Beyond Spheroids and Discs: Classifications of CANDELS Galaxy Structure at $1.4 < z < 2$ via Principal Component Analysis

Michael A. Peth$^1$, Jennifer M. Lotz$^2$, Peter E. Freeman$^3$, Conor McPartland$^2$, S. Alireza Mortazavi$^1$, Gregory F. Snyder$^2$, Norman A. Grogin$^2$, Yicheng Guo$^4$, Shoubaneh Hemati$^5$, Jeyhan S. Kartaltepe$^6$, Dale D. Kocevski$^7$, Anton M. Koekemoer$^2$, Daniel H. McIntosh$^8$, Hooshang Nayyeri$^5$, Casey Papovich$^9$, Joel R. Primack$^4$, Raymond C. Simons$^1$ and Arjen van der Wel$^{10}$

$^1$Department of Physics and Astronomy, The Johns Hopkins University, 366 Bloomberg Center, Baltimore, MD 21218
$^2$Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218
$^3$Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213
$^4$University of California Observatories/Lick Observatory, University of California, Santa Cruz, CA 95064
$^5$Department of Physics and Astronomy, University of California, Riverside, CA, USA
$^6$National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, AZ 85719, USA
$^7$Department of Physics and Astronomy, Colby College, Waterville, ME 0490, USA
$^8$Department of Physics and Astronomy, University of Missouri-Kansas City, 5110 Rockhill Road, Kansas City, MO 64110
$^9$George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843
$^{10}$Max-Planck Institut fur Astronomie, Konigstuhl 17, D-69117, Heidelberg, Germany

Accepted ??; Received ??; in original form ??

ABSTRACT
To understand the processes driving galaxy morphology and star formation, we need a robust method to classify the structural elements of galaxies. Important but rare and subtle features may be missed by traditional spiral, elliptical, irregular or S"ersic bulge/disc classifications. To overcome this limitation, we use a principal component analysis of non-parametric morphological indicators (concentration, asymmetry, Gini coefficient, $M_{20}$, multi-mode, intensity and deviation) measured at rest-frame B-band (corresponding to HST/WFC3 F125W at $1.4 < z < 2$) to trace the natural distribution of massive ($> 10^{10} M_\odot$) galaxy morphologies. Principal component analysis (PCA) quantifies the correlations between these morphological indicators and determines the relative importance of each. The first three principal components (PCs) capture $\sim$75 per cent of the variance inherent to our sample. We interpret the first principal component (PC) as bulge strength, the second PC as dominated by concentration and the third PC as dominated by asymmetry. Both PC1 and PC2 correlate with the visual appearance of a central bulge and predict galaxy quiescence. We divide the PCA results into 10 groups using an agglomerative hierarchical clustering method. Unlike S"ersic, this classification scheme separates quenched compact galaxies from larger, smooth proto-elliptical systems, and star-forming disc-dominated clumpy galaxies from star-forming bulge-dominated asymmetric galaxies. Distinguishing between these galaxy structural types in a quantitative manner is an important step towards understanding the connections between morphology, galaxy assembly and star-formation.

Key words: galaxies: structure – galaxies: evolution – methods: data analysis

* E-mail: mikepeth@jhu.edu

© 2015 RAS
1 INTRODUCTION

Galaxies today form stars at a lower rate than in the past due to many factors. However, we do not have a complete accounting of the processes quenching the star-formation in galaxies. An increase in the mass/number densities (Tomczak et al. 2014), and growth rate (van der Wel et al. 2014) of massive, red galaxies implies stars are not forming to the same extent they once were. The lack of star-formation is also seen very clearly in clusters and environmental studies (Mortlock et al. 2015). Each of these observations attempt to connect some combination of observed colour (or star-formation rate), stellar masses and morphology.

The star-formation rate - stellar mass (SFR–$M_*$) relationship shows star-forming galaxies at $z \sim 0$ follow a “main sequence” (Brinchmann et al. 2004; Wuyts et al. 2011). Galaxies on the main sequence are bluer and have lower S´ersic-indices than galaxies below the relation. Massive galaxies with low SFRs are red and have high S´ersic indices and bulge strengths. The SFR–$M_*$ relation has been shown to hold out to $z \sim 2.5$ (Wuyts et al. 2011). However, bulge strength has been described as a “necessary but not sufficient” condition for quenching star-formation for all $z \lesssim 2.2$ galaxies (Bell et al. 2012).

If the presence of a bulge is not sufficient to fully quench a galaxy other factors, such as size, may be important for shutting down star-formation. At redshifts $z \sim 1.5$ galaxies of sufficiently high mass and small size ($\sim 10.3 M_\odot$ kpc$^{-1}$) are quenched (Barro et al. 2013). This suggests a relationship between so-called “compactness” ($\Sigma_{1.5} = M/r_{e}^{2.5}$) and SFR. However, the number density of these compact galaxies has been decreasing with the age of the universe.

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

The mechanisms for quenching star-formation and transforming the morphology of galaxies are not fully understood. Proposed mechanisms include: major mergers (e.g. Naab et al. 2006; Hopkins et al. 2010); minor mergers (e.g. Hopkins & Hernquist 2009; Villforth et al. 2013; Taniguchi 1999); secular processes (for review see Kormendy & Kennicutt 2004); AGN feedback (e.g. Silk & Rees 1998; Schawinski et al. 2006); and mass quenching (Dekel & Birnboim 2006; Bell et al. 2012). Comprehensive models of galaxy formation can yield a reasonable link between galaxy morphology and star formation (e.g. Snyder et al. 2015) but we do not yet have a perfect accounting of how all these processes might contribute.

Each mechanism leaves behind a different fingerprint upon the shape and structure of a galaxy. For instance, major mergers can leave behind large tidal tails (Toomre & Toomre 1972). Following a minor merger the morphology of the galaxy is unaffected and the galaxy appears undisturbed and disc-like (i.e. fit by low S´ersic index; Schawinski et al. 2011).

To study the processes driving evolution, we need a method to effectively and efficiently characterize the structures and shapes of galaxies. Visual classifications have been used since the discovery of galaxies, and have subsequently been adapted to fit modern surveys (e.g. Galaxy Zoo,Lintott et al. 2008; Kartaltepe et al. 2011). Visual classifications can find subtle structural elements possibly missed by an automated routine. However, human classifications of galaxies can be very time consuming and subjective.

GALFIT (Peng et al. 2002; 2010) is an automated technique often used to classify galaxies that models the light profile of galaxies by S´ersic profile ($r^{-1/n}$). Discs have exponential light profiles ($n=1$), while ellipticals are best fit by a de Vaucouleurs profile ($n=4$). GALFIT is sensitive to small galaxies, can distinguish overlapping light profiles of nearby galaxies, incorporates the point spread function of a specific field/detector, and most importantly is easy to interpret. However, GALFIT assumes a symmetric and smooth light profile, which at times can be problematic. This assumption does not hold for irregular galaxies, merger remnants, and disc galaxies with bars or clumps. GALFIT is typically used to calculate a single S´ersic fit to the light profile of the galaxy. But two-component S´ersic fits has also been used to combine disc and bulge components (e.g. Simard et al. 2011; Bruce et al. 2014). However, calculations can be quite CPU intensive (sometimes taking weeks to finish).

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

However, for many galaxies these statistics can be strongly correlated. Moreover, cosmological models of galaxy formation yield a picture in which these structures can evolve quickly along diverse paths, thereby motivating a broad deep classification system (Snyder et al. 2014). Therefore we require further analysis to understand the inherent relationships among these statistics, and between these statistics and galaxy assembly processes.

Principal component analysis (PCA) is a simple way to reduce the dimensionality, break internal degeneracies and find the natural distributions of data in parameter space. To eliminate degeneracies inherent in these morphological statistics we performed a PCA using 7 non-parametric morphology measurements on 962 galaxies from 1.36 < z < 1.97. PCA has been shown to efficiently classify galaxies (e.g. Taghizadeh-Popp et al. 2012, the Zurich Estimator of Structural Types (ZEST). Scarlata et al. 2007). A few studies immediately capitalized on the ZEST classifications to study the number density evolution of disc galaxies (Sargent et al. 2007), the luminosity function evolution for elliptical...
galaxy progenitors (Scarlata et al. 2007), and the evolution of the galaxy merger rate to $z \sim 1$ (Kampczyk et al. 2007).

In this paper, we use PCA and hierarchical clustering to classify galaxies based on their structure. These classifications allow us to characterize galaxies by more subtle means than the traditional Hubble sequence scheme. We can test the mechanisms which cause galaxies to reassemble and/or influence star-formation by tracking how morphologies change across time. This places vital constraints on the types of physical mechanisms assembling galaxies and quenching star-formation.

This paper is structured as follows: §2 details the CANDELS data set and our sample selection; §3 defines the non-parametric morphologic measurements we perform, their associated error estimates and the principal component analysis as applied to our data set; §4 describes the results of our PCA, the clustering algorithm and convex hull method used for grouping galaxies, a test of the group and PCA reliability and descriptions of the final galaxy groups; §5 describes the general morphological characteristics of the galaxies in each group; §6 details the relationship between stellar mass, quenching and our group classifications. Additionally, the disagreement between Sérsic and visual classifications with PC-based classification (especially for compact/bulge-dominated galaxies) is discussed.

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

## 2 DATA

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

We use the CANDELS $H$-band (F160W) selected multi-wavelength catalogs (UDS, Galametz et al. 2013; GOODS-S, Guo et al. 2013; COSMOS, Nayyeri et al., in prep), photometric redshifts (Dahlen et al. 2013), non-parametric morphologies (this work), Sérsic parameters (van der Wel et al. 2012), visual classifications (Kartaltepe et al. 2014), rest-frame photometry, and stellar masses (this work). The limiting magnitude for HST/WFC3 F125W and F160W are 27.35 and 27.45 respectively. Both bands have a FWHM of $\sim$0.20 arcsec. Galametz et al. (2013) outlined the techniques used to create the photometric catalogs.

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

Rest-frame $U-V-J$ colours were calculated by the sED-fitting code EAZY (Brammer et al. 2008), using the empirical local galaxy templates of Brown et al. (2011 and Koekemoer et al. in prep), assumed Bruzual & Charlot (2003) delayed exponential star-formation histories, a Chabrier (2003) initial mass function, Calzetti et al. (2000) dust attenuation, and solar metallicities.

### 2.1 Sample Selection Criteria

We select bright ($H < 24.5$), massive ($M_* > 10^{10} M_\odot$) galaxies with $1.36 < z < 1.97$ galaxies measured in F125W ($J$). This band corresponds to rest-frame optical $B$-band at this redshift. This redshift range provides a large sample of galaxies measured in a constant rest-frame waveband, and offers a high enough redshift to have a different morphological distribution from a local sample. At this redshift and magnitude, the CANDELS surveys are mass-complete down to $10^{10} M_\odot$ (Wuyts et al. 2011). In our sample of UDS, COSMOS and GOODS-S there are a total of 6144 galaxies with $H < 24.5$ and $M_* > 10^{10} M_\odot$. Of those galaxies 1452 are within our redshift range ($1.36 < z < 1.97$).

The following affect our sample completeness: high signal-to-noise (per pixel) measurements ($S/N > 4$), an internal morphology quality flag = 0, and a well measured concentration (i.e. $C > 0$) requirement. The quality flag requirement removes objects from the sample with unconnected segmentation maps and bright neighbors. The concentration requirement removes the contamination from poorly measured galaxies on the overall PCA. For many galaxies, $r_{20}$ (and thus $C$) can not be accurately measured because either the object is too small, or there is a bright point source disrupting the light profile (see §3.1.2). The signal-to-noise cut reduces the sample to 1388. The FLAG requirement reduces the sample to 1203. The concentration requirement reduces the total of galaxies in the sample to 1136. The signal-to-noise, FLAG and well measured concentration requirements together reduce our final sample to 962 galaxies.

We separately test the effects of concentration and FLAG=1 galaxies on our sample. For the $z$=1.5 galaxies with $C < 0$, the problem derives from the 20 per cent light radius ($r_{20}$) calculated to be smaller than 1 pixel and thus fail the code (see §3.1.2 for more details). We assume a lower bound on $r_{20}$ fixed at 1 pixel. We determine the group characteristics (such as percentage of disc/spheroid galaxies in a single group) are still within the error limits from the original sample. These galaxies do not change our conclusions as to the nature of each group. Galaxies with FLAG = 1 are typically either low surface brightness galaxies, tidally disrupted or have a bright neighbor.
3 MORPHOLOGICAL MEASUREMENTS

3.1 Non-parametric Morphology

We focus on non-parametric morphology statistics: concentration, asymmetry, Gini coefficient, $M_{20}$, along with three new statistics from Freeman et al.\textsuperscript{2013} multi-mode, intensity and deviation. The code for calculating the morphological statistics (originally developed by Lotz et al.\textsuperscript{2008}) has been modified to include new statistics and accommodate much larger input images. The code is applied to the CANDELS F125W mosaics using the F160W detected catalogs and segmentation maps as the input.

3.1.1 Petrosian Radius

The Petrosian radius $r_p$ is the radius at which the surface brightness $\mu$ is 20 per cent of the mean interior surface brightness (\textsuperscript{Petrosian}1976, Eq. 1). The Petrosian radius is more robust to surface brightness dimming than isophotal brightness (Petrosian 1976; Eq. 1). The Petrosian radius is calculated from a sum of all pixels within 1.5 Petrosian radii from the center of the galaxy. An initial guess for the center of rotation is defined by the physical center, but is updated through an iterative process. This process continues until a global minimum value for $A$ is found (\textsuperscript{Conselice}2014).

\begin{equation}
A = \frac{\sum_{x,y} |I_{x,y} - I_{180(x,y)}|}{2 \sum |I_{x,y}|} - B_{180}
\end{equation}

Due to their uniform morphologies and lack of structure elliptical galaxies typically have small asymmetry values ($A \sim 0.02$). Meanwhile spiral galaxies usually have values between $A \sim 0.07$ to 0.2 (\textsuperscript{Conselice}2014). This statistic is most useful for identifying irregular galaxies because they appear lopsided or ragged. Visually inspected merger remnants can have $A \geq 0.3$ (\textsuperscript{Conselice}2003). The asymmetry statistic is more sensitive to gas-rich mergers than to gas-poor or minor mergers (Lotz et al.\textsuperscript{2010a,b}).

If the local background is high (due to e.g. a bright neighboring galaxy) then negative $A$ values are possible. Generally, the magnitudes of these values are meaningless and should be interpreted as $A = 0$.

3.1.4 Gini Coefficient

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

The pixels are ranked by increasing flux value, then $G$ is determined by Eq. 4, where $n$ is the number of pixels in the galaxy’s segmentation map, $X_i$ is the pixel flux at the rank $i$ pixel and $\bar{X}$ is the mean pixel value.

\begin{equation}
G = \frac{1}{X_n(n-1)} \sum_{i} (2i - n - 1) X_i
\end{equation}

A galaxy with equally distributed light will have a Gini coefficient approaching 0. Conversely, a galaxy with a large fraction of light concentrated on a few pixels will have a Gini coefficient closer to 1. Elliptical galaxies and galaxies with bright nuclei have high Gini coefficients, while discs and galaxies with a uniform surface brightness will have low Gini coefficients.

3.1.5 $M_{20}$

The second order moment of the brightest regions of a galaxy ($M_{20}$; Lotz et al.\textsuperscript{2004}) traces the spatial distribution of any bright clumps. When used in tandem with the Gini coefficient, $M_{20}$ can be an effective tool for differentiating galaxies with bright off-center clumps (such as irregular galaxies) from those with one bright central region (such as the bulge of a spiral galaxy). We define the regions representing the brightest 20 per cent of the galaxy (Eq. 3), and then calculate the spatial distribution of those pixels as an offset.
Figure 1. F125W (AB) = 22.2 CANDELS galaxy image is shown to demonstrate the $M$, $I$ and $D$ statistics. The left panel shows the image of the galaxy outlined by the segmentation map created using our morphology code. The middle panel shows red outlines describing the clumps found when calculating the $M$ statistic. The white $X$ displays the location of the brightness distribution peak, and the cyan circle represents the location of the intensity centroid used to calculate the $D$ statistic (§3.1.8). The right panel colour codes the clumps for easy identification. This galaxy is highly disturbed and is broken into 3 bright regions, with the brightness peak well separated from the intensity centroid. The threshold value ($q_l$) in this case is 0.92, which represents the threshold where the $M$ statistic was maximized.

from the central pixel (which was previously defined when calculating asymmetry Eq. $6$)

$$\sum_i f_i < 0.2 f_{tot}$$  \hspace{1cm} (5)

$$M_{tot} = \sum_i^n M_i = \sum_i^n f_i \left[(x_i - x_c)^2 + (y_i - y_c)^2\right]$$  \hspace{1cm} (6)

Finally we calculate the second order moment (Eq. $7$).

$$M_{20} = \log \left(\frac{\sum_i M_i}{M_{tot}}\right)$$  \hspace{1cm} (7)

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

3.1.6 Multi-mode

The multi-mode ($M$) statistic is the ratio, in pixels, of the two brightest regions of a galaxy (adapted from Freeman et al. 2013). Bright regions are determined via a threshold method where $q_l$ represents the normalized flux value, and $l$ per cent of pixel fluxes are less than $q_l$. This creates a new binary image $g_{i,j}$ where 1 represents fluxes larger than $q_l$ and 0 represents fluxes less than $q_l$ (Eq. $8$).

$$g_{i,j} = \begin{cases} 
1 & f_{i,j} \geq q_l \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (8)

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

$$R_l = \frac{A_{l,(2)}}{A_{l,(1)}}$$  \hspace{1cm} (9)

The previous two steps are recomputed for various normalized flux levels $l$, and the $M$ statistic is the maximum $R_l$ value (Eq. $10$). Values approaching 1 represent multiple nuclei, while values near 0 are single nuclei systems.

$$M = \max R_l$$  \hspace{1cm} (10)

This formulation is slightly revised from Freeman et al. (2013) to limit the $M$ statistic to values between 0 and 1. Freeman et al. (2013) multiplies Eq. $8$ by an additional factor of $A_{l,(2)}$ to limit the effect of hot pixels. However, this adds a size dependent factor to the calculation. Because we wish to measure $M$ values for galaxies at a variety of angular distance scales, it is important to have a size independent measure. For illustrative purposes, Fig. 1 shows an example of how the MID statistics are calculated.
3.1.7 Intensity

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

\[
I = \frac{I_{(2)}}{I_{(1)}}
\] (11)

Similar to the \(M\) statistic, elliptical galaxies with a bright bulge have \(I \sim 0\), while disc galaxies with bright clusters of star-formation are more likely to have \(I\) values approaching 1.

3.1.8 Deviation

Deviation \((D)\) measures the distance between the intensity centroid of a galaxy and the center of the brightest region (Freeman et al. 2013) Eq. [12] and Eq [13]). Disc and spheroidal galaxies have deviation values near 0 because their central bulges typical possess the brightest pixels. On the other hand, a high deviation value indicates a galaxy has bright star forming knots significantly separated from the intensity centroid (e.g. Fig. 1).

\[
(x_{cen}, y_{cen}) = \left( \frac{1}{n_{seg}} \sum_i \sum_j f_{i,j}, \frac{1}{n_{seg}} \sum_i \sum_j jf_{i,j} \right)
\] (12)

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

\[
D = \sqrt{\frac{\pi}{n_{seg}} \left( (x_{cen} - x_{(1)})^2 + (y_{cen} - y_{(1)})^2 \right)}
\] (13)

3.2 Morphological Principal Components

Principal component analysis (PCA) is a linear transformation of multivariate data. This defines a set of uncorrelated axes, called principal components (PCs), which are ranked by the variance they capture (Pearson 1901, Jevtić et al. 2013). A linear combination of the original data and eigenvector solutions (also called weights) project the original data onto the PCs. Principal component analysis is a simple way to reduce the dimensionality and find the natural distributions of data in parameter space. PCA is able to determine the correlations between the input data and can find relationships missed by other means.

We begin by “whitening” the data, i.e. we subtract the mean of each morphological measurement and divide by the standard deviation of each feature. By dividing our data by feature variance we remove the effects of mixed units. We calculate the singular value decomposition \((x_{ij} = V \Sigma V^T, SVD)\) of the “whitened” data matrix \((x_{ij})\). A SVD decomposes the original data into a diagonal matrix containing eigenvalues \((\Sigma)\) and a non-diagonal matrix \(V\) containing the expansion coefficients (aka weights). The eigenvalues determine how important each principal component is to explaining the original data set. The eigenvectors are rank ordered by their associated eigenvalue. We then project our “whitened” data onto our new eigenbasis to calculate the principal component scores, which inform us how similar are data points to each other \((PC_i, Eq. [14]).\)

\[
PC_i = \sum_{j=1}^{N} V_{ji} x_{j} (i = 1, ..., N)
\] (14)

Table 1 shows the correlations and importance of different statistics across the eigenvector solutions of the principal component analysis. The scree value is square eigenvalue divided by the squared sum of all eigenvalues and represents the amount of variance in the data captured by a single principal component. The scree values demonstrate the first 3 PCs account for \(>75\) per cent of the variance in the data. The fact that PC1 only accounts for \(40\) per cent of the variance shows that more than a single parameter is needed to define a galaxy. The error estimates are the result of the scattering method described in 4.2.

PC1 is highly dependent upon \(M, I, D, M_{20}\) and the Gini coefficient. We interpret PC1 as a “bulge strength” indicator given the correlation with \(G\) \(- M_{20}\) and the importance of the \(MID\) statistics. The PC1/\(G\) \(- M_{20}\) “bulge strength” is an indicator of a dense compact core. PC2 is highly dependent upon concentration, and is larger for galaxies with bright centers and extended envelopes. PC3 is dominated by asymmetry and is larger for disturbed galaxies. The other principal components are harder to interpret, but are also less important as evidenced by their lower scree values. It is interesting to note PC1 defines a bulge strength but is not dependent on concentration.

4 PCA-MORPHOLOGY GROUP PROPERTIES

4.1 Defining PCA morphology groups

The morphologies of galaxies are not inherently discrete, but rather lie on a continuum. However, it is often useful to bin galaxies into discrete morphological groups. Fig. 2 shows the distribution of galaxies when projected onto the first three principal axes. Except for a large distinct cluster of data points most of our sample are not well separated, requiring the need for an objective data dependent grouping method.

To classify galaxies in distinct groups, we employ the Ward hierarchical agglomerative clustering routine of scikit-learn (Pedregosa et al. 2011). Hierarchical clustering (specifically agglomerative clustering) treats each galaxy as its own cluster, which are then merged with nearby
Table 1. PC Weights with error estimates based on a bootstrap scattering method

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

Figure 2. PC1 v. PC2 v. PC3 for our sample of $M_* > 10^{10} M_\odot$, 1.36 < z < 1.97 galaxies, colour-coded by their hierarchical cluster definitions. PC1 anti-correlates with bulge strength, PC2 is dominated by concentration, and PC3 is dominated by asymmetry (see Table 1). Group -1 galaxies (black stars) are outliers from remaining groups, initially they comprised groups 3 and 7.
The hierarchical clustering algorithm defines the groups based on the distribution of the data. In order to reproduce the same group definitions for new objects with potentially different distributions (e.g. different redshifts), we use a convex hull method to define the original group boundaries in principal component space. A convex hull defines the smallest area containing a set of points. We define convex hulls using the 10 clusters determined by Ward’s method for our \( z \sim 1.5 \) galaxy sample. In practice we disregard the 2 sparsely populated clusters and instead group all of those galaxies into the outlier class.

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

4.2 Morphological Error Estimation

The Hubble Ultra Deep Field (UDF) consists of deep imaging on a portion of the shallower GOODS-S field. We measure the same galaxies using different depth images to determine reliability of morphological measurements as a function of signal-to-noise and magnitude. The non-parametric morphologies of galaxies are measured both in
Figure 4. Group classification uncertainty, based on bootstrapped morphology measurement errors. Each galaxy’s non-parametric morphologies are randomly scattered based on gaussians with widths based on errors found in Fig. 3. The principal components and group membership are redetermined 250 times. The resulting MC group distributions for each originally defined group are shown. Groups 1, 6, and -1 are the most robust to measurement errors, whereas half of Groups 2, 4, 5, and 8 galaxies are scattered into other groups. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

the deep UDF region and the GOODS-S observations. We calculate the differences of GOODS-S morphologies from UDF morphologies. We then bin galaxies in magnitude to find the average difference and median absolute deviation, which we define as the error for that morphological measurement.

In Fig. 3, brighter galaxies are (unsurprisingly) well measured. The median absolute deviations (red error bars) show a majority of galaxies have statistics that do not vary widely between shallow and deep images. However, galaxies approaching the brightness limit of 24.5 display an increased median absolute deviation from UDF morphologies. In general, the morphological offsets seen in Fig. 3 are very small. There is large scatter in the differences measured for $M_{20}$ and $C$. (For similar study see fig. 6 Grogin et al. [2011])

Now that we have calculated the principal morphological components and resulting morphology groups, we can test their robustness to measurement errors. We use Monte Carlo resampling test to randomly scatter our initial morphological measurements by Gaussians with sigma equal to the median absolute deviation for each morphological measurement (Fig. 3). We then perform a principal component analysis for this new data set and repeat this process 250 times. We project the scattered data on the original PC weights and then classify the galaxies based on the originally defined convex hulls (§4.1) each time. The group a galaxy is classified into the most times is defined as the “Monte Carlo” (MC) group. Fig. 4 can thus be seen as the probability distribution function for a galaxy of a certain group to be classified into a group via the convex hull method. Group 6 is the most robustly classified group. Only group 4 galaxies are re-classified as such following the Monte Carlo scattering to less than a majority of times (however still a large plurality of times). The plots are separated by group and ordered.
Figure 5. Rest-frame $UVJ$ diagram for $M_*>10^{10}\,M_\odot$, $1.36<z<1.97$ galaxies for each group. A $UVJ$ diagram is used to separate quenched galaxies from star-forming galaxies (Williams et al. 2009). Quenched galaxies reside in the upper left trapezoid. Star-forming galaxies follow a sequence of increasing dust from the bottom left to the upper right. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1). The majority of quenched galaxies are in group 6, with some quenched galaxies found in groups 0 and 8. As PC1 decreases and PC2 increases we observe an increase in the fraction of quenched galaxies.

Table 1 shows that the most important principal components (PC1-3) have typical bootstrapped deviations \( \leq 10^{-15} \) per cent of their weights. Higher principal component dimension display greater variability, but are also less important to our group classifications.

5 PCA MORPHOLOGY GROUPS AT $Z \sim 1.5$

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

We use a $UVJ$ colour-colour diagram (Fig. 5) to classify galaxies as “star-forming” and “quenched” using the bimodality of these two types of galaxies seen in $U-V$ and $V-J$ rest-frame colours (Labbé et al. 2005; Wuyts et al. 2007; Williams et al. 2009). Star-forming galaxies follow a sequence determined by dust extinction. The figure is arranged so that PC1 increases along the x-axis and PC2 increases along the y-axis. Most groups are primarily comprised of star-forming galaxies, however groups with lower PC1 values are more compact and quenched. Similarly, a $UV-Mass$ diagram separate star-forming from quenched galaxies (Fig. 6). Again galaxies with lower PC1 values are more mas-
Figure 6. Rest-frame $U-V$ vs. Stellar Mass diagram for $M_*>10^{10}M_\odot$, 1.36 < z < 1.97 galaxies for each cluster group. Galaxies classified by $UVJ$ as star-forming (stars) and quenched (circles) are shown for each group. The dashed line in $U-V$ represents the approximate dividing line between quenched and star-forming galaxies. Groups 6, 0, and 9 have the greatest fractions of galaxies with large masses (dashed line, $M_*>5\times10^{10}M_\odot$). The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

Previous morphological efforts at high redshift (e.g. Lee et al. 2013) and PCA-morphology efforts at low redshift (e.g. Scarlata et al. 2007) use the simple classification scheme of disc-dominated, spheroid-dominated and irregular galaxy classes. However, such a classification scheme does not account for the rare and/or subtle features of galaxy structure. Massive galaxies at $z\sim2$ show a diversity of structures and star-formation properties (Bruce et al. 2012; Bell et al. 2012; Kriek et al. 2009; Lee et al. 2013). These structures do not always coincide with the local Hubble sequence. Disc-dominated galaxies appear clumpy (Förster Schreiber et al. 2009) and bulge-dominated galaxies can be compact, very red and massive, but possess no extended envelope (van Dokkum et al. 2008). For these reasons it is important to determine a new classification scheme unencumbered by the biases from low-z galaxies. Additionally, parametric morphological statistics (such as Sérsic-n) fit a function to the light profile of a galaxy but for disturbed galaxies these light profiles may be poorly defined and constrained.

We need a way to classify and group galaxies of similar structure together while also using the available quantitative morphological statistics. Non-parametric morphologies provide an avenue for studying the inherent shape galaxy without imposing an assumption on the symmetry or profile shape of a galaxy. Principal component analysis and hierarchical clustering allow us to find the galaxies that are the most similar in quantitative non-parametric morphologies and strips away any inherent degeneracy between the statistics. PCA has been used previously to characterize and classify $z<1$ galaxies (Scarlata et al. 2007) based on non-parametric morphologies.
Figure 7. $G - M_{20}$ for each group. Overplotted are the dividing lines between: mergers (top left corner), bulge-dominated (right-most region), and disc-dominated (bottom left region) modified from Lotz et al. (2004). Group 0 fully occupies the bulge-dominated region of the plot. Symbols same as Fig. 6. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

Previous studies (e.g. Lotz et al. 2004; Conselice et al. 2000; Lee et al. 2013) utilize $G - M_{20}$ (Fig. 7) or Concentration-Asymmetry (Fig. 8) diagrams to classify galaxies into early and late-type categories. In our study we use these tools to reinforce how effective our PCA groups are at separating different classes of galaxies. In Fig. 7 the dotted lines signify classification regions adapted from Lotz et al. (2004) for z~1-2 galaxies observed by HST. Mergers are in the upper left region, late-type galaxies are in the lower region and early-type galaxies are in the wedge-shaped region on the rightmost portion of the $G - M_{20}$ diagram. C – A diagrams (for review see Conselice 2014) have been used to differentiate giant ellipticals (which live in regions of large $C$ and small $A$) from spirals (with progressively smaller $C$ and larger $A$) and from ULIRGS (which are the most asymmetric but the least centrally concentrated).

For our group descriptions in the following sections we will refer heavily to Fig. 5 - 9, the example galaxies of Fig. 11 - 19 and Tables 2 - 5. For these figures the locations of each subplot represents the approximate position of that group in PCA space. From left to right, PC1 increases which is indicative of an increase in bulge strength. From bottom to top, PC2 increases thus concentration increases.

Tables 2 - 5 describe the group demographics in terms of stellar mass, visual classification (Kartaltepe et al. 2014), Sérsic indices (van der Wel et al. 2012) and quenched fraction. These demographics are both listed in terms of the original group (as determined by the hierarchical clustering method, left columns) and in terms of the MC group (determined using the scattering method, right columns). The agreement between the galaxy demographics in the original groups and scattered MC groups shows the group characteristics are quite robust to noise. Table 2 shows that high PC1 (disc-dominated) groups have very few high mass galaxies. Meanwhile, low PC1 (compact/bulge-dominated) groups have a larger fraction of high mass galaxies.

We use CANDELS visual classifications (Kartaltepe et al. 2014) to determine the demography of the PCA groups in disc, spheroidal and irregular galaxy classes. For a galaxy to be counted as a “disc”, “spheroid” or “irregular” it must have been classified by at least two-thirds of the classifiers as such, and less than one-third as the other classes. A “disc+spheroid” is classified as both a disc and a spheroid.
Figure 8. Concentration - Asymmetry for each group. Plotting symbols same as Fig. 6. Groups 9 and 1 have the highest asymmetry, while group 0 has the highest concentration. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

by at least two-thirds of the classifiers. The “other” class represents everything that does not belong to the other 4 categories. The fractions of galaxies in each morphological type are shown in Table 3.

Sérsic fits have been used extensively to classify galaxies into early- and late-type categories (van der Wel et al. 2012; van Dokkum et al. 2010; Patel et al. 2011; Peng et al. 2002). Typically, $n=2.5$ is used to divide late-type ($n < 2.5$) and early-type ($n > 2.5$) galaxies. Table 4 shows the percentage of galaxies representing a certain classification for each group as a percentage of the group population (van der Wel et al. 2012). Similar to visual classification, the percentage of galaxies with disc-dominated morphologies decreases with decreasing PC1 values.

Table 5 and Fig. 5 show that in this redshift range ($1.36 < z < 1.97$) and mass range ($\gtrsim 10^{10} M_\odot$) only 19 per cent of galaxies are quenched. Table 5 shows that the quenched fraction for a group is anti-correlated to PC1 and PC2.

Fig. 9 shows the effective radii (kpc) - stellar mass relation for each group. In this figure, PC1 and PC2 are strongly correlated a galaxy’s compactness. Group 6 galaxies are far the most compact, with the largest fraction of quenched galaxies. As PC1 and PC2 increase the number of quenched galaxies in each group decreases.

Group 6

Constituting 33 per cent of the entire sample, group 6 is by far the most populated group at $z \sim 1.5$ (example postage stamps in Fig. 11). Group 6 galaxies are characterized by their compact sizes ($r_e \sim 1.38 \pm 0.89$ kpc) and smooth features. Many of these galaxies are barely resolved by HST WFC3 which leads to their structureless appearance. Therefore, the structural properties of this group should be interpreted with caution, since it is possible that unresolved features in these galaxies would cause them to be classified as a different group if we had access to higher resolution observations. 50 per cent of the group is quenched, which represents 75 per cent of all quenched galaxies at this redshift. Groups 0 and 4 are the only other group with a >10 per cent fraction of quenched galaxies.

Group 6 galaxies also dominate the high mass galaxies.
represent the “compact” criteria \((M/r^2 < 10.3M_\odot \text{kpc}^{-1.5})\) of Barro et al. (2013). Almost all group 6 galaxies are very compact, with most galaxies smaller than 2 kpc. Groups 0 and 8 have a number of borderline compact galaxies. The remaining groups have only a few compact galaxies at most. The panels are roughly arranged by PC1 (increasing left to right) and PC2 (increasing bottom to top, except for group -1).

**Table 2.** Group percentages by mass range for both original group and “MC Group”

| Group | 10.0 < \(\log M_*\) < 10.5 | 10.5 < \(\log M_*\) < 11.0 | \(\log M_*\) > 11.0 | Group Percentage |
|-------|--------------------------|--------------------------|----------------------|------------------|
| Low PC1 | 6 | 54.9±4.3 | 58.5±4.6 | 41.4±3.7 | 38.2±3.8 | 3.8±1.4 | 3.3±1.4 | 33.2 (319) | 30.2 (290.4) |
| 0 | 44.4±7.0 | 45.9±5.7 | 41.7±6.9 | 40.1±8.2 | 13.9±3.5 | 14.0±4.4 | 11.2 (108) | 8.0 (77.1) |
| 9 | 69.7±11.3 | 65.1±7.5 | 21.2±7.0 | 29.2±5.3 | 9.1±3.6 | 5.7±2.8 | 6.9 (66) | 13.5 (129.9) |
| Mid PC1 | 4 | 68.8±10.3 | 66.0±11.9 | 31.2±7.4 | 29.3±8.0 | 0.0±2.3 | 2.1±3.6 | 8.0 (77) | 6.8 (65.6) |
| 8 | 69.8±7.9 | 71.2±9.3 | 27.8±5.3 | 26.8±6.1 | 2.4±1.4 | 2.0±3.6 | 13.1 (126) | 9.9 (95.3) |
| High PC1 | 1 | 62.9±6.9 | 61.8±8.1 | 36.0±7.2 | 35.2±6.4 | 1.1±2.5 | 3.0±2.7 | 9.3 (89) | 11.3 (108.7) |
| 2 | 81.0±12.4 | 67.8±9.8 | 17.5±6.7 | 29.8±6.9 | 1.6±3.5 | 2.4±3.0 | 6.5 (63) | 8.7 (84.0) |
| 5 | 75.5±9.2 | 67.3±8.7 | 23.5±5.6 | 30.0±5.2 | 1.0±2.2 | 2.7±1.6 | 10.6 (102) | 10.6 (101.8) |
| Outliers | -1 | 50.0±28.7 | 48.5±35.0 | 50.0±28.7 | 51.4±35.6 | 0.0±15.0 | 0.1±20.3 | 1.2 (12) | 0.9 (8.9) |

| N Galaxies | 600 | 323 | 39 | 962 |

Note: The left hand columns for each mass range represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications.
Table 3. Demographics of Visual Classifications of Groups

| Group | Discs | Spheroids | Irregulars | D-Sph | Other |
|-------|-------|-----------|------------|-------|-------|
| Low PC1 | 6     | 13.6±2.9  | 25.8±1.1  | 52.3±5.2 | 43.0±1.1 |
|       | 0     | 44.0±2.8  | 45.5±3.6  | 21.6±6.6  | 20.0±3.5 |
|       | 9     | 45.2±10.2 | 46.6±0.9  | 9.5±7.3   | 20.9±2.3 |

Mid PC1

|        |        |           |            |       |       |
|-------|-------|-----------|------------|-------|-------|
| 4     | 62.5±13.0 | 64.1±4.6  | 8.3±6.3   | 9.6±4.3 | 6.0±3.7 |
| 8     | 75.3±10.4 | 71.5±3.3  | 8.6±4.5   | 11.1±3.0 | 4.9±3.8 |

High PC1

|       |        |           |            |       |       |
|-------|-------|-----------|------------|-------|-------|
| 1     | 71.2±11.4 | 61.6±2.6  | 15.3±3.4  | 11.1±2.4 | 16.7±2.6 |
| 2     | 76.3±16.6 | 68.0±1.4  | 7.9±4.6   | 18.7±1.1 | 5.3±1.8 |
| 5     | 95.3±13.2 | 71.4±1.3  | 0.0±0.8   | 10.2±0.9 | 3.1±2.3 |

Outliers

|        |        |           |            |       |       |
|-------|-------|-----------|------------|-------|-------|
| -1    | 50.0±38.1 | 40.0±31.1 | 25.0±31.9 | 34.0±31.1 | 12.5±27.9 |

Total Fraction

|        | 49% (313) | 24% (149) | 5% (35)  | 15% (96)  | 7% (42)  |

Note: Visual classification from Kartaltepe et al. [2014] for UDS and GOODS-S (no classifications for COSMOS galaxies). For a galaxy to be visually classified 2/3 observers need to agree. ‘Other’ classification refers to galaxies failing the 2/3 agreement requirement. The left hand columns for each visual classification represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications.

Table 4. Demographics of Sérsic Classifications of Groups

| Group | 0 < n < 1 | 1 < n < 2.5 | 2.5 < n < 4 | n > 4 |
|-------|-----------|-------------|-------------|-------|
| Low PC1 | 6     | 37.9±4.5  | 21.5±11.1  | 22.9±3.6  | 35.7±1.2  | 21.6±3.5  | 29.1±1.2  | 10.1±2.5  | 14.0±1.1  |
|       | 0     | 31.5±13.3 | 10.5±3.7   | 23.1±6.7  | 36.0±3.8  | 18.5±6.2  | 26.1±3.8  | 19.4±6.3  | 27.7±2.8  |
|       | 9     | 59.1±13.8 | 33.9±12.6  | 25.8±10.9 | 41.6±2.7  | 3.0±5.5   | 13.9±2.5  | 4.5±6.0   | 10.6±0.8  |
| Mid PC1 | 4     | 51.9±12.1 | 39.2±10.3  | 27.3±9.4  | 43.0±5.1  | 6.5±5.9   | 10.7±5.0  | 6.5±5.9   | 7.1±5.0   |
|       | 8     | 61.9±4.6  | 49.7±1.3   | 20.6±6.1  | 35.2±3.4  | 4.8±3.7   | 10.9±3.3  | 16.4±2.9  | 4.2±1.3   |
| High PC1 | 1     | 82.0±12.1 | 68.5±3.0   | 3.4±4.0   | 12.9±2.8  | 0.0±2.7   | 10.9±2.8  | 11.3±2.7  | 7.6±2.8   |
|       | 2     | 69.8±15.6 | 43.8±3.9   | 14.3±8.8  | 33.2±3.9  | 1.6±5.5   | 12.2±1.3  | 3.2±6.1   | 11.6±3.8  |
|       | 5     | 86.3±12.6 | 68.3±4.3   | 4.9±4.5   | 14.5±1.2  | 0.9±2.8   | 9.5±3.1   | 6.1±2.8   | 7.6±1.1   |
| Outliers | -1    | 83.3±44.5 | 66.0±3.1   | 0.0±2.5   | 0.0±8.6   | 0.0±2.25  | 0.0±2.0   | 16.7±29.3 | 34.0±31.1 |

Total Fraction

|        | 60% (527) | 20% (179) | 12% (103)  | 8% (68)  |

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

at this epoch, constituting 41 per cent of galaxies with 5 × 10^{10} M_{\odot} < M_\star < 10^{11} M_{\odot} and 31 per cent of galaxies with M_\star > 10^{11} M_{\odot}.

Group 6 galaxies have low concentrations (C ~ 2.97 ± 0.42), moderate Gini coefficients (G ~ 0.51 ± 0.03), low M_{20} (~ -1.64 ± 0.14), extremely low MID values (M = 0.00, I ~ 0.00 ± 0.02, D ~ 0.07 ± 0.04), and low asymmetry values (A ~ 0.06 ± 0.06). The G ~ M_{20} diagram classifies the majority of these galaxies as borderline disc/spheroidal (with occasional irregular classification). However, M_{20} values are potentially biased because the 20 per cent light is not resolved. These galaxies have large average Sérsic indices (\bar{n} ~ 2.75).

Group 6 is comprised of the highest percentage of visually identified spheroids (52 percent) and disc+spheroids (22 percent), and also has the lowest percentage of discs (13 percent) of any group.

Group 0

Group 0 galaxies are characterized by a strong bulge component which is surrounded by a faint smooth extended component (example postage stamps in Fig. [12]). A significant fraction of group 0 galaxies are quenched galaxies (22 percent; Table [3]). Although group 0 galaxies make up only 11 per cent of the galaxies in the sample, they constitute 38 per cent of the galaxies more massive than 10^{11} M_{\odot} (Table [2]).

These galaxies have high concentration values (C ~ 3.99
Table 5. UVJ Quenched Fractions of Groups

| Group | Quenched | Star-Forming |
|-------|----------|--------------|
| Low PC1 | 6 | 45.8±3.9 | 37.3±3.8 | 54.3±4.2 | 62.4±4.7 |
| | 0 | 18.5±4.9 | 17.7±4.2 | 81.7±8.2 | 82.3±11.1 |
| | 9 | 4.5±3.4 | 16.2±3.4 | 95.5±13.0 | 83.7±8.4 |
| Mid PC1 | 4 | 10.4±3.9 | 9.3±3.6 | 89.6±11.6 | 90.7±12.7 |
| | 8 | 7.1±3.1 | 7.8±3.1 | 92.9±9.0 | 92.2±10.4 |
| High PC1 | 1 | 1.1±2.5 | 10.4±3.9 | 98.9±11.2 | 89.5±9.6 |
| | 2 | 0.0±0.8 | 12.7±3.8 | 100.0±12.0 | 87.1±9.7 |
| | 5 | 1.0±1.1 | 9.1±2.9 | 99.0±9.4 | 90.8±9.0 |
| Outliers | -1 | 0.0±15.0 | 0.0±20.2 | 100.0±36.4 | 100.0±44.2 |
| Total Fraction | 20% (188) | 80% (773) |

Note: Quenched/star-forming classifications based on Fig. 3. The left hand columns for quenched/star-forming classifications represent the demographics based upon the original group based on hierarchical clustering. The right hand columns are based on the total group probabilities based on the scattering technique classifications.

This group is the most visually irregular group (26 percent), and has a relatively low disc fraction (45 percent), spheroid fraction (9 percent) and disc+spheroid fraction (14 percent). These statistics and visual classifications imply many galaxies have bright off-center clusters, in addition to bright central bulges.

Group 9

Group 9 is characterized by their asymmetric, irregular morphologies and strong bulge component (example postage stamps in Fig. 13). These galaxies make up a significant portion of the M* > 10^11 M⊙ galaxies (15 percent). However, most of these galaxies are lower mass (M* < 3 × 10^10 M⊙). Only 6 per cent of group 9 galaxies are quenched.

These galaxies have moderate concentrations (C ~ 3.70 ± 0.62), moderate Gini coefficient (G ~ 0.51 ± 0.05), moderate M20 (~ -1.44 ± 0.25), moderate MID values (M ~ 0.15 ± 0.16, I ~ 0.25 ± 0.19, D ~ 0.20 ± 0.10) and high asymmetry (A ~ 0.22 ± 0.12). These galaxies lie along the G = M20 merger/disc galaxy dividing line and also overlap with the spheroidal region.

Group 9 galaxies have large radii (re ~ 4.30 ± 2.24 kpc) and moderately low average Sérsic indices (n ~ 1.82). This group contains some quenched galaxies (~11 percent). Some galaxies are extended and also quenched; meaning they are rare “red disc” population. None of the group 4 galaxies are more massive than M* > 10^11 M⊙. Primarily these galaxies are lower mass (M* < 3 × 10^10 M⊙).

Group 4 has moderate concentrations (C ~ 3.77 ± 0.66), moderate Gini coefficients (G ~ 0.49 ± 0.04), high M20 (~ -1.14 ± 0.22), low intensities (I ~ 0.06 ± 0.07), small multi-mode values (M ~ 0.07±0.08), low deviations (D ~ 0.10 ± 0.07), and low asymmetry (A ≲ 0).

Group 4 galaxies have moderate effective radii (re ~ 3.43 ± 1.09 kpc) and medium average Sérsic indices (n ~ 2.20).

Group 4 members are primarily visually classified as discs (62 percent) or disc+spheroids (21 percent) and are far less classified as spheroids (8 percent) or irregulars (0 percent).
stamps in Fig. [15]. This class is dominated by low-mass star-forming galaxies, but also includes low-mass ($< 3 \times 10^{10} M_\odot$) quenched galaxies ($\sim$ 8 percent). Very few galaxies have stellar masses > $5 \times 10^{10} M_\odot$.

Group 8 galaxies have small concentrations ($C \sim 3.07 \pm 0.42$), moderate Gini coefficients ($G \sim 0.46 \pm 0.03$), moderate $M_{20}$ ($\sim -1.56 \pm 0.16$), low but non-zero MIDs ($M \sim 0.06 \pm 0.06$, $I \sim 0.10 \pm 0.11$, $D \sim 0.09 \pm 0.05$), and low asymmetry values ($A \sim 0.08 \pm 0.06$). On the $G - M_{20}$ diagram these galaxies fall within the disc-dominated region but are close to the spheroidal/disc dividing line.

Sérsic fits to this class find moderate sizes ($r_e \sim 3.41 \pm 1.72$ kpc) and low average Sérsic indices ($\bar{n} \sim 1.31$).

Group 8 is dominated by visually-classified discs (75 percent) with only a modest fraction of spheroids (9 percent). A small number of galaxies are quenched and compact which overlaps with groups 0 and 6.

**Group 1**

Group 1 galaxies are primarily large discs and irregulars with bright off-center star-forming knots (example postage stamps in Fig. [16]). Only 2 per cent of these galaxies are quenched based on their $UVJ$ colours. The distribution of masses is heavily weighted towards lower mass galaxies with very few objects more massive than $3 \times 10^{10} M_\odot$.

Group 1 galaxies have low concentration values ($C \sim 2.74 \pm 0.39$), low Gini coefficients ($G \sim 0.43 \pm 0.04$), high $M_{20}$ ($\sim -1.07 \pm 0.17$), moderately high asymmetry values ($A \sim 0.13 \pm 0.11$), large multi-mode values ($M \sim 0.43 \pm 0.27$), high deviations ($D \sim 0.37 \pm 0.13$) and large intensities ($I \sim 0.63 \pm 0.23$). The high $A$ and MIDs statistics indicate many of these galaxies have bright off-center clusters and are potentially irregular.

The visual classifications and Sérsic indices primarily classify this group as disc galaxies and/or irregulars. Group 1 is dominated by visually-classified discs (71 percent) and has a relatively large fraction of irregulars (16 percent). This group has very few spheroids or bulge-dominated disc galaxies. Their effective radii are large for this redshift and masses ($r_e \sim 5.13 \pm 1.60$ kpc). This group has low average Sérsic indices ($\bar{n} \sim 0.65$) imply a large disc and irregular population.

**Group 2**

Group 2 galaxies are primarily low-mass, star-forming, smooth disc galaxies with high central concentrations and few visually detected star-forming knots (example postage stamps in Fig. [17]). Very few of these galaxies are quenched (3 percent). The mass distribution for this group is a steeply declining function where there are only a few galaxies with masses > $3 \times 10^{10} M_\odot$.

Group 2 galaxies have large concentrations ($C \sim 4.84 \pm 0.62$), low Gini coefficients ($G \sim 0.45 \pm 0.04$), moderate $M_{20}$ ($\sim -1.23 \pm 0.24$), low asymmetry ($A \sim 0.06 \pm 0.08$), low deviations ($D \sim 0.16 \pm 0.09$), moderate multi-modes ($M \sim 0.15 \pm 0.19$), and a wide spread of intensity values ($I \sim 0.30 \pm 0.32$). On the $G - M_{20}$ diagram these galaxies fall within the disc-dominated and irregular portion of the diagram. However, their high $C$ values suggest a bright nuclear component.

The visual classifications show this group is dominated by discs (76 percent) and only small fractions of irregular galaxies (~5 percent) and disc+spheroid galaxies (~8 percent). They have mid-sized effective radii ($r_e \sim 3.58 \pm 0.85$ kpc) and mid-to-low average Sérsic indices ($\bar{n} \sim 1.13$).

**Group 5**

Group 5 galaxies are primarily low-mass, star forming, extended disc galaxies with a weak bulge component (example postage stamps in Fig. [18]). This group has a negligible fraction of quenched galaxies (~1 percent). These galaxies are primarily lower mass ($M_{20} < 3 \times 10^{10} M_\odot$).

Group 5 is mostly comprised of low concentration values ($C \sim 2.91 \pm 0.41$), low Gini coefficients ($G \sim 0.40 \pm 0.03$), low/moderate $M_{20}$ ($\sim -1.20 \pm 0.18$), a wide spread in multi-mode ($M \sim 0.25 \pm 0.24$), large intensity values ($I \sim 0.51 \pm 0.28$), low deviation values ($D \sim 0.12 \pm 0.05$), and low asymmetry values ($A \sim 0.03 \pm 0.12$). On the $G - M_{20}$ diagram these galaxies fall solidly within the disc-dominated region.

The defining feature of this group is its large typical size ($r_e \sim 5.35 \pm 1.58$ kpc). Group 5 galaxies have very low average Sérsic indices ($\bar{n} \sim 0.64$); implying a disc-dominated/irregular population.

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

**Group -1**

The original groups 3 and 7 were comprised of only a few galaxies each (12 in total, example postage stamps in Fig. [19]). They were outliers from the remaining groups and are combined into a single outlier group. These galaxies are most likely outliers because they have at least one poorly measured (or missing) morphological parameter (especially the multi-mode statistic). These galaxies have a low surface brightness, very large radii ($r_e \sim 6.55 \pm 2.74$ kpc), very low concentration ($C \sim 1.92 \pm 0.85$), high intensity ($I \sim 0.49 \pm 0.39$), high $M_{20}$ ($\sim -1.02 \pm 0.24$), low Gini coefficient ($G \sim 0.36 \pm 0.08$), extremely high deviations ($D \sim 0.71 \pm 0.44$) and high multi-modes ($M \sim 0.51 \pm 0.39$). The deviation values can separate group -1 galaxies from all the other groups.

6 DISCUSSION

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

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

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

### 6.1 Stellar Mass - Quenching Connection for groups

Fig. 10 shows the cumulative distribution of the quenched fraction rank-ordered by “compactness” ($M/r_1^{1.5} < 10^9 M_\odot$ kpc$^{-1.5}$). Barro et al. 2013, Stellar mass, Sérsic-n, PC1, PC2 and PC3. For every galaxy we assign a 0 to star-forming galaxies and 1/1$N_{\text{quenched}}$ for quenched galaxies and these values are cumulatively summed. We observe a flat trend in stellar mass and PC2, and a much steeper trend in PC1 and Sérsic-n. Sérsic-n has previously been shown to correlate well with quenching (e.g. Bell et al. 2012). The similarities in steepness between the PC1 and Sérsic-n curves show PC1 is an equivalently useful predictor of quenching.

We also investigate the relationship between quenching and PC1 through the colour-mass relation. In Fig. 9 we observe a correlation between the increase in the fraction of massive galaxies ($> 5 \times 10^9 M_\odot$) for a specific group and the magnitude of PC1 (bulge strength). The amount of quenched galaxies correlates more strongly with PC1 (bulge strength) than PC2 (concentration). Similar results have been found for $z \sim 1$-2 galaxies (Bell et al. 2012; Barro et al. 2013; Lang et al. 2014). Unsurprisingly, the most massive galaxies are also the most likely to be quenched. For instance, group 6 has the largest amount of red galaxies and many massive galaxies ($> 5 \times 10^9 M_\odot$). The only other groups with a substantial number of quiescent galaxies are groups 0 (22 percent) and 4 (11 percent). Group 0 galaxies are primarily bulge-dominated with a faint disc. However, group 6 galaxies are more massive ($> 5 \times 10^9 M_\odot$) than group 0 galaxies. Furthermore, a much larger percentage of these massive galaxies in group 6 are quenched. Group 9 galaxies are slightly less massive ($M_\odot \sim 10^4$), but still generally have a strong bulge component (as determined by PC1).

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

### 6.2 The relationship between PCA Classes and Visual/Sérsic Classifications

PCA, in conjunction with our group finding algorithm, provides a distinct picture of galaxy structure from Sérsic index and visually based classifications. This is especially apparent for bulge-dominated systems. Across many groups we see a deficit of galaxies with Sérsic-$n > 2.5$ when compared to the amount of PC-classified spheroid and disc+spheroid galaxies in our sample. In Table 4 we observe a trend of decreasing PC1 and increasing number of high Sérsic galaxies.

This classification scheme also separates quenched compact galaxies (group 6) from larger, smooth proto-elliptical systems (group 0), and star-forming disc-dominated clumpy galaxies (group 1) from star-forming bulge-dominated asymmetric galaxies (group 9). Separating clumpy star-formers and bulge dominated star forming galaxies has great importance for understanding the mechanisms that formed these galaxies and the potential avenues for evolution available to them.

Our groups belong to 3 distinct types: the “disc-dominated” galaxies of groups 1, 2, and 5; the “compact/bulge-dominated” galaxies of groups 0, 6, and 9; and the “intermediate” galaxies of groups 4 and 8.

---

**Figure 10.** Cumulative quenched fraction rank ordered by various metrics: PC1, PC2, PC3, stellar mass, Sérsic-n and “compactness”. The green solid line represents no correlation between quenched fraction and rank. Sérsic-n and PC1 have a similar CQF compactness. The green solid line represents no correlation between quenched fraction and rank. Sersic-n and PC1 have a similar CQF.
purposes of our discussion we refer the reader to Fig. 11-19 and Tables 2-5.

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

Galaxies in groups 0, 6, and 9 display a variety of visual classifications, but have a single unifying characteristic: many of these galaxies are bulge-dominated. Group 6 galaxies are very small and compact ($r_e \sim 1.38 \pm 0.89$ kpc) with no discernible stellar envelope. Group 0 galaxies are slightly larger ($r_e \sim 3.20 \pm 2.08$ kpc) than group 6, and display evidence for an extended stellar envelope. While group 9 galaxies are the most visually disturbed group (26 per cent Irregular), they also have bright central bulges as found by PC1 (but not by visual classifications and Sérsic indices).

Group 6 galaxies are visually classified as bulge-dominated (either pure spheroid or disc+spheroid morphology) >74 per cent of the time. However, a Sérsic cut of $n > 2.5$ yields only 31 per cent. While PC1 is a strong indicator of a compact core, it cannot measure the Sérsic profile for radii smaller than the PSF. Similarly for group 0 galaxies, 54 per cent of galaxies are visually classified as bulge-dominated, but a Sérsic classification only indicates 38 per cent are bulge-dominated galaxies.

Groups 0 and 6 display some distinguishing characteristics. Group 6 galaxies are actually less concentrated ($C \sim 2.97 \pm 0.42$) than those in group 0 ($C \sim 3.99 \pm 0.87$). Although the small sizes and lower concentrations for group 6 are due to the fact that $r_{20}$ can be smaller than the PSF of the instrument and are not resolved or just barely resolved.

The size-mass (Fig. 9) relation for these two groups is different as well. Group 6 galaxies are smaller but still as massive as group 0 galaxies. Thus many group 6 galaxies are compact using the Barro et al. (2013) definition. Most of the compact galaxies in group 6 are quenched, whereas the quenched galaxies of group 0 are more extended (only a few are compact). Group 6 also has a number of compact, star-forming galaxies, while only a few group 0 galaxies are both compact and forming stars.

Classifications based on PCs provide a slightly different picture from those based on Sérsic-$n$ or visual inspection. A PC classification determines ~51 per cent of galaxies are compact/bulge-dominated (groups 0, 6 and 9) while visual classifications determine ~39 per cent of galaxies are bulge-dominated (either pure spheroids or disc+spheroids) and Sérsic indices classify ~20 per cent of galaxies as bulge-dominated ($n > 2.5$). The differences between the classification schemes are subtle but important because they mean each is probing a slightly different subset of galaxies.

The compact/bulge-dominated nature and high masses of these 3 groups could imply an evolutionary connection. In this scenario, galaxies begin as group 6 galaxies, a naked core with no extended envelope or structure. Following a gas-rich merger disturbed tidal features become visible and the galaxy becomes classified as group 9. After a sufficient time for the gas to settle in a disc or spheroidal envelope ($\gtrsim 1.5$ Gyr) the galaxy would appear as a group 0 galaxy.

6.2.2 The Disc-dominated Galaxies: Groups 1, 2 and 5

Groups 1, 2 and 5 all have an overwhelmingly large percentage of visually classified disc galaxies (71 per cent, 76 per cent, and 95 per cent respectively). Sérsic classifications largely agree with the visual classifications for these groups. The only difference is that Sérsic classifications yield more disc-dominated galaxies ($1 < n < 2.5$) than visual classifications would indicate. Non-parametric morphologies determine these disc galaxies have varying degrees of clumpiness and disturbances.

Group 1 galaxies are the most disturbed of the “disc-dominated” groups. They have the largest asymmetries ($A \sim 0.13 \pm 0.11$), multi-modes ($M \sim 0.43 \pm 0.27$), intensities ($I \sim 0.63 \pm 0.23$) and deviations ($D \sim 0.37 \pm 0.13$). They are more often visually classified as irregular (16 percent), but have a weaker bulge component (indicated by their larger $M_{20}$ values, $\sim 1.07 \pm 0.17$) than groups 2 and 5.

Of the remaining disc-dominated groups, group 5 galaxies have much higher $M$ and $I$ statistics ($M \sim 0.25 \pm 0.24$ and $I \sim 0.51 \pm 0.28$) than those in group 2 ($M \sim 0.15 \pm 0.19$ and $I \sim 0.3 \pm 0.32$). However, these two groups have similar asymmetry values ($A \sim 0.05$), $M_{20}$ values ($\sim 1.2$), and deviations ($D \sim 0.1$).

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

6.2.3 The Intermediate Galaxies: Groups 4 and 8

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

Both groups 4 and 8 have a large fraction of galaxies with $n < 2.5$ (79 per cent and 82 per cent, respectively). However, group 8 galaxies are more likely to be visually classified as discs than group 4 galaxies (75 per cent compared to 62 percent). Meanwhile, group 4 galaxies are more likely be visually classified as bulge-dominated (28 per cent compared to 16 percent). However, the differences between groups should taken with caution as the small numbers of galaxies in these groups reduces the significance of the percentages.

For groups 4 and 8 the classifications based upon non-parametric morphologies do not always agree with classifications based on Sérsic indices or visual inspection. Group 8 has a much smaller average $M_{20}$ value ($M_{20} \sim 1.56 \pm 0.16$) than group 4 ($M_{20} \sim 1.14 \pm 0.22$). This indicates the bulges of group 8 galaxies are large and possibly dominate.
the morphology. However, Sersic indices and visual classifications would suggest there is no sizable bulge component for most of these galaxies. Group 4 galaxies have high concentrations, low Sersic indices and are the least well defined group by bootstrap measures (see Fig. 11). Meanwhile, the \( G-M_{20} \) diagram suggests a population of irregular galaxies while visual classifications find no irregular galaxies. The bright nuclear components may be the result of an AGN or starburst activity.

6.2.4 Comparing the Irregular Galaxies of Groups 1 and 9

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

Group 9 galaxies display tidal features and irregular discs but their strong central bulge is missed by Sersic fits. Group 9 itself, shows the power of our PCA classifications to find interesting subtypes of galaxy morphology. Group 9 galaxies are visually classified as discs (45 percent), irregulars (26 percent) and bulge-dominated discs (14 percent). However, small PC1 values would indicate group 9 galaxies possess a strong central bulge. Meanwhile, group 1 galaxies much more likely to be visually classified as a pure disc galaxy (71 percent), slightly less likely to be irregular (16 percent) and much less likely to be bulge-dominated (1 percent). Group 1 galaxies also have higher PC1 values, indicating a weaker bulge component. Using Sersic index classifications, both groups 1 and 9 have a very large fraction of these galaxies are disc-dominated (85 percent) as opposed to bulge-dominated (15 percent). Groups 1 and 9 would be considered very similar in a Sersic classification and the differences between these groups are more subtle.

We observe subtle differences between these two groups in many statistics; group 9 galaxies are more asymmetric (0.22 ± 0.12 vs. 0.13 ± 0.11) and have lower \( M_{20} \) values (-1.44 ± 0.25 vs. -1.07 ± 0.17) than galaxies found in group 1. Group 9 galaxies are also more concentrated (3.7 ± 0.62 vs. 2.74 ± 0.39). Meanwhile \( M, I \) and \( D \) statistics all display an increased enhancement in group 1 galaxies because these statistics probe the existence of off-center clumps.

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

7 SUMMARY

We use a principal component analysis of non-parametric morphology measurements (\( G, M_{20}, C, A, M, I \) and \( D \)) and agglomerative hierarchical clustering to group galaxies into a more descriptive schema than the traditional spiral, elliptical, and irregular categories. The PCA weights we calculate (Table 1) show that non-parametric morphological correlations vary in importance: PC1 is based upon \( M, I, D, M_{20} \) and Gini thus it is interpreted as a bulge strength indicator; PC2 is dominated by concentration; and PC3 is dominated by asymmetry; the remaining PCs are less important and difficult to interpret.

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

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

- **Compact or Bulge-dominated/low PC1, ~ 51 per cent**
  - Group 6: Most populated group (~33 per cent of sample, examples seen in Fig. 11). Very compact and most massive galaxies; and contains the largest spheroidal (based on Sersic and visual classifications) and quenched fraction.
  - Group 0: Large bulge+disc population, has prominent bulge with faint disc component. (~11 per cent, Fig. 12). Contains a sizable fraction of massive and quenched galaxies, not to the same extent as group 6 however.
  - Group 9: Large and massive galaxies with a substantial irregular population. Visually, these galaxies possess tidal tails, bright star-forming knots and a large bulge. (~7 per cent, Fig. 13)

- **Bulge+Disc/intermediate PC1, ~ 21 per cent**
  - Group 4: Galaxy population of both small, compact spheroids and larger, bulge-dominated disc galaxies. (~8 per cent, Fig. 14)
  - Group 8: Small, primarily disc galaxies without much evidence of disturbances. (~13 per cent, Fig. 15)

- **Disc-dominated/high PC1, ~ 26 per cent**
  - Group 1: Large galaxies with prominent (albeit) irregular discs. (~9 per cent, Fig. 16)
  - Group 2: Compact and small discs galaxies. (~6 per cent, Fig. 17)
  - Group 5: Large and low mass disc galaxies with evidence of disturbances and interactions. (~10 per cent, Fig. 15)
Group -1, \( \sim 1 \) per cent: Low surface brightness galaxies (Fig. 19) with outlier PC values.

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

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

ACKNOWLEDGEMENTS

This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA Hubble Space Telescope. Support for HST Programs GO-12060 and GO-12099 was provided by NASA through grants from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support also came from the HST Archival program through HST-AR-12856. We also acknowledge support from the STScI DDRF. We would like to thank Harry Ferguson and Sandy Faber for their insightful input and many fruitful discussions.

REFERENCES

Abraham R. G., van den Bergh S., Nair P., 2003, ApJ, 588, 218
Barro G., Faber S. M., Pérez-González P. G., et al., 2013, ApJ, 765, 104
Bell E. F., van der Wel A., Papovich C., et al., 2012, ApJ, 753, 167
Bershady M. A., Jangren A., Conselice C. J., 2000, AJ, 119, 2645
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brinchmann J., Charlot S., White S. D. M., et al., 2004, MNRAS, 351, 1151
Brown M. J. I., Moustakas J., Smith J.-D. T., et al., 2014, ApJS, 212, 18

Bruce V. A., Dunlop J. S., Cirasuolo M., et al., 2012, MNRAS, 427, 1666
Bruce V. A., Dunlop J. S., McLure R. J., et al., 2014a, MNRAS, 444, 1001
Bruce V. A., Dunlop J. S., McLure R. J., et al., 2014b, MNRAS, 444, 1660
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., et al., 2000, ApJ, 533, 682
Chabrier G., 2003, PASP, 115, 763
Cisternas M., Jahnke K., Insikp K. J., et al., 2011, ApJ, 726, 57
Conselice C. J., 2014, ArXiv e-prints
Conselice C. J., Bershady M. A., Jangren A., 2000, ApJ, 529, 886
Conselice C. J., Chapman S. C., Windhorst R. A., 2003, ApJ Lett., 596, L5
Dahlen T., Mobasher B., Faber S. M., et al., 2013, ApJ, 775, 93
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
Förster Schreiber N. M., Genzel R., Bouché N., et al., 2009, ApJ, 706, 1364
Freeman P. E., Izbicki R., Lee A. B., et al., 2013, MNRAS, 434, 282
Galametz A., Grazian A., Fontana A., et al., 2013, ArXiv e-prints
Grogin N. A., Kocevski D. D., Faber S. M., et al., 2011, ApJS, 197, 35
Guo Y., Ferguson H. C., Bell E. F., et al., 2015, ApJ, 800, 39
Guo Y., Ferguson H. C., Giavalisco M., et al., 2013, ApJS, 207, 24
Hopkins P. F., Bundy K., Croton D., et al., 2010, ApJ, 715, 202
Hopkins P. F., Hernquist L., 2009, ApJ, 694, 599
Ivezić Z., Connolly A., Vanderplas J., Gray A., 2013, Statistics, Data Mining and Machine Learning in Astronomy. Princeton University Press
Kampczyk P., Lilly S. J., Carollo C. M., et al., 2007, ApJS, 172, 329
Kartaltepe J. S., Mozena M., Kocevski D., et al., 2014, ArXiv e-prints
Koekemoer A. M., Faber S. M., Ferguson H. C., et al., 2011, ApJS, 197, 36
Kormendy J., Kennicutt Jr. R. C., 2004, ARA&A, 42, 603
Kriek M., van Dokkum P. G., Franx M., et al., 2009, ApJ Lett., 705, L71
Labbé I., Huang J., Franx M., et al., 2005, ApJ Lett., 624, L81
Lang P., Wuyts S., Somerville R. S., et al., 2014, ApJ, 788, 11
Lee B., Giavalisco M., Williams C. C., et al., 2013, ApJ, 774, 47
Lintott C. J., Schawinski K., Slosar A., et al., 2008, MNRAS, 389, 1179
Lorenz M. O., 1905, Publications of the American Statistical Association, Volume 9, Number 70, p. 209-219, 9,

© 2015 RAS, MNRAS 000, 0-27
Wuyts S., Labbé I., Franx M., et al., 2007, ApJ, 655, 51

van der Wel A., Franx M., van Dokkum P. G., et al., 2014, ApJ, 788, 28
van Dokkum P. G., Franx M., Kriek M., et al., 2008, ApJ Lett., 677, L5
van Dokkum P. G., Whitaker K. E., Brammer G., et al., 2010, ApJ, 709, 1018
Villforth C., Hamann F., Koekemoer A., et al., 2013, ArXiv e-prints
Williams R. J., Quadri R. F., Franx M., et al., 2009, ApJ, 691, 1879
Wuyts S., Förster Schreiber N. M., van der Wel A., et al., 2011, ApJ, 742, 96
Figure 11. Group 6 F125W $1.36 < z < 1.97$ galaxies, shown in F160W/F125W/F814W RGB 6'' x 6'' postage stamps. $p(\text{Group})$ represents the percentage of times a galaxy is classified into group 6 after the scattering test. Very compact and small spheroidal galaxies. This group contains the largest spheroidal and quenched fraction. Many of these galaxies are barely resolved which leads to their structureless appearance.
Figure 12. Group 0: These galaxies are characterized by a strong bulge component surrounded by a fainter smooth disc.

Figure 13. Group 9: These galaxies are characterized by their asymmetric, irregular morphologies and strong bulge component.
Figure 14. Group 4: These galaxies consist of low-mass smooth galaxies with moderate central concentrations.

Figure 15. Group 8: These galaxies represent class of bulge-disc systems with dominant smooth discs.
Figure 16. Group 1: These galaxies are primarily large discs and irregulars with bright off-center star-forming knots.

Figure 17. Group 2: These galaxies appear to be primarily low-mass star-forming disc galaxies with higher central concentrations and few detected star-forming knots.
Figure 18. Group 5: Many of these galaxies are low-mass extended star forming disc galaxies with weak (if any) bulge components.

Figure 19. Group -1: Low surface brightness galaxies originally in groups 3 and 7, and only have a combined 12 galaxies which are outliers from all other groups.