The galaxies within said research are taken at a redshift of $z \sim 0.7$, where they have been plotted within morphological parameters in figure 6. As can be seen in this figure, early-type galaxies are clustered very distinctly when compared to late-type, which is especially visible for the plots of the Concentration, Asymmetry and Smoothness (CAS) parameters. The other 2 visible parameters are the Gini coefficient and M_{20} statistic, where the Gini coefficient indicates how concentrated a galaxy's light distribution is (Genel et al. 2015) and the M_{20} statistic is the second moment of the galaxy for where 20% of the galaxy's flux is located. This in turn can also be described as a statistic for how far the pixel with high flux are from the centre of the galaxy (Genel et al. 2015). These are elaborated upon more in section 3.2 and form a core component of the project.

1.3 Kinematic

Within the category of kinematics, galaxies are separated commonly on the binary between rotation dominated and dispersion dominated galaxies. Rotation dominated galaxies are commonly associated with disk galaxies, as the shared rotation in those galaxies is far greater than the random motion (dispersion). This is due to high prevalence of dust and gas within the galaxy (Holwerda et al. 2019). On the other hand, for elliptical galaxies there is commonly a greater amount of dispersion than rotation present, as there is very little gas or dust present within the galaxy (Georgakakis et al. 2001). A cause for this change in gas content within the galaxy can be attributed to the stronger star formation rate found when looking at galaxies at high redshift, as seen in figure 3.

While dispersion dominated galaxies do evolve distinctly from rotation dominated galaxies, it is hypothesised that rotation dominated galaxies can evolve into the dispersion dominated ones through merging with other galaxies (Hopkins et al. 2008). In these cases the interacting of multiple galaxies with different initial rotational velocities mix, causing a more random distribution of velocity to occur.

1.4 Colours

The division of galaxies based on their colours can be explained due to elliptical galaxies possessing a higher amount of red stars, with a low star formation rate. This causes a typically red colouration of the galaxy, while for spiral galaxies there is a higher star formation rate due to a larger concentration of dust and gas within the arms (Hopkins et al. 2008) (Cassata et al. 2007). This in turn causes newer, blue stars to be created giving the galaxy overall a more blue hue. Furthermore, due to light being scattered through dust present within the spiral galaxies there is a higher prevalence of higher wavelength light directly from the emission source, while shorter wavelengths get scattered through the dust into different directions. The different distributions of galaxies based on different samples and colour can be seen in figure 7, where it can be seen that late-type galaxies are more prevalent in the blue colour range while for early-type they are more found in more red colour spectra.

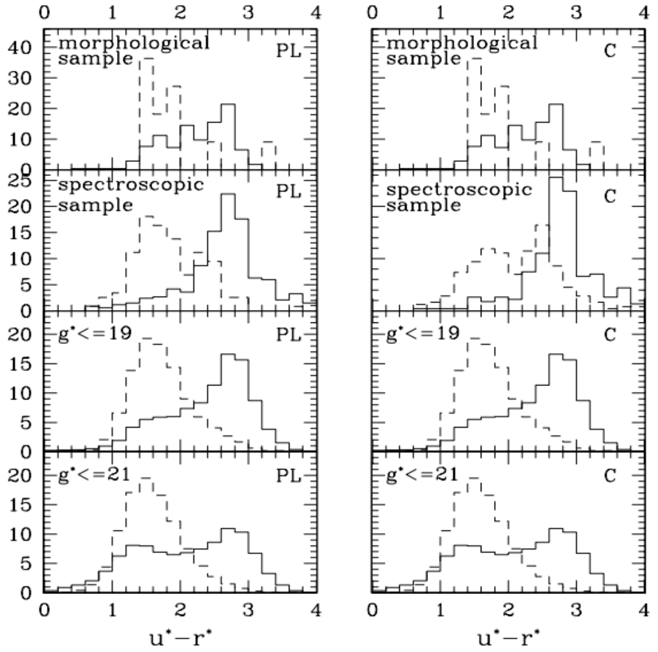

Figure 7: Colour histograms of early-type and late-type galaxies from the measurements of Strateva et al. 2001, indicated through solid and dashed lines respectively. Early- and late-type galaxies on the left side are separated through profile likelihood and right side through the Concentration index.

1.5 This Project

This project has as the main focus the subject of exploring a method as to determine kinematic parameters of galaxies based on photometric input. To this the question is posed: Is it possible to classify galaxies kinematically through purely morphological and colour properties? For this project a linear regression model is created using simulated galaxies as training data, where the end goal is to use several input parameters of morphology and colour and return a kinematic parameter that shows if a galaxy is supported through rotation or dispersion. This model would in turn allow other researchers in the future determine the kinematic classification of galaxies using photometric parameters instead of time-consuming spectroscopic observation and analysis. This model can then be used to build upon it as well, making it more accurate over time and predict other parameters based on what the program is trained on.

In section 2 the data will be elaborated on, explaining the parameters and specifications of the simulation. Section 3 will detail the computational steps to go from visual data and initial parameters to acquire the morphological parameters. Lastly section 4 will showcase the computed data and linear regression model, with section 5 discussing prospects of the model.

2 Data

In this section the data of the thesis project is discussed, along with the specification of IllustrisTNG from which the data is acquired. Lastly the kinematic and filter parameters of the simulated data are explained and which parameters are of specific interest to the computation.

2.1 IllustrisTNG

For this research project galaxy simulations were used from IllustrisTNG (TNG50), and compared to measurements from the LSST telescope. TNG50 is a hydrodynamic baryonic simulation of the universe, ranging from moments after the big bang to the present day. IllustrisTNG consists out of 3 different volumes, TNG50, TNG100 and TNG300. The distinction between these three volumes is because of the cubic volume in which each simulation occurs, where the number dictates the length of the sides of the cubic volume in Mpc. For this project TNG50 is used with all galaxies taken at present day with redshift z=0, as the major focus is on the properties of each individual galaxy and not clusters as a whole. TNG50 in turn also gives a far greater resolution, being more than 100 times as high in mass resolution than TNG300 according to their statistics (Nelson et al. 2019). These different volumes of IllustrisTNG are displayed in figure 8. With multiple different methods to view the galaxies (e.g. gas temperature, dark matter density), IllustrisTNG can be a useful tool in examining phenomena in cosmology at a greater resolution with far less noise than through real observations.

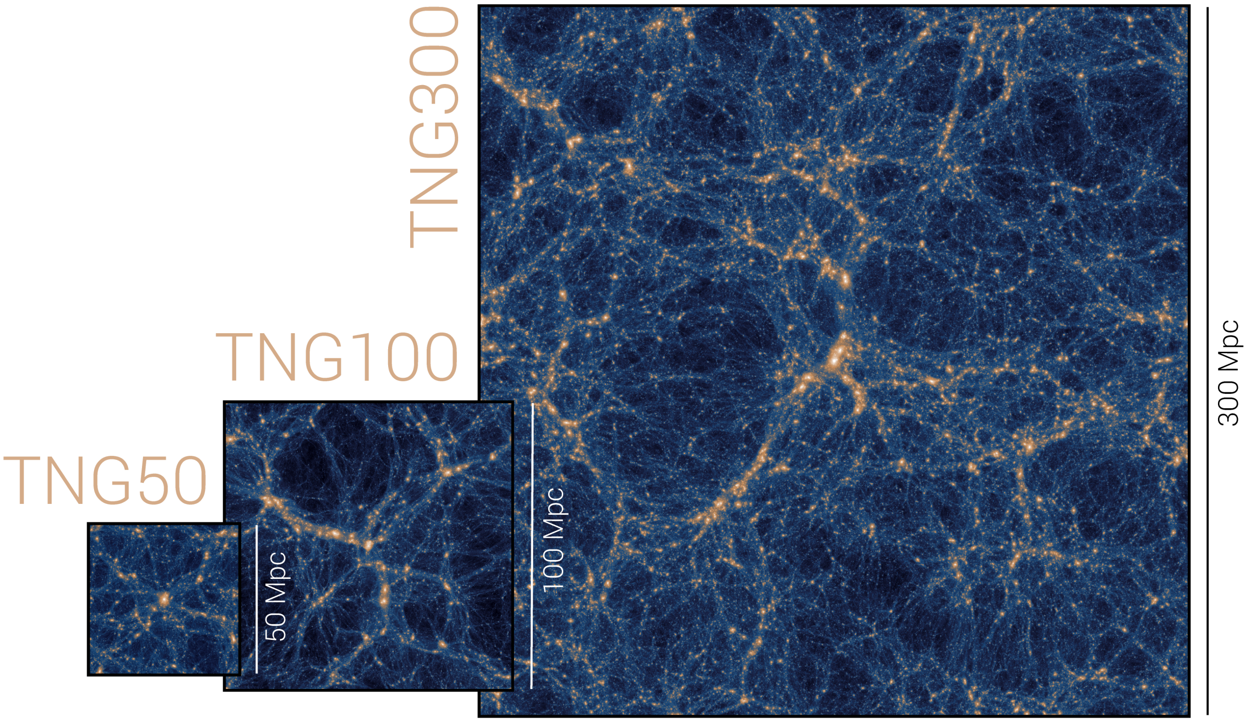

Figure 8: The 3 different volumes of IllustrisTNG, displayed at roughly their respective sizes (not exactly however). Each image displays the dark matter density throughout the entire volume (Nelson et al. 2019).

For the end goal of the research project, that being a trained linear regression model for galaxies based on their kinematic properties from morphological and visual data, 1160 simulated galaxies are used from TNG50 with a cutout section containing the galaxy of 1600 by 1600 pixels at a calculated distance of 206 Mpc distance. The resolution of the cutout is to match the Euclid VIS resolution which is used as a basis for the resolution metrics. Each of these galaxy images use the Euclid VIS band, with Monte Carlo noise. Absorption and scattering of interstellar dust are included, but dust emission is not. All galaxies are within the mass range of 6×10^9 to 1×10^{12} solar masses at redshift z=0. Of the images each pixel is the size of 0.1 arcsecond using the statistics acquired from Euclid VIS (ESA, Euclid VIS Instrument 2019). While multiple viewing angles are available of each galaxy, only 1 angle per galaxy is used for the model. The subhalo ID is given for each galaxy in order to be able to call upon specific galaxies and merge table such that each galaxy is allocated the correct parameters.

2.1.1 Kinematic Parameters

From TNG50 a few kinematic and filter parameters of each galaxy were given. The kinematic parameters from this catalogue that are relevant for the model are the specific angular momentum of the stars J_Z and the circularity parameter ϵ. The specific angular momentum is derived from the angular momentum with respect to the mass of the stellar objects, which are both available from the input data of the galaxy. For the calculations of this, the galaxy is aligned firstly such that the z-axis of it is aligned with the angular momentum vector, after which the angular momentum of the stars in the system is taken from within 10 times the radius from wherein half the total mass of the galaxy is contained. After this the angular momentum is divided by the mass of the stars through $J_Z = J/M_*$ where J_Z is the specific angular momentum and M_* is the stellar mass. Thus this way the specific angular momentum is derived for each stellar particle (Genel et al. 2015).

Following this the circularity parameter ϵ, the fraction of specific angular momentum for a star to the maximum angular momentum, is calculated by using the aligned galaxies and applying the following formula:

$$\epsilon = \frac{J_Z}{J(E)} \tag{1}$$

Where J_Z is the specific angular momentum of each stellar particle and $J(E)$ is the maximum angular momentum of stellar particles within a range of 50 particles above and below the stellar particle in question. Here the order in which these particles are arranged is through their binding energy, defined by $U_{grav} + v^2$ where U_{grav} is the gravitational binding energy and v is the velocity of the stellar particles. From this circularity parameter it is then possible to calculate the fractional mass of stars with $\epsilon > 0.7$, the fractional mass of stars with $\epsilon > 0.7$ minus the fractional mass of $\epsilon < -0.7$ and the fractional mass with $\epsilon < 0$ times two. These circularity parameters are commonly used to kinematically classify galaxies based on the value they return. For dispersion supported galaxies an ϵ is expected of below 0, with rotationally supported galaxies an ϵ is expected of above 0.7. Often the ϵ below -0.7 is subtracted from the rotationally supported galaxies as to rule out stars that are commonly found in the bulge that would contribute to a dispersion supported galaxy (Nelson et al. 2019). Within observations these fractions derived from the circularity parameters are used, where the fraction of ϵ above 0.7 minus ϵ below -0.7 is the disk fraction f_d (thus the fraction of stars that follow the bulk rotation). This disk fraction is then used to kinematically classify the observed galaxies.

2.1.2 Filter Parameters

The catalogue provided for the galaxies lastly has 8 columns for the filters griz and UBVK. These filters are measured in AB magnitudes (the spectral flux densities of the filter) at rest frame and are used to determine the colours of the galaxies. The colours here are determined by taken 2 different magnitudes observed in the bands and subtracting them from each other in order to return the flux ratio between the 2 bands. The g, r, i and z filters are based on the Sloan Digital Sky Survey filters, measuring wavelengths at $\lambda = 0.469\mu m$, $\lambda = 0.617\mu m$, $\lambda = 0.748\mu m$ and $\lambda = 0.893\mu m$ respectively. For the U, B, V and K filters the Johnson-Bessel filters' wavelengths are used of $\lambda = 0.36\mu m$, $\lambda = 0.435\mu m$, $\lambda = 0.55\mu m$ and $\lambda = 2.22\mu m$ respectively. For this research the g, r, i, B, V and K filters were used, as the U filter is on nearly the same wavelength as the u filter, differing only by $0.005\mu m$, and the z filter within the simulated runs had its curve extend beyond range of allowed wavelengths (Nelson et al. 2019). In order to proceed from filter magnitudes to colours, the flux ratio is determined through the subtraction from 2 filters, with the longer wavelength filter commonly being subtracted from the shorter one (e.g. g-r, B-V). For this thesis the colours g-r, B-V and g-K are used, as they are commonly used within photometry and display large differences between galaxies and their colours.

3 Method

In this section the visual classification and morphological parameter estimation are explained, along with how the data from the previous section is convolved for the visual inspection. Lastly, the morphological parameters acquired are explained in depth, ranging from their calculation to their significance to the galaxy's properties.

3.1 Visual Classification

Starting with the received image data from TNG50, the images are not yet convolved with the appropriate PSF. In order to handle each galaxy fits file, the astropy module is used along with pandas for table management. Using the arcseconds per pixels of EUCLID VIS from ESA (ESA, Euclid VIS Instrument 2019), the next step is determining the standard deviation of the PSF based on the PSF's given FWHM. Using a FWHM of 0.16", the standard deviation is approximated through $\sigma_{PSF} = FWHM/2.355$. From this the PSF is calculated using the following formula on a 40x40 pixel grid:

$$PSF = exp(-\frac{x^2+y^2}{2\sigma_{PSF}^2}) \tag{2}$$

Using equation 2, each individual galaxy is convolved with the PSF and stored in a new fits file. The convolution itself uses the cupy module with the convolve2d method to speed up convolution through using a GPU instead of CPU. In order for this to work the image data first has to be converted to a cupy array for convolution, and after convolution converted back into a numpy array in order for it to be able to be stored within the fits files.

Each galaxy is visually inspected through both an image visualisation software called DS9 and image plots using matplotlib, both in logarithmic scale. Both methods are used as some of the simulated galaxies had odd 'tails' that were cut off at certain parts, making it difficult to discern them in either DS9 or the matplotlib plots.

Each galaxy image after being visually classified is stored in an array where a value is assigned for each type. The values are 0 for elliptical, 1 for spiral, 2 for mergers/irregular galaxies and 3 for poorly rendered galaxies.

Some example galaxies are given in figure 9 where they are displayed in logarithmic scale using matplotlib's viridis colour map. As can be seen in the 3rd elliptical image and the 4th spiral image, some galaxies were more compact when comparing to the other galaxies. This makes visual classification for these smaller galaxies more difficult than for the larger ones due to a lack of resolution. In the bad render galaxies it can be seem how the galaxies are either 'boxed' in or have maximum brightness to the corner of them. Especially in the 4th bad render galaxy it is visible how this box around the galaxy makes it impossible to accurately classify them. The fits files of all the galaxies are then used in Statmorph for the morphological parameters.

Out of all the galaxies given, there were a few galaxies found that seemed poorly rendered. The poorly rendered galaxies either have very localised boxed areas around the galaxies with values as high as the galaxy or are so small that the galaxy is a similar size to the PSF. In both cases it is not possible to acquire meaningful data or visually classify these galaxies. This also seemed to be unlikely to be the actual size of the galaxies, as the solar mass minimum limit would imply a larger size. The visual classification listed 262 elliptical galaxies, 866 spiral galaxies, 32 mergers/irregular galaxies and 9 poorly rendered galaxies. The 9 total poorly rendered galaxies were removed from the catalogue and image fits files, as to not use

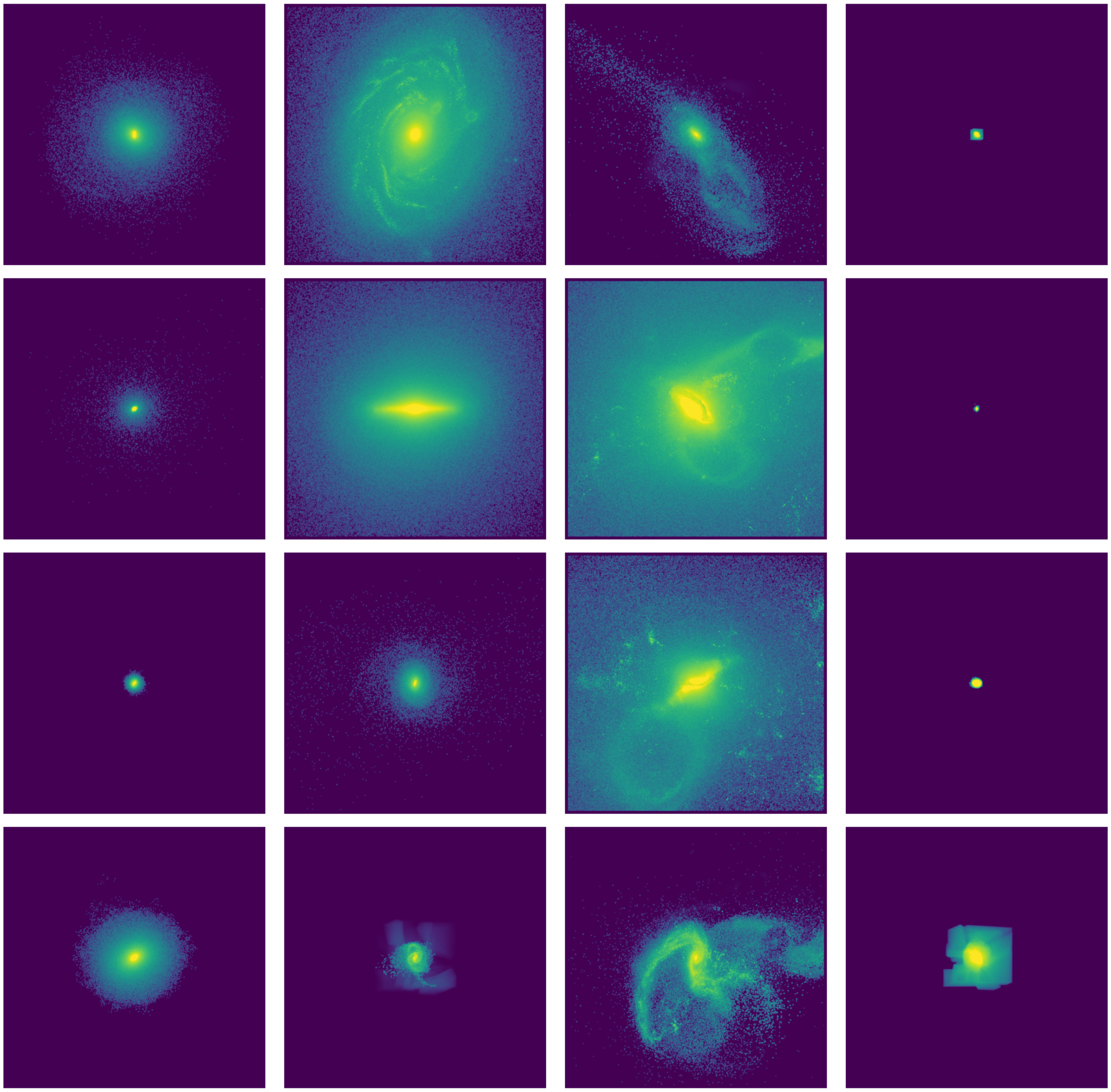

Figure 9: From left to right (columns): Elliptical galaxies, Spiral galaxies, Merger/Irregular galaxies, Bad render galaxies

them as training data for the model. This in turn leaves 1151 galaxies to be used for parameter estimation using Statmorph

For the visual classification it would be advisable to use citizen science or a similarly large group of people to have multiple attempts in the future for classification. In this instance all classification was done by 1 person and thus is prone for a random error in classifying, where a larger group of people would reduce this possible error.

3.2 Statmorph

In order to estimate the remaining morphology parameters necessary of each galaxy, Statmorph is used (Rodriguez-Gomez et al. 2019). For this estimation, Statmorph requires the thresholds and segmentation maps. For the threshold the segmentation.detect_threshold method is used from the photutils module with a standard deviation per pixel of 5. The segmentation maps uses photutils's segmentation.detect_sources method with npixels set to 100, where npixels are the minimum required amount of connected pixels per segment. These parameters gave the highest amount of working computations. Lastly, Statmorph's source_morphology() method is used which returns an object from which all parameters for the morphology can be extracted. This does require a gain or weight map parameter, where in this project a gain of 1 is used as it could not be determined from the simulation images. This does result in the sersic data being incorrectly calculated, but these parameters are not used in the model or remainder of the project. All of these morphological parameters are subsequently stored in a fits table based on the halo ID, after which they are merged with the TNG50 catalogue and visual classification array through the pandas merge() method. The total time for the morphological parameters to be determined took approximately 17 hours for all 1151 galaxies, making this the most time-consuming step for computation. This computation time was done with computer hardware of an Nvidia RTX 3070 Ti GPU, AMD Ryzen 5 5600x CPU and 16 gigabytes of DDR4 3200 MHz RAM. The largest bottleneck during calculation was the RAM, and using a computer with a greater amount of allocated memory space for the code will likely result into a smaller computation time.

From the Statmorph data multiple parameters are used, but only the ones directly relevant to training the linear regression model. The first 4 parameters are 2 different estimates of the ellipticity and elongation of gaussian function such that they have the same second-order moment as the galaxies, both relative to the centroid and to the point of minimised asymmetry. While these return similar values they do have larger discrepancies in mergers and irregular galaxies, and thus are all 4 used. The next 4 parameters used are the circular radius and elliptical semimajor axis of both the half-light radius and the Petrosian radius, assuming that the center minimises asymmetry. The half-light radius is defined by the radius as to where half the total emitted light of the galaxy is contained. For the Petrosian radius, the mean surface brightness is taken where it is equal to η, a fraction of the mean surface brightness (Rodriguez-Gomez et al. 2019). Beyond this the 20% and 80% light radius are used, along with the Gini coefficient, the M_{20} coefficient, the Gini bulge and merger statistic, and the Concentration, Asymmetry and Smoothness indices.

For the Gini coefficient the following formula is used to compute the parameter:

$$G = \frac{1}{|\overline{X}|n(n-1} \sum ni = 1(2i - n - 1)|X_i|$$

Where n is the set of pixels, X_i is the flux value of the individual pixels and $\overline{X}$ is the mean of the pixel values (Genel et al. 2015). The coefficient ranges from 0 to 1, where 0 is a homogeneous distribution and 1 is akin to a dirac delta function, with all flux contained within a single pixel. This formula is an edited one from the originally created gini coefficient formula due to the use of absolute values. These are put in place as to work with galaxies with a high amount of noise (J. M. Lotz, Primack, and Madau 2004).

For the M_{20} calculation the total second-order central moment μ_{tot} is calculated first, from which the M_{20} statistic is calculated through:

$$M_{20} \equiv log_{10}(\frac{\sum_i \mu_i}{\mu_{tot}}), while \sum I_i < 0.2 I_{tot}$$

Where μ_i are the individual second-order central moments, I_i are the individual flux values for each pixel and I_{tot} is the total flux of all pixels, within the bounds of the segmentation map given (Genel et al. 2015). The resulting M_{20} statistic is thus the second moment of the galaxy wherein 20% of the galaxy's total flux is contained.

Following this the bulge and merger statistics are calculated through using the calculated Gini coefficient and M_{20} statistic. To this end the bulge statistic is calculated through $F(G, M_{20}) = -0.693M_{20} + 4.95G - 3.96$ (Snyder, J. Lotz, et al. 2015) and the merger statistic through $S(G, M_{20}) = 0.139M_{20} + 0.990G - 0.327$ (Snyder, Torrey, et al. 2015). The bulge statistic here is defined through indicating a more bulge-dominated galaxy through a higher value in F where as for the merger statistic a higher S indicates a merger dominated galaxy.

Lastly the Concentration, Asymmetry and Smoothness (also called Clumpiness) indices are calculated. For the Concentration index the formula $C = 5log_{10}(\frac{r_{80}}{r_{20}})$ is used (Bershady, Jangren, and Conselice 2000), where r_{80} and r_{20} are the 80% and 20% light radii respectively. For the Asymmetry index the galaxy is rotated by 180 degrees and subsequently subtracted from the original image through the formula $A = \frac{\sum_{i,j}|I_{ij} - I_{ij}^{180}|}{\sum_{i,j}|I_{ij}|} - A_{bgr}$, where I_{ij} is the pixel flux of the image, I_{ij} the pixel flux of the rotated image, and A_{bgr} is the background's average symmetry. The Smoothness index is calculated through first smoothing the image by a boxcar filter of a width σ, defined by $\sigma = 0.25 r_{petro}$ where r_{petro} is the Petrosian radius (J. M. Lotz, Primack, and Madau 2004). From this the Smoothness index is determined by the formula $S = \frac{\sum_{i,j} I_{ij} - I_{ij}^S}{\sum_{i,j} I_{ij}} - S_{bgr}$, where I_{ij}^S is the pixel flux of the smoothed image and S_{bgr} is the background's average smoothness (Conselice, Bershady, and Jangren 2000).

4 Results and Discussion

In this section the results of the thesis are discussed, along with a discussion on why certain observed results are as such. Specifically the disk fraction of the galaxies and morphological parameters are inspected with respect to the visual classification done in the previous section. Lastly the linear regression model is explained along with the results from said model.

4.1 Disk Fraction

Using the disk fraction (f_d) parameter of $\epsilon > 0.7$ minus $\epsilon < 0$ and the colour g-r based on the given fluxes, histogram plots are made with a separation of elliptical galaxies having an f_d of 0.5 or lower, while spiral galaxies have an f_d higher than 0.5. Comparing these histograms as can be seen in figure 10 with visually separated galaxies as in figure 11, it can be seen that within the kinematically classified galaxies there is a high prevalence of rotationally dominated galaxies with relation to dispersion dominated galaxies, with visually classified galaxies sharing a similar division between disk and spheroid galaxies. However, visual classification shows a greater difference in prevalence between disk galaxies and spheroid galaxies. Furthermore, visual classification contains a greater fraction of red spiral galaxies than through kinematic galaxies. The mergers which are present in the visual classification galaxies are negligible and have a random distribution of colour. These counts of a colour range share similar profiles as within Strateva et al. 2001 which is visible in figure 7, where they have been divided through morphological and spectroscopic sampling. With this it can be concluded that there appears to be correlation between visual and kinematic properties of galaxies, but that they are not identical in division between their respective groups.

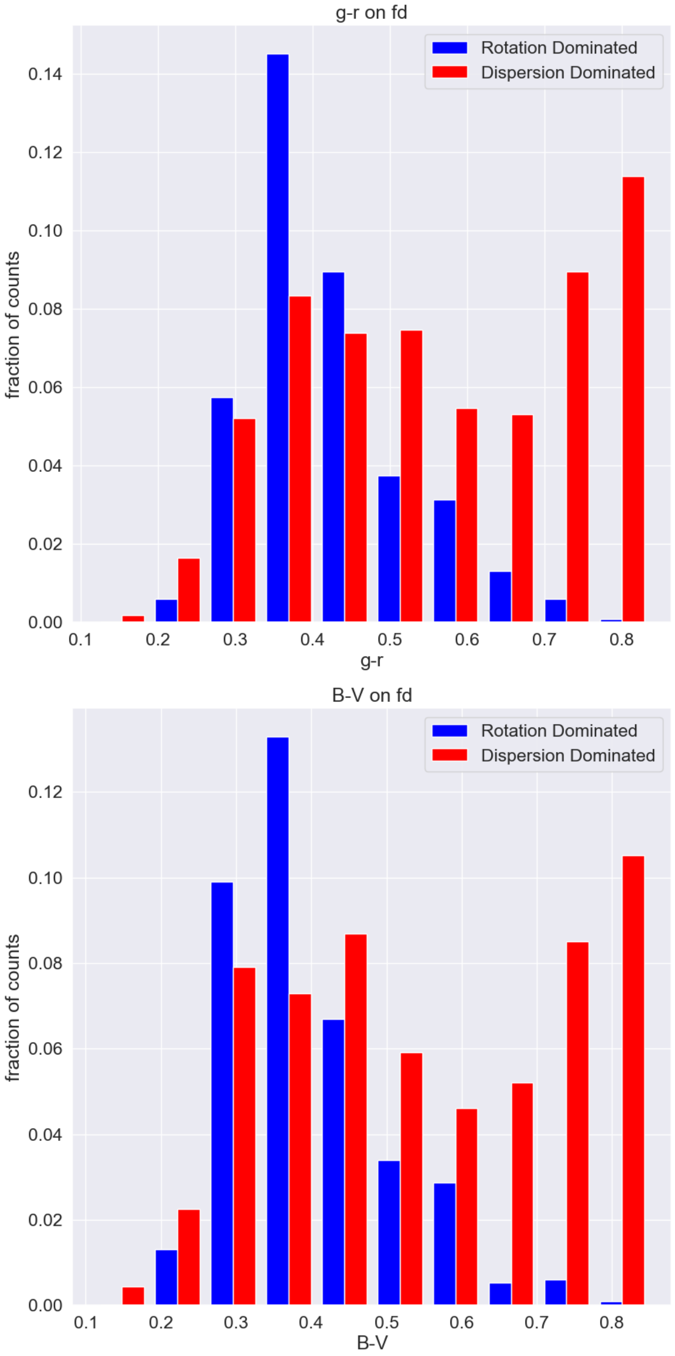

Figure 10: Histograms for colours g-r and B-V respectively, separated by f_d values

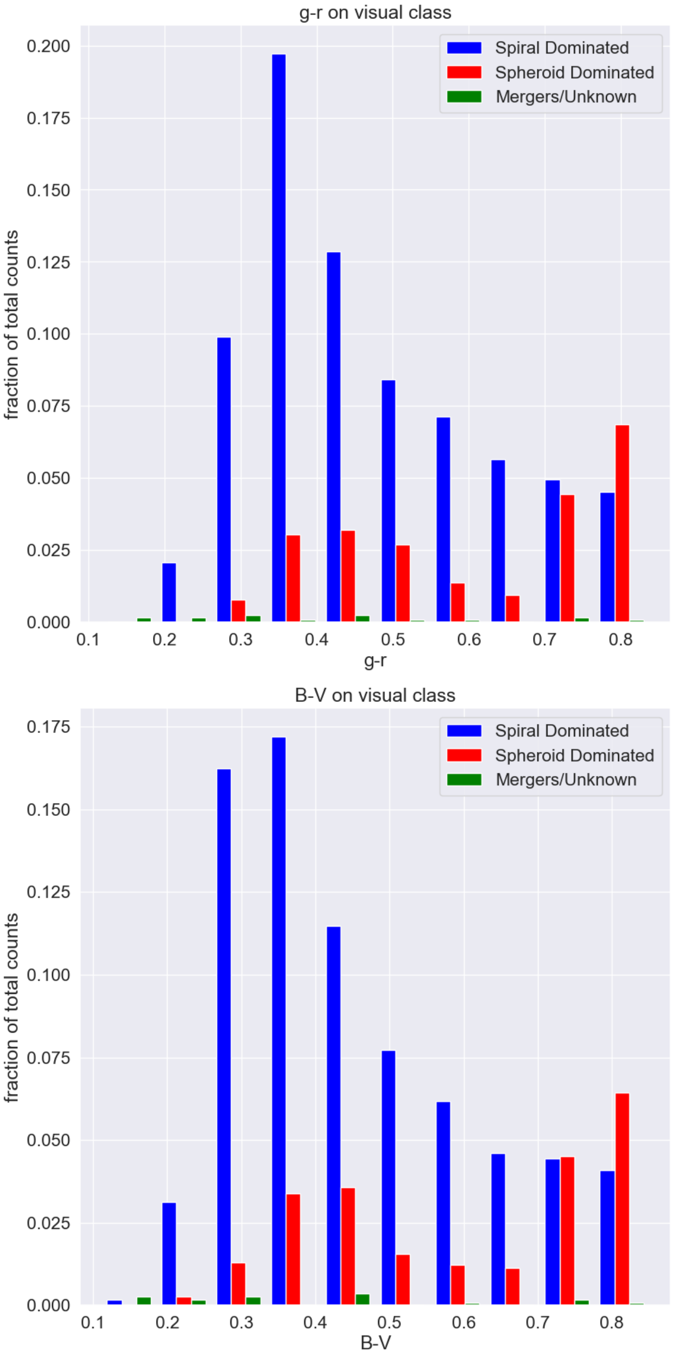

Figure 11: Histograms for colours g-r and B-V respectively, separated by visual classification

4.2 Morphological Parameters

During the computation of each galaxy in Statmorph, 29 galaxies returned errors during computation. While being fine on other machines when running those galaxies individually with the exact same initial parameters, they returned a NonFiniteValueError when applying the source_morphology() method over each galaxy sequentially. The convolved image and detected segmentation map were both checked for infinite, Nan and None values through built in python and numpy method which returned a boolean array. These arrays were subsequently summed up, where if any of these non-finite values were present it would return a nonzero total, yet every array returned a value of zero, indicating that they did not have non-finite values. It is also very unlikely that this error occurred through 0 division as multiple other 'fine' galaxies had data points in the image with a value of 0.

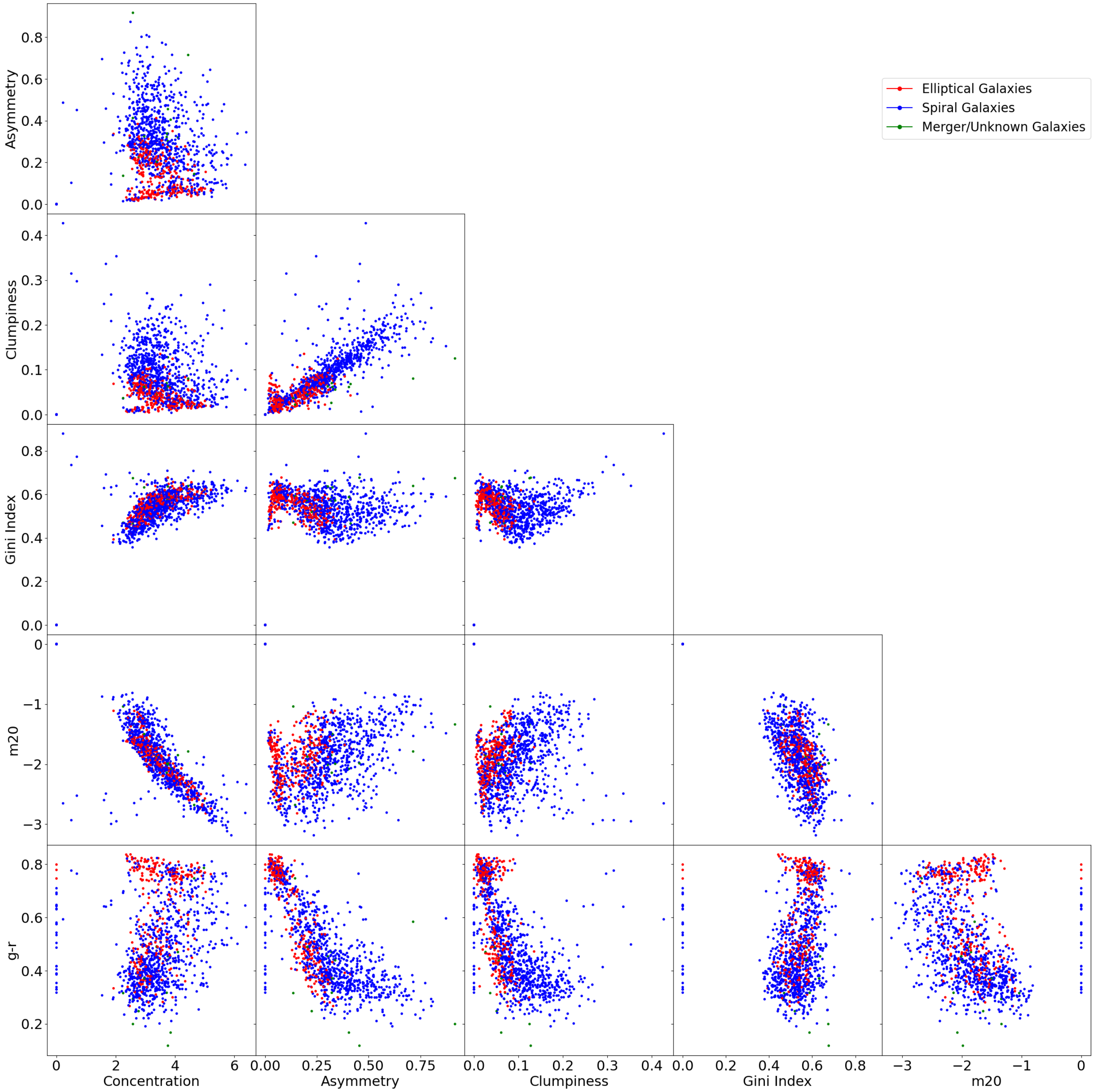

Figure 12: Multiple morphological parameters mapped to each other along with one colour parameter, in order to compare correlation and distinction between the groups. Each galaxy is separated into the 3 groups from visual classification.

Some of the morphological parameters along with the colour parameter g-r for comparison are plotted in figure 12, where the galaxies are each divided based on their visual classification. As can be seen in nearly each subplot, there is a clear distinction between elliptical and spiral galaxies on only some of the parameters, with mergers being randomly scattered. While the Gini and M20 index do not seem to have a strong distinction between spirals and elliptical galaxies, the CAS parameters along with the g-r colour seem to have a very strong distinction between them. This is most visible when comparing Asymmetry (A) with Clumpiness (S), showing a very strong positive correlation with elliptical galaxies being concentrated at low values, whereas the spiral galaxies are more spread out to higher index values. On either the origin of the subplots or for the g-r row there are the 29 galaxies visible which returned an error during the Statmorph computation. While it was not clear as to why this occurred, as they individually rendered fine, real data will hopefully not return any of these errors.

Unlike in the paper of J. M. Lotz, Primack, and Madau 2004, where the Gini coefficient was observed to be high for early-type galaxies and low for late-type galaxies, figure 12 appears to not show a distinction on the Gini coefficient between visually classified elliptical and spiral galaxies. While in (Hopkins et al. 2008) it is discussed how Gini on itself does often disagree with visual classification, it is still a significant oddity through the perspective of galaxy classification through different categories. This could possibly be due to their paper using high redshift early-type galaxies, which would have different properties as in figure 3 early-type galaxies have a higher star formation rate at higher redshifts. This in turn could shift the distribution of pixel flux measured and shift the Gini coefficient. An alternative explanation could be simply due to the simulation possessing a systematic error in terms of pixel flux distribution of galaxies, but visual inspection did not indicate as such. Lastly there is one possible reason, though it is far more unlikely to the simulated galaxies possessing a far greater resolution than imaging at high redshifts. Taking the PSF of 0.16" of euclid into account for with the simulation, the PSF would influence lower resolution images more in terms of their pixel flux distribution. This possible cause does suffer from the fact that the observations of (J. M. Lotz, Primack, and Madau 2004) used the Hubble Space Telescope while this thesis uses a simulated version. In the case of more future measurements done through both real and simulation observed media, it should be possible to either confirm or refute this.

In Cassata et al. 2007, the CAS parameters were used alongside Gini and M_{20} in order to compare morphological with visual classification. As can be seen in their data in figure 6, Concentration against Asymmetry and Clumpiness against Asymmetry have a very similar shape to the subplot in figure 12 (take in mind the different point of origin in both figures). However, the Gini coefficient and M_{20} statistic here too show a distinction between visual classifications, while the morphological parameters from Statmorph did not do so as mentioned before. This could once again indicate a difference between real observed data and simulated galaxies.

With all the data combined, a linear correlation matrix between all relevant parameters was made as can be seen in figure 13. As can be seen in this figure, the ellipticity and elongation seem to have near identical correlation to the other parameters between the two different methods of acquiring those properties. This same pattern can be observed for the half-light, Petrosian, 20% and 80% radii. This is according to what is to be expected from the estimation, as they all depends similarly on the brightness values of each pixel on the images. This matrix in turn also assists in finding unexpected correlations or anti-correlations, with for example the colours very strongly anti-correlating to the Asymmetry index, showing that galaxies that are more blue or shorter wavelength have a higher asymmetry index than redder galaxies. This is visible in figure 12, as mergers are more prevalent at a lower g-r value, but more significantly spiral galaxies appear to have a far higher asymmetry index than elliptical ones.

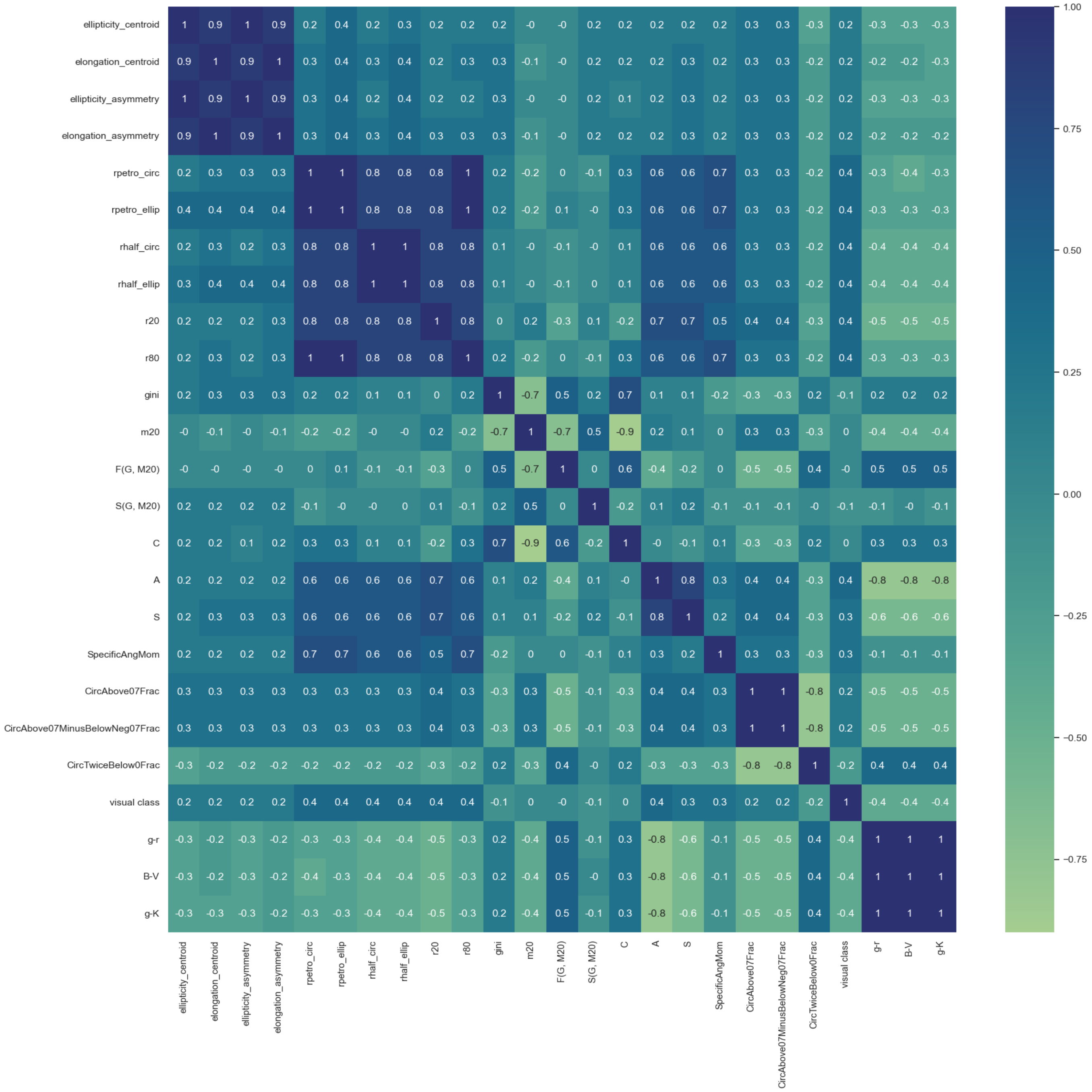

Figure 13: A correlation matrix heatmap of each parameter, with the correlation coefficient rounded to 1 decimal. Note the identical rows and columns of similar parameters.

4.3 The Kinematic Linear Regression Model

The last step towards determining the model is through running a linear regression over multiple parameters to train for separating the galaxies by some bound of the disk fraction. After this sklearn's linear_model.LinearRegression() method is used to return the f_d for a set of 7 input variables. For testing the uncertainty of the model when predicting data, the data is split between training and testing data where 20% of the data set is split for testing. The parameters put in are both morphological (e.g. ellipticity, Gini coefficient) and visual (e.g. colours). The returned f_d value can subsequently be used to kinematically classify the galaxies without a need for complex calculation of kinematic properties. From the linear regression method 3 data sets are returned: the coefficient of determination R^2, the coefficients of each input parameter and the intercept of the linear function. R^2 is calculated through the following formula:

$$R^2 = 1 - \frac{u}{v}$$

, where $u = \sum(y_{data} - y_{model})^2$, $v = \sum(y_{data} - \mu_{data})^2$ of which y_{data} and y_{model} are the y values of the data and model respectively and μ_{data} is the mean of the data. This determination coefficient ranges from some negative value to 1, where 0 disregards input data and 1 disregards prediction data. The parameters used to train the data were limited to elongation with relation to the centroid, the Petrosian elliptical radius, the Gini coefficient, the M_{20} statistic, the Asymmetry and Smoothness indices and the color g-r, as to limit the amount of parameters that are necessary for the model to work. These parameters were chosen based off the correlation matrix in figure 13 where parameters with similar to identical correlation were left out as to include as many differently related parameters.

The linear regression returned an R^2 of 0.4444, which could be improved by more training data or more input parameters. While more input parameters could assist this, it is more preferable to have more training data instead, as the model becomes more of an arbitrary step to add to the calculation if more parameters have to be calculated first. For the slopes, see table 1. The intercept was found to be 0.6484. Lastly the Mean Squared Error (MSE) was calculated to be at 0.02250 and the Mean Absolute Deviation (MAD) at 0.1242.

Elongation Centroid	$r_{petro,ellipse}$	Gini coeff.	M_{20}	A	S	g-r
9.573e-2	-1.484e-4	-9.047e-1	-3.068e-2	2.123e-1	7.793e-1	-1.442e-1

Table 1: Slopes for the input parameters of the model

A current limitation for the model is that due to the training data being done at a redshift of z=0, it will only work within the local group when applied to real observations. This can be easily fixed however by applying a correction to the colour based on the redshift, but this will be very important in order to give accurate results. The other input data rely only on the data received on the CCD of the telescope and the subsequent produced image. Along with that possible limitation, the difference between the observed Gini coefficient and M_{20} statistic as mentioned in section 4.3 does mean that there is a possible error in the model when using those parameters. In the case that this is on the basis of the simulation having a greater resolution with a small PSF, this would be solved through having greater resolution telescopes on the level of the Euclid telescope in the future.

5 Conclusion and Outlook

Within the project 1160 TNG50 simulated galaxies were given in both images of 1600x1600 pixel resolution and as a data table of kinematic and colour parameters of each galaxy at z=0 with mass ranging from 6×10^9 to 1×10^{12} solar masses. Each images was convolved using the Euclid VIS statistics for the PSF. Each galaxy was manually inspected and visually classified between disk, elliptical and merger/irregular galaxies. From these images at high resolution morphological parameters were computed using Statmorph. From these parameters, multiple plots were made comparing morphological, kinematic and colour-based properties of the galaxies, dividing them based on the visual classification. The morphological, colour and visual classification parameters were subsequently inspected through their correlation matrix to remove any parameters that already were very similar to other parameters. After this the remaining parameters were used as training data through linear regression for the f_d. Finally, the coefficients and intercept from the linear regression are used for the model, such that the kinematic property f_d can be returned from morphological and colour parameters. From this, the following conclusions are drawn:

1. There is a direct correlation between visual classification and kinematic classification, as can be seen in figure 10 and 11 respectively.

2. The Gini coefficient and M_{20} statistic from the simulated galaxies do not agree with the research in the paper of J. M. Lotz, Primack, and Madau 2004 and Cassata et al. 2007. The parameters in the data in this project show no distinction of the galaxies through visual classification. The CAS parameters however do agree with the other papers in that they show a clear distinction between the different galaxies as can be seen in figure 12.

3. The model produced from the linear regression produced usable coefficients and intercepts (see table 1) with an MSE of 0.02250 and MAD of 0.1242. The score of determination of 0.4444 indicates that more training data or more input parameters are advisable, where more training data is preferable as to not make the model require too much computation before use.

Within the computed data there were 9 poorly rendered galaxies which were removed as they did not appear to have any distinct visual properties. Within the use of Statmorph there were 29 more galaxies that returned an error during computation. Within the method source_morphology() a non-finite value was encountered, but there were no oddities observed within these galaxies when compared to the working samples. Lastly, as the gain of the images was unknown, it was set to 1. This results in the sersic parameters returned by Statmorph to be inaccurate and thus they were removed from the morphological data set. No further errors were encountered during the project. The model of the project is expected to be able to be used for real observations, allowing researchers to determine kinematic properties more efficiently. As the model has not been trained on real data yet, this still has to be done to test the accuracy of it.

6 Acknowledgments

I would like to thank my supervisors for both of their continuous support during the thesis project and for giving constructive advice during the writing of it. Specifically, I want to thank Dr. Lingyu Wang for helping me with the start of the project as I struggled to find a project. For Antonio La Marca I thank him for helping me with the code provided for convolution along with other codes and the steps to take when this proceeded to not function in the slightest. Furthermore I would like to thank my sister for giving additional feedback on the writing, allowing a perspective on the thesis that does not necessarily have in-depth knowledge on the project already.

This project makes use of the IllustrisTNG simulation, with specific galaxies from the TNG50 Volume.

In order to find references ChatGPT was used to recommend a list of research articles which were subsequently read to be positive that the recommendations were relevant.

References

Bershady, Matthew A., Anna Jangren, and Christopher J. Conselice (June 2000). “Structural and Photometric Classification of Galaxies. I. Calibration Based on a Nearby Galaxy Sample”. In: 119.6, pp. 2645–2663. DOI: 10.1086/301386.

Cassata, P. et al. (Sept. 2007). “The Cosmic Evolution Survey (COSMOS): The Morphological Content and Environmental Dependence of the Galaxy Color-Magnitude Relation at z ~0.7”. In: 172.1, pp. 270–283. DOI: 10.1086/516591.

Conselice, Christopher J., Matthew A. Bershady, and Anna Jangren (Feb. 2000). “The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies”. In: 529.2, pp. 886–910. DOI: 10.1086/308300.

Genel, Shy et al. (May 2015). “Galactic Angular Momentum in the Illustris Simulation: Feedback and the Hubble Sequence”. In: 804.2, L40, p. L40. DOI: 10.1088/2041-8205/804/2/L40.

Georgakakis, A. et al. (Oct. 2001). “Cold gas in elliptical galaxies”. In: 326.4, pp. 1431–1440. DOI: 10.1111/j.1365-2966.2001.04677.x. arXiv: astro-ph/0105435 [astro-ph].

Holwerda, Benne W. et al. (Sept. 2019). “The Frequency of Dust Lanes in Edge-on Spiral Galaxies Identified by Galaxy Zoo in KiDS Imaging of GAMA Targets”. In: 158.3, 103, p. 103. DOI: 10.3847/1538-3881/ab2886. arXiv: 1909.07461 [astro-ph.GA].

Hopkins, Philip F. et al. (Apr. 2008). “A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. II. Formation of Red Ellipticals”. In: 175.2, pp. 390–422. DOI: 10.1086/524363.

Hubble, Edwin. (Jan. 1926). “No. 324. Extra-galactic nebulae.” In: *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington* 324, pp. 1–49.

Kraljic, Katarina, Frédéric Bournaud, and Marie Martig (Sept. 2012). “The Two-phase Formation History of Spiral Galaxies Traced by the Cosmic Evolution of the Bar Fraction”. In: 757.1, 60, p. 60. DOI: 10.1088/0004-637X/757/1/60. arXiv: 1207.0351 [astro-ph.GA].

Lintott, Chris J. et al. (Sept. 2008). “Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey”. In: 389.3, pp. 1179–1189. DOI: 10.1111/j.1365-2966.2008.13689.x.

Lotz, Jennifer M., Joel Primack, and Piero Madau (July 2004). “A New Nonparametric Approach to Galaxy Morphological Classification”. In: 128.1, pp. 163–182. DOI: 10.1086/421849.

Loveday, Jon (Feb. 1996). “The APM Bright Galaxy Catalogue”. In: 278.4, pp. 1025–1048. DOI: 10.1093/mnras/278.4.1025.

Mandelbaum, Rachel et al. (May 2006). “Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies”. In: 368.2, pp. 715–731. DOI: 10.1111/j.1365-2966.2006.10156.x. arXiv: astro-ph/0511164 [astro-ph].

Melvin, Thomas et al. (Mar. 2014). “Galaxy Zoo: an independent look at the evolution of the bar fraction over the last eight billion years from HST-COSMOS”. In: 438.4, pp. 2882–2897. DOI: 10.1093/mnras/stt2397. arXiv: 1401.3334 [astro-ph.GA].

Mo, Houjun, Frank C. van den Bosch, and Simon White (2010). *Galaxy Formation and Evolution.*

Naab, Thorsten et al. (Apr. 2007). “Formation of Early-Type Galaxies from Cosmological Initial Conditions”. In: 658.2, pp. 710–720. DOI: 10.1086/510841.

Nelson, Dylan et al. (May 2019). “The IllustrisTNG simulations: public data release”. In: *Computational Astrophysics and Cosmology* 6.1, 2, p. 2. doi: 10.1186/s40668-019-0028-x.

Rodriguez-Gomez, Vicente et al. (Mar. 2019). “The optical morphologies of galaxies in the IllustrisTNG simulation: a comparison to Pan-STARRS observations”. In: 483.3, pp. 4140–4159. doi: 10.1093/mnras/sty3345.

Scoville, N. et al. (Sept. 2007). “The Cosmic Evolution Survey (COSMOS): Overview”. In: 172.1, pp. 1–8. doi: 10.1086/516585.

Snyder, Gregory F., Jennifer Lotz, et al. (Aug. 2015). “Diverse structural evolution at z ¿ 1 in cosmologically simulated galaxies”. In: 451.4, pp. 4290–4310. doi: 10.1093/mnras/stv1231.

Snyder, Gregory F., Paul Torrey, et al. (Dec. 2015). “Galaxy morphology and star formation in the Illustris Simulation at z = 0”. In: 454.2, pp. 1886–1908. doi: 10.1093/mnras/stv2078.

Strateva, Iskra et al. (Oct. 2001). “Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data”. In: 122.4, pp. 1861–1874. doi: 10.1086/323301.

van der Wel, A. et al. (Sept. 2005). “Mass-to-Light Ratios of Field Early-Type Galaxies at z ˜1 from Ultradeep Spectroscopy: Evidence for Mass-dependent Evolution”. In: 631.1, pp. 145–162. doi: 10.1086/430464. arXiv: astro-ph/0502228 [astro-ph].