Synergies between low- and intermediate-redshift galaxy populations revealed with unsupervised machine learning

Sebastian Turner, Malgorzata Siudek, Samir Salim, Ivan K. Baldry, Agnieszka Pollo, Steven N. Longmore, Katarzyna Malek, Chris A. Collins, Paulo J. Lisboa, Janusz Krywult, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **503,** 3010–3031 (2021) doi:10.1093/mnras/stab653
Advance Access publication 2021 March 8

Synergies between low- and intermediate-redshift galaxy populations revealed with unsupervised machine learning

Sebastian Turner [1]★ Malgorzata Siudek [2,3]★ Samir Salim,[4] Ivan K. Baldry [1] Agnieszka Pollo,[3,5] Steven N. Longmore,[1] Katarzyna Malek,[3,6] Chris A. Collins,[1] Paulo J. Lisboa,[7] Janusz Krywult,[8] Thibaud Moutard [9] Daniela Vergani[10] and Alexander Fritz[11]

[1]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[2]*Institut de Física d'Altes Energies, The Barcelona Institute of Science and Technology, E-08193 Bellaterra (Barcelona), Spain*
[3]*National Centre for Nuclear Research, ul. Hoza 69, PL-00-681 Warsaw, Poland*
[4]*Department of Astronomy, Indiana University, Bloomington, IN 47405, USA*
[5]*Astronomical Observatory of the Jagiellonian University, ul. Orla 171, PL-30-244 Kraków, Poland*
[6]*Aix Marseille Univ. CNRS, CNES, F-13388 LAM Marseille, France*
[7]*Department of Applied Mathematics, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK*
[8]*Institute of Physics, Jan Kochanowski University, ul. Swietokrzyska 15, PL-25-406 Kielce, Poland*
[9]*Department of Astronomy & Physics and the Institute for Computational Astrophysics, Saint Mary's University, 923 Robie Street, Halifax, NS B3H 3C 3, Canada*
[10]*INAF - OAS Bologna, Via P. Gobetti 93, I-40129 Bologna, Italy*
[11]*Lichtenbergstraße 8, D-85748 Garching, Germany*

Accepted 2021 March 2. Received 2021 February 8; in original form 2020 October 5

ABSTRACT

The colour bimodality of galaxies provides an empirical basis for theories of galaxy evolution. However, the balance of processes that begets this bimodality has not yet been constrained. A more detailed view of the galaxy population is needed, which we achieve in this paper by using unsupervised machine learning to combine multidimensional data at two different epochs. We aim to understand the cosmic evolution of galaxy subpopulations by uncovering substructures within the colour bimodality. We choose a clustering algorithm that models clusters using only the most discriminative data available, and apply it to two galaxy samples: one from the second edition of the GALEX-SDSS-WISE Legacy Catalogue (GSWLC-2; $z \sim 0.06$), and the other from the VIMOS Public Extragalactic Redshift Survey (VIPERS; $z \sim 0.65$). We cluster within a nine-dimensional feature space defined purely by rest-frame ultraviolet-through-near-infrared colours. Both samples are similarly partitioned into seven clusters, breaking down into four of mostly star-forming galaxies (including the vast majority of green valley galaxies) and three of mostly passive galaxies. The separation between these two families of clusters suggests differences in the evolution of their galaxies, and that these differences are strongly expressed in their colours alone. The samples are closely related, with star-forming/green-valley clusters at both epochs forming morphological sequences, capturing the gradual internally driven growth of galaxy bulges. At high stellar masses, this growth is linked with quenching. However, it is only in our low-redshift sample that additional, environmental processes appear to be involved in the evolution of low-mass passive galaxies.

Key words: methods: statistical – galaxies: evolution – galaxies: general – galaxies: star formation – galaxies: statistics – galaxies: stellar content.

1 INTRODUCTION

The composition of a galaxy is subject to the influence of an ever-changing balance of astrophysical and cosmological processes acting upon it. Hence, chronicling of the evolutionary history of a galaxy requires a precise knowledge of its present contents. A galaxy expresses its contents (stars, gas, dust, etc.) in its spectral energy distribution (SED). Therefore, inventory of the composition of a galaxy requires measurement of the radiation that it emits as a function of wavelength (Conroy 2013). It is impractical to measure full galaxy spectra that span large wavelength ranges (e.g. ultraviolet-through-infrared), especially for the large number of galaxies needed for a robust statistical study of galaxy evolution. Instead, their SEDs must be inferred from curtailed, summary measurements.

Colours are the simplest such measurements. Optical colours have been used to probe the contents of galaxies since the infancy of extragalactic astrophysics (Roberts 1963). Early studies matched sums of individual stellar spectra (i.e. synthetic composite spectra) to the observed optical colours of galaxies in order to discern their stellar content (e.g. Spinrad 1962; Spinrad & Taylor 1971; Faber 1972).

★ E-mail: seb.turne@gmail.com (ST); msiudek_at_ifae.es (MS)

This method was superseded by stellar population synthesis (SPS), which uses theoretical models of stellar evolution to set astrophysical constraints upon these synthetic composite spectra (e.g. Bruzual & Charlot 2003; Maraston 2005). The advancement of the scope of SPS out to ultraviolet wavelengths and the incorporation of infrared emission models has facilitated the estimation of the full ultraviolet-through-infrared SEDs of galaxies from their observed colours (e.g. Ilbert et al. 2006; Da Cunha, Charlot & Elbaz 2008; Boquien et al. 2019). SEDs spanning these wavelength regimes are governed in their shapes chiefly by stellar emission, and by attenuation (in the ultraviolet and optical) and re-emission (in the infrared) of stellar emission by interstellar dust.

The discovery of a bimodality in the two-dimensional optical colour distribution of galaxies (Strateva et al. 2001; Baldry et al. 2004) has begotten a simple empirical paradigm of galaxy evolution. Galaxies generally go from being blue and star-forming to being red and passive. This change in their colours (and quenching of their star formation) is accompanied for the most part by a change in their morphologies from disc- ('late-type') to spheroid-dominated ('early-type') and an increase in their local environmental densities (Baldry et al. 2006; Bamford et al. 2009). A variety of processes have been proposed as drivers of galaxy evolution (see reviews by Kormendy & Kennicutt 2004 and Boselli & Gavazzi 2006) but their interplay is poorly understood. Furthermore, exceptions to this paradigm (Schawinski et al. 2009; Masters et al. 2010) complicate the issue. Studies aiming to disentangle the interplay of evolutionary processes have focused on galaxies between the two peaks of the colour bimodality (Fritz et al. 2014; Schawinski et al. 2014; Smethurst et al. 2015; Moutard et al. 2016b; Gu et al. 2018; Manzoni et al. 2021; Krywult et al., in preparation), a region called the 'green valley' (Martin et al. 2007). As galaxies under the direct influence of evolutionary processes, they are ideally poised to enable an understanding of how galaxies transition from blue to red.

Bimodalities of galaxies have since also been observed in colours involving ultraviolet and near-infrared magnitudes (Wyder et al. 2007; Williams et al. 2009; Arnouts et al. 2013). Different colours, though, yield slightly different bimodalities; for example, galaxies occupying the blue peak of the $g - r$ bimodality may instead occupy the green valley of the $NUV - r$ bimodality (Salim 2014), because optical–optical colours probe star formation over longer time-scales than ultraviolet-optical colours do. Hence it is clear that, for a complete description of the evolution of galaxies in the context of the bimodality and the green valley, several colours spanning the ultraviolet-through-near-infrared wavelength regime must be considered simultaneously. Machine learning techniques, which can parse multiple features at once, are well suited to the task. Exploration of the multidimensional ultraviolet-through-near-infrared colour distribution of galaxies may overcome degeneracies that exist in two-dimensional colour distributions, uncover substructures to the established bimodality, and reveal the extent to which the ultraviolet-through-near-infrared colours of galaxies express their evolution and assembly histories.

The adoption of machine learning techniques within astronomy and astrophysics was primarily a response to the enormous data volumes anticipated from forthcoming surveys (e.g. 20 TB per night from the Legacy Survey of Space and Time; Ivezić et al. 2019). While fulfilling the demand for automated data analysis methods, these techniques also invite a renewed examination of our understanding of astrophysics due to their ability to distill interpretable models from complex, multidimensional input data that may be difficult to fully visualize. Supervised techniques are useful for mapping existing domain knowledge on to new data. A supervised classification algorithm, for example, may assign labels to previously unseen observations after being trained on pre-labelled observations. Unsupervised techniques, on the other hand, demonstrate substantial promise for exploration and discovery because they are less reliant on prior knowledge than supervised techniques. An unsupervised clustering algorithm, for example, assigns labels to observations in accordance with their intrinsic similarity to one another (i.e. the distances between observations in terms of the features used to represent them). Unsupervised techniques, then, construct models that are driven purely by the structure of input data, and require no training. They may therefore be said to express the 'natural' structure of the input data rather than expressing structures imposed upon it by assumptions that are explicitly built into the use of supervised techniques. The use of unsupervised techniques does, though, incorporate implicit assumptions, and the precise definition of similarity can vary between techniques. Ensuring the astrophysical utility of these models hence requires carefully considered choices of algorithm and features.

A growing literature has emerged in recent years, reporting the results of the application of unsupervised techniques to various astrophysical contexts (see Baron 2019 and Ball & Brunner 2010 for comprehensive reviews). Clustering has been used, for example, to partition galaxies on the basis of their pixel data (Hocking et al. 2017, 2018; Martin et al. 2020), their spectra (Sánchez Almeida et al. 2010; de Souza et al. 2017), their SEDs (Siudek et al. 2018a, b), and their derived astrophysical features (Barchi et al. 2016; Turner et al. 2019). Dimensionality reduction, which can extract important or discriminative information from large ensembles of input features, has been used, for example, to produce simplified projections of galaxy samples based on their multiwavelength photometry (Steinhardt et al. 2020) and their estimated SEDs (Davidzon et al. 2019; Hemmati et al. 2019), and to classify their spectra (Yip et al. 2004; Marchetti et al. 2013).

In this paper, we describe work that builds on that of Siudek et al. (2018a, b). They applied a clustering algorithm to partition galaxies observed by the VIMOS Public Extragalactic Redshift Survey (VIPERS; Scodeggio et al. 2018). They chose the Fisher Expectation-Maximization (`FEM`) algorithm, which implements a clustering approach called the 'Discriminative Latent Mixture' (DLM) model. The algorithm incorporates dimensionality reduction as it iterates rather than as a part of any preparation of the input data ahead of clustering. This ensures that improvements to the estimated parameters of the model are adaptive, and that the clustering uses only the most important information available from the input features. They aimed to establish the ability of `FEM` to determine a naturally defined, astrophysically meaningful partition in a feature space of high dimensionality (i.e. containing more potentially discriminative information than lower dimensionalities). Their feature space was defined by spectroscopic redshifts and 12 rest-frame ultraviolet-through-near-infrared colours. The 12 clusters that they determined revealed substructure to the established colour bimodality of galaxies, distinguishing subpopulations of galaxies that overlapped in two-dimensional colour distributions. In addition, their clusters correlated with a variety of astrophysical features including stellar masses, morphologies, and emission-line strengths.

We adapt the approach of Siudek et al. (2018a, b) to compare samples of galaxies at two different redshifts. Our aim is to use clustering to characterize the structures of the samples in a common feature space of high dimensionality, to examine similarities and differences between these structures at the two cosmic epochs, and to interpret these similarities and differences in the context of theories of galaxy evolution. While each cluster will constitute a class of

galaxies that are intrinsically similar to one another, connections between clusters will chart the evolution of galaxies through the feature space. Hence, we also aim to establish how strongly the evolutionary histories of galaxies, which are ordinarily inferred using a combination of various types of features (e.g. photometric, spectroscopic, morphological), are encoded in just their ultraviolet-through-infrared colours. Our sample of galaxies at low redshift ($z \sim 0.06$) is drawn from the second edition of the GALEX-SDSS-WISE Legacy Catalogue (GSWLC-2; Salim, Boquien & Lee 2018), and our sample of galaxies at intermediate redshift ($z \sim 0.65$) is based on the VIPERS sample of Siudek et al. (2018b). We prepare our samples carefully to ensure a fair comparison of galaxies from different cosmic epochs and different surveys, and to mitigate methodological influences on the clustering outcomes. We also adjust the input features, defining nine neighbouring rest-frame colours that, together, represent the shapes of the ultraviolet-through-near-infrared SEDs of the galaxies in our samples, and thus enable insight into their evolution.

The remainder of this paper proceeds as follows. In Section 2, we introduce our samples, the data we use to represent and analyse the galaxies that they contain (including the estimation of their SEDs), and the measures that we take to ensure a fair comparison between them. In Section 3, we explain the DLM model and how `FEM` algorithm implements it, and we describe the feature space within which we cluster our samples. In Section 4, we present the outcomes of the clustering, and in Section 5, we offer our interpretation thereof. Finally, in Section 6, we summarize, make concluding statements, and suggest future directions for our work. Where required, we assume a $(H_0, \Omega_m, \Omega_\Lambda) = (70 \text{ km s}^{-1} \text{ Mpc}^{-1}, 0.3, 0.7)$ cosmology for our calculations.

2 DATA

2.1 GALEX-SDSS-WISE Legacy Catalogue 2

The second edition of the GALEX-SDSS-WISE Legacy Catalogue (GSWLC-2; Salim et al. 2016, 2018) was assembled using Data Release 10 (DR10; Ahn et al. 2014) of the Sloan Digital Sky Survey (SDSS; York et al. 2000). GSWLC-2 aimed to characterize the star formation activity and dust content of galaxies in the local Universe. It contains all SDSS DR10 galaxies that meet the following criteria:

(i) have apparent *r*-band Petrosian magnitudes <18,
(ii) have spectroscopic redshifts within the range $0.01 < z < 0.3$,
(iii) lie within the Galaxy Evolution Explorer (*GALEX*; Martin et al. 2005; Morrissey et al. 2007) observation footprint, whether they were detected by *GALEX* or not.

The lower redshift limit was imposed to exclude foreground stars, and particularly close galaxies with potentially unreliable photometry and/or distance estimates. Retaining galaxies that were not actually detected by *GALEX* itself preserves the optical selection of SDSS. In all, these criteria select 659 229 SDSS DR10 galaxies.

u-, *g*-, *r*-, *i*-, and *z*-band optical photometry for galaxies in GSWLC-2 was drawn from SDSS. `modelMag` magnitudes, which are based on profile fits, were selected due to the accuracy of their colours. These `modelMag` magnitudes were corrected for extinction due to Milky Way dust using the empirical Yuan, Liu & Xiang (2013) coefficients.

The SDSS optical photometry was supplemented with near- (*NUV*) and far-ultraviolet (*FUV*) photometry from *GALEX*'s final data release (GR6/7). *GALEX* conducted surveys at varying depths: an All-sky Imaging Survey (which observed several targets per orbit), a Medium Imaging Survey (one target per orbit), and a Deep Imaging Survey (several orbits per target). These surveys were nested, such that it is possible for a galaxy to have been observed at more than one depth (although an observation of a galaxy at a given depth does not guarantee an observation of the same galaxy at shallower depths). Here, we use the UV photometry for galaxies in GSWLC-2 based on the deepest available observation of each galaxy (catalogue GSWLC-X2). Salim et al. (2016) applied corrections to mitigate systematic offsets between the SDSS and *GALEX* photometry, which arose mostly due to the blending of sources in *GALEX*'s low-resolution images. Peek & Schiminovich (2013) corrections for extinction due to Milky Way dust were applied to the UV photometry. UV photometry in at least one of *GALEX*'s two bands (almost always *NUV* if just one) is available for 65 per cent of GSWLC-2 galaxies, and for 80 per cent of the galaxies in our final GSWLC-2 sample (Section 2.1.2).

Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) observations at 12 and 2 μm (channels W3 and W4, respectively) were used to provide mid-infrared (MIR) photometry for GSWLC-2 galaxies. Salim et al. (2018) opted for unWISE (Lang, Hogg & Schlegel 2016) forced photometry, which was based directly on SDSS source positions and profiles. MIR photometry in at least one of channels W3 and W4 is available for 78 per cent of GSWLC-2 galaxies, and for 87 per cent of the galaxies in our final GSWLC-2 sample (Section 2.1.2).

2.1.1 GSWLC-2 rest-frame SEDs

The rest-frame SEDs of GSWLC-2 galaxies were estimated using the Code Investigating GALaxy Emission (CIGALE; Noll et al. 2009; Boquien et al. 2019). Synthetic spectra generated by CIGALE were validated against the available observed UV-through-optical photometry in order to constrain the SEDs. Details of this fitting procedure are described at length in Salim et al. (2016, 2018); here, we offer a brief summary.

Synthetic spectra were generated using Bruzual & Charlot (2003) simple stellar population templates, based on a Chabrier (2003) initial mass function and with metallicities of $\log_{10}(Z) = -2.4$, -2.1, -1.7 ($\sim Z_\odot$), or -1.3. These templates were combined with Myr-resolution star formation histories (SFHs) consisting of two exponentially declining episodes of star formation, producing an old and a young population. Absorption of stellar emission by dust was implemented via a Noll et al. (2009) generalization of the Calzetti et al. (2000) attenuation curve, modified to allow its slope to vary and to add a UV bump (see section 3.4 of Salim et al. 2018).

The SED estimation was additionally constrained by the galaxies' total IR luminosities (i.e. matching the energy absorbed by the dust within galaxies with the energy it re-emits; see section 3.2 of Salim et al. 2018). Total IR luminosities were derived from the 22 μm WISE photometry (if available, 12 μm if not) using Chary & Elbaz (2001) templates, further corrected based on Herschel (Valiante et al. 2016) IR photometry (see section 3.1 of Salim et al. 2018). The overall quality of fit was measured by its reduced chi-squared value (χ^2_r).

Astrophysical features including rest-frame absolute magnitudes, colour excesses [$E(B-V)$], stellar masses (M_*), stellar metallicities (Z), mass-weighted stellar ages (*MWSA*), and specific star formation rates [*sSFR* (SED)] were derived from the full ensemble of possible synthetic spectra via a Bayesian approach (Salim et al. 2007). The likelihood of the fit of each synthetic spectrum to the photometry of each galaxy was used to generate a probability density function for each feature, with the likelihood-weighted means of the functions

being quoted as the best estimates of the features, and the likelihood-weighted standard deviations as the errors.

2.1.2 *Final low-redshift sample*

Our final GSWLC-2 sample is subject to the following selections. First, we only retain galaxies whose best-fitting CIGALE SEDs produce $\chi^2_r <= 11.07$ (i.e. the mean plus two standard deviations of the logarithmic GSWLC-2 distribution in χ^2_r), in order to omit particularly poorly constrained fits. Spectroscopic redshifts are limited to the range $0.02 < z < 0.08$, and stellar masses (as estimated via Bayesian analysis of the synthetic CIGALE spectra) to $>10^{9.5}$ $M_\odot$. These two restrictions ensure completeness above the imposed stellar mass limit. Finally, broad-line active galactic nuclei are removed by asserting `flag_sed` = 0. Our final GSWLC-2 sample has a median redshift of 0.06 and contains 177 362 galaxies.

As additional, CIGALE-independent indicators of the stellar populations in GSWLC-2 galaxies, we invoke Brinchmann et al. (2004) specific star formation rates [*sSFR* (ind.)] and 4000 Å break strengths [*D*(4000)]. The SFRs sum two components: a spectroscopic fibre SFR, and a photometric SFR outside the fibre, given by an optical SED fit (Salim et al. 2007). The fibre SFR is given by either a Hα calibration (Charlot & Longhetti 2001) or, in the case of spectra that have a contribution from an active galactic nucleus, a *D*(4000)-based estimate (itself calibrated on the emission lines of pure star-forming galaxies). These SFRs are then normalized by photometrically determined stellar masses to give *sSFR* (ind.). The time-scale probed by *sSFR* (ind.) lies between the 10 Myr time-scale of the Hα-calibrated fibre SFRs, and the 1 Gyr time-scale of optical SED-based SFRs (Salim et al. 2016). The *D*(4000) measurements apply to fibre region only. Both of these features are available for 97 per cent of the galaxies in our GSWLC-2 sample.

We obtain Sérsic indices (n_g) and circularized half-light radii ($R_{1/2}$) for the galaxies in our GSWLC-2 sample from catalogues assembled by Simard et al. (2011). Both were derived from fits of singular Sérsic (1963, 1968) profiles to *r*-band images of galaxies in SDSS. The Sérsic indices have minimum and maximum allowed values of 0.5 and 8, respectively. Sérsic indices and half-light radii are available for 96.2 per cent of the galaxies in our final GSWLC-2 sample. We also use Simard et al. (2011) *r*-band bulge-to-total ratios (B/T_r) for these galaxies, which were based on fits consisting of two components: a Sérsic bulge (fixed at an index of 4) and an exponential disc. Local environmental densities, available for 92.1 per cent of our GSWLC-2 galaxies, come from Baldry et al. (2006). They averaged the surface densities of SDSS galaxies with respect to their fourth- and fifth-nearest density-defining neighbour within 1000 km s^{-1} along the line of sight. We calculate local overdensities (δ) using $\delta = (\Sigma - \bar{\Sigma})/\bar{\Sigma}$, where Σ is the local surface density and $\bar{\Sigma}$ the average surface density of the sample.

2.2 VIPERS

The VIMOS Public Extragalactic Redshift Survey (VIPERS; Garilli et al. 2014; Guzzo et al. 2014; Scodeggio et al. 2018) aimed to match the statistical fidelity of low-redshift surveys like SDSS, but at intermediate redshifts ($z \sim 0.7$). The survey was conducted using the VIMOS spectrograph (Le Fèvre et al. 2003) of the European Southern Observatory's Very Large Telescope. Its targeting was based on the Canada–France–Hawaii Telescope Legacy Survey Wide (CFHTLS-Wide) photometric catalogue,[1] with objects qualifying for VIPERS if they had extinction-corrected *i*-band magnitudes $i_{AB} < 22.5$. An additional *ugri* colour cut was applied to remove low-redshift ($z \lesssim 0.5$) galaxies from the survey (Guzzo et al. 2014). PDR2, the second and final public data release of VIPERS, comprises spectroscopy for 97 414 objects (Scodeggio et al. 2018). 52 114 of these objects (51 522 galaxies and 592 broad-line active galactic nuclei) have 'secure' (>99 per cent confidence) redshifts. This secure-redshift sample was the subject of the Siudek et al. (2018b) study, and is the basis of our present VIPERS sample.[2]

Photometry for this sample was taken from a catalogue prepared by Moutard et al. (2016a). The CFHTLS-Wide photometric catalogue (i.e. the basis of the targeting for VIPERS) provided optical photometry for this sample in $u*$, *g*, *r*, *i*, and *z* bands. Moutard et al. (2016a) derived total magnitudes for the galaxies in this sample by rescaling their isophotal magnitudes. These isophotal magnitudes were chosen for the accuracy of their colours with a view to photometric redshift estimation; this choice now benefits our SED estimation as well.

Like for our GSWLC-2 sample, UV photometry came from *GALEX*. Moutard et al. (2016a) supplemented existing Deep Imaging Survey observations of VIPERS galaxies with deep *GALEX* observations of their own in order to improve UV coverage within the VIPERS footprint. Coverage is complete in the W1 field of VIPERS, but not in the W4 field (see fig. 1 of Moutard et al. 2016a). UV photometry was then measured using a Bayesian approach with the $u*$-band profiles of galaxies as priors (Conseil et al. 2011), which mitigated the confusion of sources due to their blended UV profiles. UV photometry in at least one of *GALEX*'s two bands (almost always *NUV* if just one) is available for 52 per cent of galaxies in the Siudek et al. (2018b) sample and in our final VIPERS sample (Section 2.2.2).

Near-infrared (NIR) K_s-band photometry came from a dedicated CFHT WIRCam (Puget et al. 2004) follow-up survey of VIPERS galaxies (Moutard et al. 2016a). This K_s-band photometry was validated against NIR photometry from the VISTA Deep Extragalactic Observations (VIDEO) survey (Jarvis et al. 2013), exhibiting good agreement. We also take VIDEO survey *Z*, *Y*, *J*, *H*, and K_s NIR photometry for our sample where available (11 per cent of the Siudek et al. 2018b sample, 10 per cent of our final VIPERS sample; Section 2.2.2). CFHT K_s-band photometry is available for 91 per cent of galaxies in the Siudek et al. (2018b) sample, and for 93 per cent of galaxies in our final VIPERS sample (Section 2.2.2).

2.2.1 *VIPERS rest-frame SEDs*

The SEDs of VIPERS galaxies are estimated via a full fit of synthetic CIGALE spectra to the available UV-through-NIR photometry. This differs slightly from the method used for the GSWLC-2, whose NIR SEDs were constrained not by their shapes but simply by their total IR luminosities (Section 2.1.1). While we use the same stellar templates (Bruzual & Charlot 2003, with Chabrier 2003 initial mass functions and metallicities of 0.004, 0.008, 0.02, or 0.05) for VIPERS as were used for GSWLC-2 , the SFHs are adjusted to reflect the change in cosmic epoch between samples and to account for the possibility

[1] http://www.cfht.hawaii.edu/Science/CFHLS

[2] The use of these secure redshifts is recommended by Garilli et al. (2014) and Scodeggio et al. (2018) for scientific analyses. Approximately 75 per cent of all VIPERS galaxies within and throughout the redshift range of our final VIPERS sample (see Section 2.2.2) have secure redshifts (see fig. 9 of Scodeggio et al. 2018).

of very recent bursts of star formation.[3] Astrophysical features are derived for VIPERS galaxies using the same Bayesian approach as for GSWLC-2 galaxies (see Section 2.1.1).

2.2.2 *Final intermediate-redshift sample*

We make the following selections to yield our final VIPERS sample. Galaxies are kept if the χ^2_r of their best-fitting CIGALE SED has a value less than or equal to the mean plus two standard deviations (=18.85) of the overall logarithmic VIPERS distribution. Spectroscopic redshifts are restricted to being within the range $0.5 < z < 0.8$, balancing our intent to define a co-eval population of galaxies against the need to keep the sample as large as possible. Like our GSWLC-2 sample, stellar masses are limited to $>10^{9.5}$ $M_\odot$ with a view to mass completeness (though see Sections 4.4.2 and 5.2, where we discuss shortcomings). Broad-line active galactic nuclei and serendipitous secondary spectral sources are removed using `zflag` <10. Ultimately, this gives us a final VIPERS sample consisting of 31 889 galaxies, with a median redshift of 0.65.

Emission-line SFRs, which are independent of our CIGALE SED estimation, were calculated from the [O II] λ3727 fluxes of the galaxies in our VIPERS sample using the calibration (which includes empirical stellar-mass-based corrections) of Gilbank et al. (2010, 2011a, b). These [O II] λ3727 fluxes are available for 27 537 of the galaxies in our VIPERS sample, and they probe short time-scales of star formation (∼10 Myr). We normalize these [O II] SFRs by our CIGALE stellar masses to yield specific star formation rates[4] [*sSFR* (ind.)]. *D*(4000) was measured from VIPERS spectra by Garilli et al. (2014), using the same Balogh et al. (1999) method as was used for SDSS (Brinchmann et al. 2004). Sérsic indices and circularized half-light radii for the galaxies in our VIPERS sample are given by Krywult et al. (2017), who fitted the *i*-band light distributions of galaxies with single Sérsic (1963, 1968) profiles. These features are available for 96.2 per cent of the galaxies in our final VIPERS sample. We winsorize the Sérsic indices to values of 0.5 and 8 in order to match our GSWLC-2 sample. The overdensities of 91.7 per cent VIPERS galaxies were derived by Cucciati et al. (2017), based on fifth-nearest neighbour surface densities.

3 CLUSTERING METHOD

We apply the `FEM` algorithm, which estimates the parameters of the DLM model. Bouveyron & Brunet (2012) offer full, rigorous, mathematical derivations of both the DLM model and the `FEM` algorithm in their paper; here, we offer brief summaries of the model (Section 3.1), and of its implementation via the algorithm (Section 3.2). In Section 3.3, we discuss some additional relevant practicalities to the use of the model and algorithm, and in Section 3.4, we describe the shared feature space within which we cluster our two samples.

[3]Consequences of this adjustment are discussed in Section 4.4.2; the properties of most VIPERS galaxies appear accurate, except for those a subpopulation of passive VIPERS galaxies.

[4]Our use of stellar masses given by CIGALE means that these *sSFR* (ind.) estimates are not entirely independent of CIGALE; however, we expect that CIGALE's stellar masses would be consistent with those estimated via other methods, given that stellar mass estimates are generally quite robust (Bell & de Jong 2001).

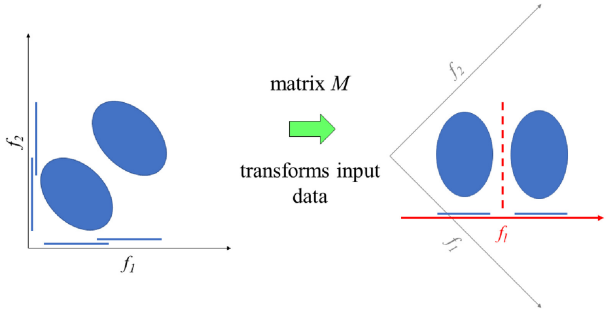

Figure 1. A simple demonstration of the principles behind subspace clustering. Here, a sample consisting of two clusters (represented by the two blue ellipses) is represented in a two-dimensional full space defined by features f_1 and f_2. Matrix M enables the transformation of the sample to a one-dimensional subspace, defined by latent feature f_l, in which the two clusters are easily discriminated.

3.1 The DLM model

The DLM model is a clustering approach that incorporates dimensionality reduction on the fly to determine a frugal fit to the structure of an input sample, which is assumed to consist of k clusters. Selection of the value of k is discussed in Section 3.3.

The key premise of the DLM model is thus: a sample represented in a D-dimensional space that is defined by observed features actually occupies an intrinsic d-dimensional subspace ($d < D$; the 'empty space phenomenon'; Scott & Thompson 1983) that is defined by unobserved, latent features. Hence, the clustering structure of the sample should be fitted in this intrinsic subspace.

The subspace has two important properties in the context of the DLM model. First, of all possible d-dimensional subspaces, it is the one that best discriminates the k clusters in the sample. The model assumes $1 \leq d \leq k - 1$: that k clusters may be distinguished in $k - 1$ dimensions or fewer (see Section 3.3 for further explanation). Secondly, the subspace is linearly related to the full D-dimensional space, such that the unobserved, latent features are linear combinations of the observed features. Hence there exists a matrix M, common to all of the k clusters, that enables the transformation of the sample between the full space and the subspace. This transformation matrix is constrained by the condition that the basis vectors of the subspace must be orthonormal. Estimation of the transformation matrix M is explained in Section 3.2. Selection of the value of d is explained in Section 3.3. Fig. 1 demonstrates these two important properties of the subspace.

The DLM model assumes that the sample is distributed among a mixture of k Gaussian density functions within the discriminative latent subspace. The functions, each of which corresponds to a cluster, are defined by three parameters: a mean vector (μ_k), a covariance matrix (Σ_k), and a scalar relative mixture proportion (π_k). The matrix M enables the transformation of these parameters back to the full space. For the covariances, this includes the addition of Gaussian 'noise' (δ_k; unique to each of the clusters), which is defined as non-discriminative structure that exists in the full space but not in the subspace. While Σ_k captures the cluster covariances inside the discriminative latent subspace, δ_k captures the cluster covariances outside the subspace. Full space covariances are the sum of both. Estimation of the cluster means, covariances, and noise terms is discussed in Section 3.2.

Implementation of the DLM model hence requires the estimation of the following parameters:

Table 1. Integrated Completed Likelihood (ICL) scores reported by our search over all possible combinations of submodel (see Section 3.1 for further explanation) and k for our samples. The uncertainties span the full range of ICL scores (i.e. from minimum to maximum) registered over 100 initializations for each combination. As mentioned in Section 3.3, only nine of the 12 submodels are available in the version of FEM that we use for our fitting. The score of the best-fitting combination is highlighted using bold text. While submodel Σ, δ produces the highest score for our GSWLC-2 sample (at $k = 9$), we reject it for reasons given in Appendix C. Blank entries correspond to combinations for which FEM did not converge (see Appendix C). The entries listed in this table are subject to the multipliers at the right-hand side of each section. The ICL scores for our GSWLC-2 sample are systematically higher than those for our VIPERS sample because it contains more galaxies.

		Submodel								
		Σ_k, δ_k	Σ_k, δ	Σ, δ_k	Σ, δ	$\alpha_{k,j}$, δ_k	$\alpha_{k,j}$, δ	α_k, δ_k	α_k, δ	α, δ
GSWLC-2	$k = 2$	1.8 ± 0.0	1.5 ± 0.0	0.4 ± 0.0	−6.2 ± 0.0	1.8 ± 0.0	−4.7 ± 0.0	1.8 ± 0.0	−4.7 ± 0.0	−6.2 ± 0.0 × 10^5
	$k = 3$	8.2 ± 0.0	7.8 ± 0.0	−141.0 ± 0.0	0.0 ± 0.0	3.8 ± 0.0	−5.1 ± 0.0	3.7 ± 0.0	−5.3 ± 0.0	−4.5 ± 0.0
	$k = 4$	11.7 ± 0.0	11.3 ± 0.0	2.7 ± 0.0	5.2 ± 0.0			4.9 ± 0.0	−4.2 ± 0.0	−3.8 ± 0.0
	$k = 5$	13.4 ± 1.4	13.4 ± 0.2	−46.2 ± 51.5	8.7 ± 0.4	6.7 ± 1.0		6.0 ± 0.0	−2.2 ± 0.0	−5.7 ± 1.3
	$k = 6$	16.7 ± 0.0		9.4 ± 0.0	13.0 ± 0.1			6.8 ± 0.0	0.7 ± 0.0	−5.2 ± 0.0
	$k = 7$	**17.9 ± 0.2**		11.8 ± 2.0	14.2 ± 1.6			7.2 ± 0.0	2.3 ± 0.0	−5.6 ± 0.0
	$k = 8$				16.3 ± 1.3			8.1 ± 0.0	1.1 ± 0.0	−5.7 ± 0.0
	$k = 9$			17.1 ± 0.0	~~**18.1 ± 0.9**~~			8.0 ± 0.0	3.9 ± 0.0	−6.0 ± 0.0
VIPERS	$k = 2$		2.1 ± 0.0	4.3 ± 0.0	−8.4 ± 0.0		−8.0 ± 0.0		−8.0 ± 0.0	−8.4 ± 0.0 × 10^4
	$k = 3$		11.1 ± 0.0	−421.0 ± 0.0	−0.4 ± 0.0		−4.7 ± 0.0	6.8 ± 0.0	−5.3 ± 0.0	−7.9 ± 0.0
	$k = 4$			−294.0 ± 0.0	8.3 ± 0.0		−6.7 ± 0.0	8.3 ± 0.0	−7.4 ± 0.0	−9.0 ± 0.0
	$k = 5$	32.9 ± 0.0	32.4 ± 0.0	15.6 ± 0.0	16.3 ± 0.2			10.5 ± 0.0	−2.9 ± 0.0	−7.0 ± 0.0
	$k = 6$			20.9 ± 1.5	23.6 ± 0.0			13.0 ± 0.0	2.9 ± 0.0	−3.8 ± 0.1
	$k = 7$		**41.8 ± 0.1**	26.4 ± 2.3				15.9 ± 0.8	7.1 ± 0.4	−6.7 ± 0.0
	$k = 8$							14.6 ± 0.0	3.2 ± 0.0	−10.8 ± 0.0
	$k = 9$							12.5 ± 0.0		−10.8 ± 0.0

(i) $k - 1$ relative mixture proportions (π_k; given that one cluster has a proportion of 1);

(ii) kd parameters for the mean vectors (μ_k) in the subspace;

(iii) $kd(d + 1)/2$ parameters for the covariance matrices (Σ_k) in the subspace (fewer than kd^2 parameters because covariance matrices are symmetric);

(iv) $d(D - (d + 1)/2)$ parameters for the transformation matrix M (the number of free parameters, given the constraint that the basis vectors of the subspace must be orthonormal);

(v) k noise terms (δ_k; given that this non-discriminative structure is Gaussian and spherical, and may therefore by parametrized by a single value in reference to the Gaussian density function estimated for each cluster).

The total number of parameters ($q_{\rm DLM}$) is most strongly influenced by the value of d. The maximum $q_{\rm DLM}$ at a certain combination of D and k is given by setting d to its maximum value of $k - 1$ (based on the aforementioned assumption that k clusters may be distinguished in $k - 1$ dimensions or fewer). $q_{\rm DLM}$ is smaller than the number of parameters that must be estimated for a Gaussian Mixture Model in the full space ($q_{\rm GMM}$), especially if $d \ll D$ ($q_{\rm GMM}$ is given by the sum of $k - 1$ relative mixture proportions, kD parameters for the mean vectors, and $kD(D + 1)/2$ parameters for the covariance matrices).

Parameter $q_{\rm DLM}$ may be further reduced by imposing additional constraints upon the DLM model. For example, the covariance matrices (Σ_k) may be assumed to be the same for all Gaussians (Σ; the Gaussians all have the same shape). Alternatively, they may be assumed to be diagonal ($\alpha_{k,j}$, where the subscript j indicates a different variance in each dimension of the subspace), meaning the latent features that define the subspace are uncorrelated. These diagonal covariance matrices may then also be assumed to be isotropic (α_k; spherical Gaussians in the subspace), the same for all Gaussians (α_j), or both (α). The noise terms (δ_k) may be assumed to be the same for all Gaussians (δ) as well. Constraints like these may be imposed to speed up the clustering, in anticipation of a particular clustering structure, or (as in our case) to compare fits of models of varying complexities (see also Section 3.3). The various combinations of these constraints on the covariance matrices and noise terms yield 11 submodels of the full Σ_k, δ_k DLM model. They are listed in full in table 1 of Bouveyron & Brunet (2012) (and listed partially in Table 1 of this paper).

3.2 The FEM algorithm

The FEM algorithm estimates the parameters (π_k, μ_k, Σ_k, M, δ_k) of the DLM model, fitting a sample of N observations, observed in a D-dimensional space (the 'full' space, defined by D observed features), with k Gaussian density functions in a d-dimensional discriminative latent subspace ($1 \leq d \leq k - 1$). FEM comprises the following steps:

(i) Initialization: k starting points are selected within the extent of the sample in the full space;

(ii) Expectation (E): transform the parameters of the mixture of Gaussians to the full space, and calculate the probability of each observation having originated from each Gaussian;

(iii) Fisher (F; based on discriminant analysis): using the observation probabilities, find the subspace that best separates the Gaussians;

(iv) Maximization (M): update the parameters of the mixture of Gaussians (including non-discriminative structure, termed 'noise') within the subspace.

The Expectation, Fisher, and Maximization steps are iterated such that FEM improves its estimates of the DLM model parameters as it proceeds. FEM is slow to run on our large samples and, unlike traditional expectation-maximization algorithms, does not always converge perfectly (such that there are no changes between successive iterations; due to the Fisher step). We therefore terminate FEM at the completion of 25 iterations; changes between iterations become negligible well before this number (see Appendix A). The final output of FEM is a series of k probabilities for each of the observations: probabilities of each observation having originated from each of the k Gaussians. Final cluster labels are given by assigning each

observation to the Gaussian with the highest probability of having originated it.

While successive iterations of FEM improve its estimates of the DLM model parameters, these estimates improve only towards local maxima of the likelihood function. FEM is hence run with varying initializations, which may intuitively be considered as 'exploring the surface' of the likelihood function of the model parameters. This encourages optimization towards different local maxima and, hopefully among these, the global maximum, corresponding to the very best estimate of the DLM model parameters.

Initialization techniques may be as simple as a uniform random selection of k observations from the sample. We opt to use the k-means algorithm (MacQueen 1967; Lloyd 1982), which implements a simple centroid-based clustering approach, to generate initializations for FEM. k-means is an expectation-maximization algorithm and, like FEM, only optimizes to local maxima. We therefore initialize k-means *itself* 100 times in the hope of encouraging optimization towards the global maximum of *its* objective function (which measures how separated the clusters are). Use of varying initializations provided by a heuristic like k-means leads to 'pre-optimization' of FEM because the separated centroids are likely to span the full extent of the sample in its full space. This facilitates improvement of FEM's estimates of the DLM model parameters towards the global maximum of their likelihood functions. Following this initialization, FEM proceeds to the Fisher step, in which it finds the subspace that best separates the final k-means clusters, and to the Maximization step, in which it fits the observations with a mixture of Gaussians within this subspace. FEM then loops back around to the Expectation step and begins iterating proper.

The Expectation step uses the parameters estimated in the Maximization step (π_k, μ_k, Σ_k, δ_k) to calculate the conditional probability of each observation having originated from each of the k Gaussians. These parameters are transformed from the subspace, within which they are estimated in the Maximization step, to the full space using matrix M, found in the Fisher step.

The Fisher step finds the d-dimensional discriminative latent subspace that best separates the new partition calculated in the Expectation step. Bouveyron & Brunet (2012) base this step on discriminant analysis, which finds the linear combination of the input features that maximizes the ratio of the scatter *between* clusters to the scatter *within* clusters. Similar principles have been applied for the visualization of multidimensional clusters as well (e.g. Lisboa et al. 2008). These scatters are weighted by the probabilities calculated in the Expectation step. A constraint of the DLM model is that the d basis vectors that define the subspace must be orthonormal, which is not necessarily a property of the d basis vectors that linear discriminant analysis (LDA) provides. Bouveyron & Brunet (2012) assert this constraint by applying the orthonormal discriminant vector method (ODV; Okada & Tomita 1985). ODV uses LDA to find the d basis vectors in succession while also ensuring the orthonormality of each new basis vector with respect to all of those that have already been calculated. The first basis vector, which is free of this constraint, is given by the direct application of LDA to the sample in the full space. The d orthonormal basis vectors constitute the columns of M, the matrix that enables the transformation of the sample between the full space and the subspace.

The Maximization step updates the estimates of the means, covariances, and relative mixture proportions (π_k, μ_k, Σ_k) of the k Gaussians in order to maximize the likelihood of the fit. These estimates are measured within the subspace found in the Fisher step, and are weighted by the probabilities calculated in the Expectation step. This step also updates the estimates of the noise terms (δ_k), which is given by the differences between the full-space variances (again weighted by the probabilities calculated in the Expectation step) and the newly updated subspace variances.

3.3 Practicalities

We do not presume a DLM submodel or value of k with which to fit our samples. Instead, we conduct a search over all of the DLM submodels and over a range of values of k to determine the best-fitting combination. Three of the DLM submodels (α_j, δ_k; α_j, δ; α, δ_k) are not available for use in the version of FEM[5] that we use for our fitting. This reduces the total number of available submodels from 12 (including the full Σ_k, δ_k model) to nine.

We identify the best-fitting combination of DLM submodel and value of k by using the Integrated Completed Likelihood criterion (ICL; Biernacki, Celeux & Govaert 2000):

$$\mathrm{ICL} = \ln(L) - \frac{q_{\mathrm{DLM}}}{2}\ln(N) - \left[-\Sigma_{i=1}^{N}\,\Sigma_{l=1}^{k}\, z_{i,l}\ln(p_{i,l})\right], \qquad (1)$$

where L is the likelihood of the fit, $p_{i,\,l}$ is the probability of observation i belonging to cluster l, and $z_{i,\,l}$ denotes cluster membership, taking a value of 1 when $p_{i,\,l} = \max(p_{i,\,:})$ and a value of 0 otherwise. The ICL is related to the popular Bayesian Information criterion (BIC; Schwarz 1978). While both the BIC and ICL criteria penalize the likelihood using the number of model parameters (to avoid overfitting), the ICL criterion also rewards separated clusters (a general aim of clustering). The combination of submodel and k that returns the highest ICL score is deemed the best fit.

The dimensionality of the discriminative latent subspace is constrained by the number of clusters being fitted: $1 \leq d \leq k-1$. The maximal $d = k-1$ case may intuitively be understood as setting the origin of the subspace at one of the k cluster centres so that the full-space vectors to each of the remaining $k-1$ cluster centres define the basis vectors of the subspace. If multiple clusters lie along the same direction in the full space, the number of basis vectors needed to define the subspace is reduced. In our application of FEM, we hold d at its maximum value of $k-1$. This is recommended by Bouveyron & Brunet (2012) to avoid omitting any discriminative structure from the subspace and to ease convergence of FEM (which may become unstable or fail to converge if d is too small in comparison with k and/or D). Hence, the maximum value of k in our model selection search is 9 (set by $d = 8$, given $D = 9$).

3.4 Input features to the clustering

The fitting of the clustering structures of both of our samples is conducted within a nine-dimensional feature space defined by UV-through-NIR colours. We opt for colours because of their widespread use in studies of galaxy evolution, and because of the relative ease with which they may be measured. While clustering in terms of derived astrophysical features may facilitate a more direct interpretation of resultant clusters in terms of theories of galaxy evolution, their derivation is much more model-dependent than that of colours. Clustering in terms of colours ensures the generalizability of our outcomes.

The colours that we use are calculated not from the observed photometry that is used as input to the SED fitting, but from rest-frame magnitudes estimated by CIGALE. This ensures homogeneity among the input features, and that the feature space is defined by

[5]Version 1.5.1, for the R statistical computing environment.

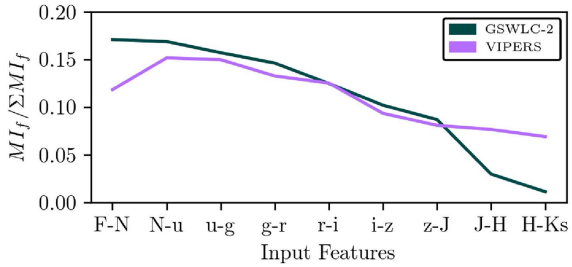

Figure 2. The relative importance of each of the input features to the clustering. 'F' stands for FUV, and 'N' for NUV. The mutual information (see Section 4.2 and equation 2) of each of the input features with respect to the cluster labels has been normalized by the sum across all of the input features for each sample.

rest-frame colours (which is more difficult to ensure using colours calculated directly from observed photometry). In addition, the SED estimation can infer the rest-frame magnitudes of galaxies in bands for which there is no observed photometry. The full list of rest-frame colours used for the clustering is: $FUV - NUV$, $NUV - u$, $u - g$, $g - i$, $i - r$, $r - z$, $z - J$, $J - H$, and $H - K_s$. These rest-frame colours are intended to represent the shape of each galaxy's UV-through-NIR SED, and to remove the influence of the intrinsic brightnesses of the galaxies on the clustering outcomes. The rest-frame magnitudes of GSWLC-2 galaxies (but not VIPERS galaxies) are subject to some smoothing (see Appendix B). In addition, the rest-frame NIR colours of GSWLC-2 galaxies were inferred from UV and optical photometry (given the lack of input NIR photometry). Use of the term 'colour' from this point forward in this paper is intended in reference to these rest-frame colours, as estimated by CIGALE.

These colours differ from those used by Siudek et al. (2018b); they used rest-frame colours defined with reference to the rest-frame i-band magnitudes of galaxies ($FUV - i$, $NUV - i$, etc.), also with the aim of removing the influence of galaxy intrinsic brightnesses on their clustering outcomes. However, their UV colours, defined across the largest distances in wavelength among their features, exhibited large spreads (up to a factor of 10 larger than the spreads of other colours) and dictated much of their clustering. Preliminary tests of clustering with these i-band based colours for our present, carefully prepared samples confirmed this. The $\alpha_{k,j}$ and $\alpha_{k,j}$ submodels achieved the highest ICL scores for these i-band colours, but gave only relatively crude segmentations of our samples (see also Appendix C). Our colours, defined using magnitudes in filters at neighbouring effective wavelengths, mitigate this effect and encourage FEM to converge to more detailed partitions (although, as shown in Fig. 2, bluer colours are still most important).

4 RESULTS

4.1 FEM submodel selection

As outlined in Section 3.3, we conduct a search for the best-fitting FEM submodel and number of clusters for our samples. We identify the best-fitting combination using the ICL criterion (equation 1), which penalizes the number of parameters of the submodel while favouring separated clusters. Table 1 lists ICL scores reported for both samples. The uncertainties on these scores, which span the *full* variation (i.e. from minimum to maximum) over 100 initializations, show that FEM is extremely stable and self-consistent, robustly converging to highly similar outcomes over successive runs that use the same combination of submodel and number of clusters. The best-fitting combinations for each sample are highlighted using bold text. We briefly describe patterns of behaviour of the various submodels and explain the large range in ICL scores in Appendix C. Despite it registering the highest score for the GSWLC-2 sample, we reject the $k = 9$, Σ, δ combination due to its inclusion of empty clusters (explained further also in Appendix C).

Both samples are best partitioned into seven clusters, within a six-dimensional discriminative latent subspace. The Gaussian density functions representing the clusters are each described by their own unique, full covariance matrices (Σ_k); the clusters each have different shapes, and the use of full covariance matrices indicates correlations (as expected) among the input features within the subspaces. While the best-fitting submodel for the GSWLC-2 sample uses unique noise terms for each cluster (δ_k), the best-fitting submodel for the VIPERS sample does not (δ), owing to the smoother distribution of the VIPERS sample in the feature space (see e.g. Fig. 3). Submodels Σ_k, δ_k and Σ_k, δ report similar ICL scores and produce similar clustering structures in general and may therefore readily be compared with one another (see also Appendix C). That FEM has converged to highlighting these closely related submodels as being optimal for describing both samples is encouraging, and gives us confidence that we are conducting a fair comparison.

4.2 Feature importance

In Fig. 2, we show the relative importance of each input feature to the clustering. Specifically, we calculate the mutual information (MI) between each input feature and the output cluster labels:

$$MI(f, l) = D_{KL}(p_{f,l} || p_f p_l). \quad (2)$$

Here, $D_{\rm KL}$ is the Kullback–Leibler divergence (Kullback & Leibler 1951; also known as the relative entropy) between the joint probability distribution of input feature f and output label l, and their independent distributions. For Fig. 2, $MI_{f,l}$ is normalized by its sum across all input features to give a relative value.

The lines in Fig. 2 are broadly similar, indicating that, on the whole, FEM uses the nine features in a similar way to determine its best partitions. This is further confirmed by noting that the subspaces within which FEM determined these best partitions have the same dimensionality (6) for both samples. The lines are especially consistent among the optical colours, which is expected given that optical photometry is ubiquitously available for galaxies in both samples. Altogether, the optical regime is the most important to the clustering. Individually, colours from the UV region of the SEDs of the galaxies in both samples are most strongly related to the output cluster labels. This highlights, as expected, the star formation activity and the dust content of galaxies as major influences on the shapes of their UV-through-NIR SEDs.

UV colours are slightly more important for the clustering in our GSWLC-2 sample, which reflects the increased UV coverage of its galaxies by *GALEX* (80 per cent, as opposed to 52 per cent for our VIPERS sample). NIR colours are less important for distinguishing clusters within our GSWLC-2 sample than within our VIPERS sample, which is likely due to their having been inferred purely from UV and optical input photometry. This is in contrast with the galaxies in our VIPERS sample, whose NIR SEDs (more important to the clustering) were instead constrained by K_s-band photometry.[6]

[6]While the Two Micron All-Sky Survey (Skrutskie et al. 2006) has NIR photometry for ∼50 per cent of GSWLC-2 galaxies, it is shallow and would not have provided strong constraints upon their NIR SEDs.

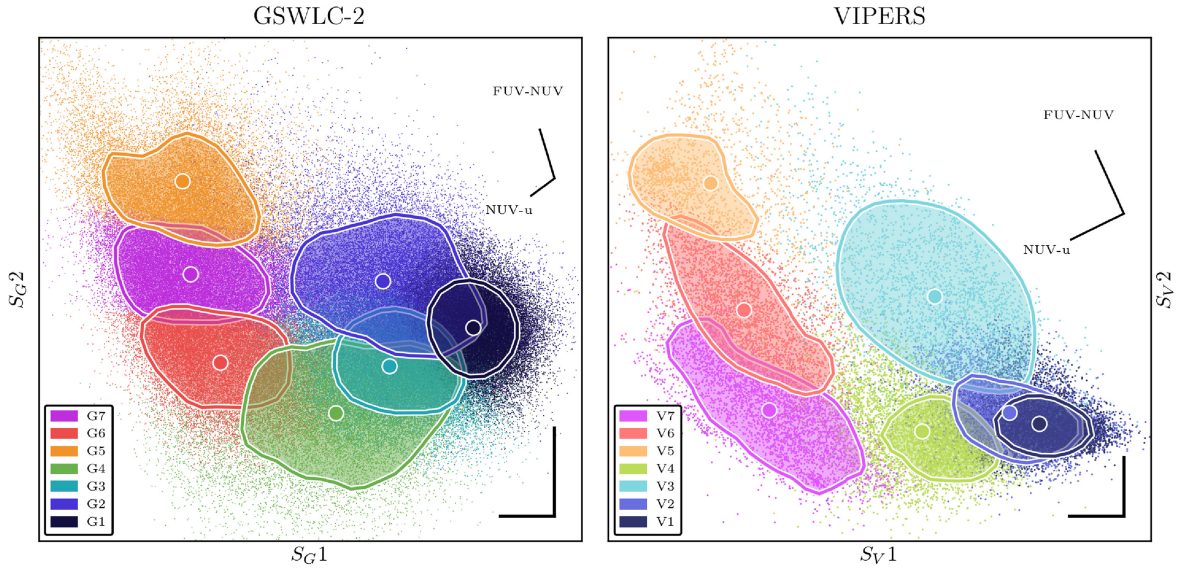

Figure 3. Projections of both samples on to the two dimensions that best separate their clusters. The axes of each plot are determined by FEM and are unique to each sample (as indicated by their labels; e.g. S_G1 represents the first axis of the subspace of our GSWLC-2 sample), but the resultant projections are mostly similar none the less. The distributions of clusters within this plane are shown using coloured, filled contours (drawn at a relative density of 0.4), and the coloured, circular markers show their means. The perpendicular black lines at the lower right of each plot show the extent to which the y-axis has been stretched relative to the x-axis to yield the projections as shown. The vectors at the upper right of each plot show the projections of the two input features that correlate most strongly with the axes of these projections.

For galaxies with incomplete photometry, the array of templates and synthetic spectra with which CIGALE may fit them is reduced, leading to reduced variation in the shapes of their SEDs. In addition, the rest-frame magnitudes (and hence, rest-frame colours) that CIGALE must infer from photometry at other wavelengths have larger uncertainties. Hence, availability of photometry with which to constrain the SEDs of galaxies is advantageous to the clustering. Nevertheless, Fig. 2 shows that, for the most part, FEM uses the features similarly to model both samples despite slight differences in this availability, which is driven mostly by the ubiquitous availability of optical photometry for both samples.

4.3 Clustering structures

Table 2 profiles the clusters determined within both samples. Features are derived from the same SEDs as the colours used for the clustering (see Sections 2.1.1 and 2.2.1) as well as from ancillary sources (see Sections 2.1.2 and 2.2.2). Clusters are named using two-part notation that will be used throughout the remainder of this paper; prefixes 'G' or 'V' denote clusters determined within the GSWLC-2 and VIPERS samples, respectively. Clusters names have been ordered by their mean $NUV - r$ colours for ease of reference.

Fig. 3 shows projections of our samples on to the two principal dimensions of their respective six-dimensional discriminative subspaces. These projections, which offer direct views of the structures of the clustering outcomes, are determined uniquely for each sample by FEM: *the axes of the two plots do not correspond exactly to one another*. Nevertheless, these projections are broadly similar in terms of the shapes of the overall samples within them. Both samples exhibit continua in these projections, running from the lower right to the upper left of each plot, which have been segmented by FEM. That this segmentation is robustly reproducible over successive runs of FEM (Table 1) indicates that FEM has captured astrophysically meaningful structures in the samples. In addition, both samples exhibit a cluster which extends into the sparser region to the upper right of each plot. This overall similarity suggests that the evolution of galaxies at the epochs of the two samples is mostly similar. It also gives us confidence in the success of the measures taken to ensure a fair comparison between samples at different redshifts and from different surveys (see Sections 2.1.2 and 2.2.2), and reinforces our conclusion that FEM has overall used the input features similarly for both samples in spite of slight differences in the availability of photometry between them (see Section 4.2). The subtler differences between clusters in these projections are subject to the distributions of galaxies *within* the shapes of their respective samples. We comment on these differences where relevant in Section 4.4. Cluster colours in the plots in this paper, like their names, are assigned based on their mean $NUV - r$ colours.

We break down the analysis of our clusters by using the two-dimensional colour bimodality of galaxies as a simple framing device. The colour bimodality is a steady property of the galaxy population throughout cosmic time, having been observed among galaxies with redshifts as high as 4 (Wuyts et al. 2007; Williams et al. 2009; Ilbert et al. 2010, 2013). Hence, we may use it to separate clusters that are more strongly associated with the blue peak (containing mostly star-forming galaxies) from clusters that are more strongly associated with the red peak (containing mostly passive galaxies) in a way that is independent of redshift.

This two-dimensional separation is marked by the black lines in Fig. 4. The $NUV - r - K_s$ colour–colour plane (Arnouts et al. 2013; Moutard et al. 2016b) is a useful tool to probe galaxy subpopulations due to its ability to separate star-forming (low $NUV - r$), passive (high $NUV - r$), and also dusty (high $r - K_s$) galaxies. It has been applied in several studies of galaxy evolution using data from VIPERS (e.g. Fritz et al. 2014; Davidzon et al. 2016; Moutard et al. 2016b; Siudek et al. 2017, 2018b; Vergani et al. 2018). The form of the black lines is inspired by Fritz et al. (2014) and Moutard et al. (2016b); they are placed independently in each panel, without

Table 2. Profiles, in terms of averages, of the clusters determined within each of our samples. See the main text for an explanation of the cluster naming scheme. We list cluster means in columns $NUV - r$ and $r - K_s$. For the remaining features, which are less directly linked to the clustering, we opt for medians to mitigate the potential influence of outliers on the cluster profiles. Column '%' lists the percentage of galaxies contained within each cluster for each sample. The data in the next seven columns [$NUV - r$ to $\log_{10}(sSFR/\mathrm{yr}^{-1})$ (SED)] originate from the same CIGALE SEDs as the rest-frame colours that were used as inputs to the clustering. Features listed in this table include colour excesses [$E(B - V)$], stellar masses (M_*), stellar metallicities (Z), mass-weighted stellar ages ($MWSA$), and specific star formation rates ($sSFR$). We list sSFRs both determined by CIGALE (SED; averaged over 100 Myr time-scales) and determined from galaxy spectra (and hence independent of CIGALE; ind.; see Sections 2.1.2 and 2.2.2). Medians marked with asterisks have unexpected values given their corresponding $NUV - r$ colour and are discussed in Section 4.4.2.

Cluster	%	$NUV - r$	$r - K_s$	$E(B - V)$	$\log_{10}(M_*/\mathrm{M}_\odot)$	$\log_{10}(Z)$	$\log_{10}(MWSA/\mathrm{Myr})$	$\log_{10}(sSFR/\mathrm{yr}^{-1})$ (SED)	(ind.)
G1	24.0	2.39	0.42	0.11	9.90	− 2.22	3.80	− 9.87	− 9.87
G2	15.2	3.29	0.91	0.20	10.26	− 1.81	3.85	− 10.02	− 10.19
G3	17.3	3.51	0.78	0.14	10.37	− 2.11	3.89	− 10.38	− 10.47
G4	8.5	4.31	1.16	0.13	10.70	− 1.75	3.92	− 10.87	− 11.22
G5	9.7	5.07	0.67	0.22	10.35	− 2.30	3.90	− 10.78	− 11.97
G6	11.3	5.24	0.78	0.08	10.57	− 2.11	3.93	− 11.92	− 11.93
G7	14.0	5.27	0.73	0.11	10.54	− 2.20	3.93	− 11.85	− 12.02
V1	26.8	1.86	0.25	0.01	9.87	− 2.12	3.52	− 9.34	− 9.25
V2	18.4	2.17	0.60	0.02	10.14	− 1.90	3.55	− 9.22	− 9.34
V3	9.3	2.62	0.75	0.05	10.10	− 1.40	3.52	− 8.99	− 9.35
V4	18.5	3.26	1.05	0.12	10.67	− 1.80	3.58	− 9.71	− 9.92
V5	5.2	4.75	0.91	∗0.15	10.61	∗ − 1.51	∗3.52	∗ − 9.43	− 10.09
V6	10.3	4.81	0.90	∗0.15	10.69	∗ − 1.86	∗3.61	∗ − 9.90	− 10.29
V7	11.5	4.86	0.96	0.02	10.91	− 2.05	3.74	− 11.27	− 10.42

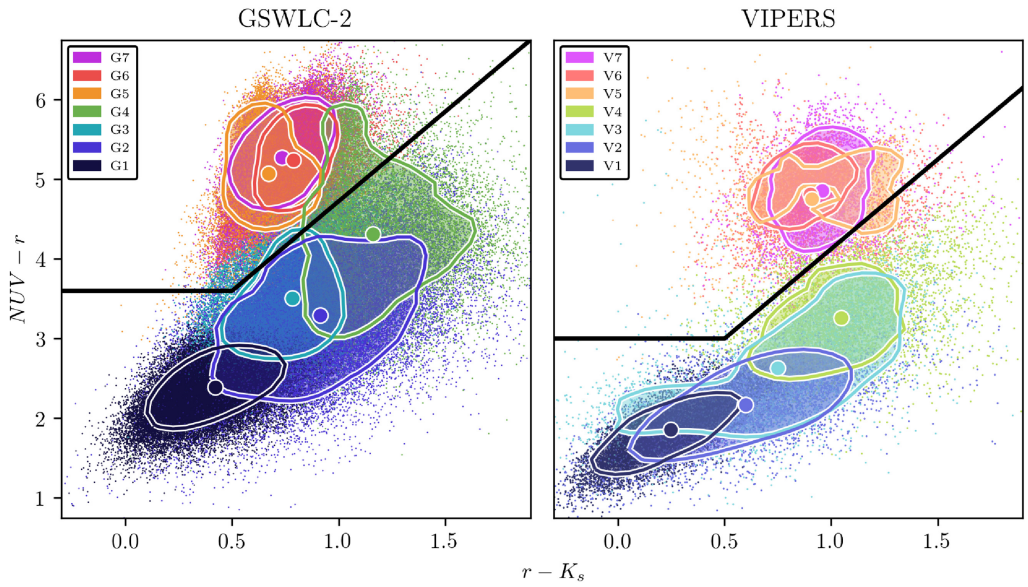

Figure 4. Colour–colour plots of our samples. Colours are derived from CIGALE SED estimation. The distributions of clusters are shown using coloured, filled contours (drawn at a relative density of 0.4), and the coloured, circular markers show their means. The black line in each plot (inspired by Moutard et al. 2016b; see main text) marks the boundary between star-forming galaxies (below the line) and passive galaxies (above the line).

reference to the positions of the clusters, to simply demarcate the star-forming and passive regions of the $NUV - r - K_s$ plane. Clusters whose means then lie below the black line in each plot are selected as 'blue', 'star-forming' clusters, and clusters whose means then lie above the black lines are selected as 'red', 'passive' clusters. As a result, both samples break down into four blue clusters and three red clusters. Deviations of the structures of the clusters from this simple blue/red (star-forming/passive) division that we enforce (e.g. clusters that overlap or span this division) will highlight limitations of a purely two-dimensional view of the galaxy population and its bimodality. The separation between these two main families of clusters suggests differences in the evolution processes influencing the galaxies that they contain.

The blue peak of the bimodality corresponds closely with the star-forming main sequence (SFMS; Noeske et al. 2007; Salim et al. 2007), which is the tight correlation between the SFRs and the stellar masses of actively star-forming galaxies. The SFMS, like the bimodality, is ubiquitous throughout cosmic time (Speagle et al. 2014). It has a lower normalization with decreasing redshift; this cosmological decline of star formation (Madau et al. 1996; Madau & Dickinson 2014; Driver et al. 2018) is visible as a vertical offset between the samples in Fig. 4. In this paper, the

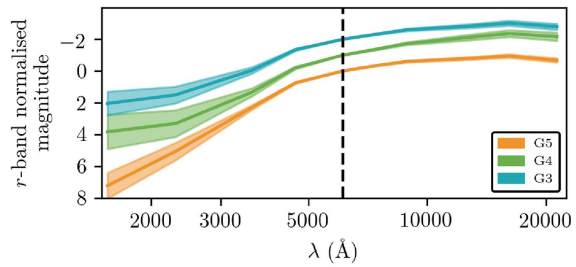

Figure 5. A comparison of the shapes of the mean (± standard deviation) estimated SEDs of galaxies in clusters G3, G4, and G5. Clusters G3 and G5 are chosen as they neighbour G4 in terms of their average $NUV - r$ colour. The estimated SEDs of individual galaxies are normalized by their r-band magnitudes (the effective wavelength of which is marked by a dashed black line) before the mean estimated SEDs are calculated. The y-axis applies to the mean SED of G5; those of G3 and G4 are vertically offset by −1 and −2, respectively, to more clearly show the differences in their shapes.

terms 'blue peak' and 'SFMS' are synonymous, and we use them interchangeably.

The stronger $NUV - r$ split between star-forming and passive VIPERS clusters in comparison with those of GSWLC-2 (Figs 4 and 3) is likely to result from two factors. First is the difference in the rest-frame wavelength coverage of *GALEX* photometry for the two samples; some rest-frame UV emission is redshifted out of the bandwidths of *GALEX*'s filters at $z \sim 0.65$. Second is the difference in the completeness of UV photometry for each sample. *GALEX* observations exist for ∼80 per cent of galaxies in clusters G1–4. This proportion falls to ∼55 per cent in clusters G5–7, but this is expected given that these galaxies would be fainter in the UV regime. Meanwhile, ∼65 per cent of V1, V2, and V4 galaxies were observed by *GALEX*. Interestingly, only ∼20 per cent of galaxies in V3 have observed UV photometry, which may explain its separation from the other star-forming VIPERS clusters. Passive VIPERS clusters are ∼25 per cent complete in observed UV photometry. Together, these factors mean we are likely to miss low levels of UV emission from more evolved VIPERS galaxies with more intermediate colours. On the other hand, Fig. 2 shows that rest-frame $NUV - u$ colours are similarly important to the clustering structures of *both* samples, with *NUV* emission expected to be a particularly accurate tracer of star formation (Salim 2014).

4.4 Cluster identities

4.4.1 Clusters of star-forming galaxies

Our $NUV - r - K_s$ cut (Section 4.3) yields the following blue clusters: G1, G2, G3, and G4 for the GSWLC-2 sample; and V1, V2, V3, and V4 for the VIPERS sample. While dominated by blue galaxies, clusters G4 and V4 also contain a significant number of galaxies with green or red $NUV - r$ colours (including the vast majority of green valley galaxies). Fig. 5 shows that the SEDs of G4 galaxies are generally more similar to those of actively star-forming galaxies, being flatter in the UV regime (e.g. G3 galaxies) than those of typically passive galaxies (e.g. G5 galaxies). Hence, in terms of the influence of their evolutionary histories on the shapes of their SEDs, G4 galaxies appear more closely related to G1–3 galaxies than to G5–7 galaxies, despite some G4 galaxies occupying the passive region of the $NUV - r - K_s$ plane in Fig. 4. Similarly, the SEDs of V4 galaxies more closely resemble those of V1–3 galaxies rather than V5–7 galaxies (not shown).

Given that the SFMS is a smooth continuum, it is important where possible to establish why FEM has distinguished clusters within it, and to interpret the significance of these distinctions in terms of galaxy evolution. The position of a galaxy along the $NUV - r - K_s$ SFMS (Fig. 4) is governed by a combination of its stellar mass and its dust content (Moutard et al. 2016a, b). The lobe at high $r - K_s$, which preferentially consists of edge on galaxies, is known to capture the excess reddening of high-mass star-forming galaxies (Arnouts et al. 2013), but it is more difficult to disentangle this combination of stellar mass and dust elsewhere within the SFMS. Hence, we see an overlap of star-forming clusters in Fig. 4. In Fig. 3, though, these clusters are more clearly separated.

G1 and V1 capture equivalent subpopulations of galaxies. Both clusters contain the galaxies with the bluest colours and the lowest masses (Fig. 4, Table 2) within their respective samples; star-forming galaxies at relatively early stages of their evolution. The remaining star-forming clusters have higher masses and lie further along the SFMSs of each sample.

Clusters G2 and G3 overlap with one another in the left-hand panel of Fig. 4, as do clusters V2 and V3 in the right-hand panel of the same figure. Fig. 3 shows that G2 and V3 both extend away from the main continua within the subspace projections of their respective samples. The feature vector projections in Fig. 3 show that the galaxies in these clusters have particularly red $FUV - NUV$ colours in comparison with other SFMS clusters. However, the astrophysical meaning behind this is unclear. CIGALE alternately attributes this reddening to high colour excesses for galaxies in G2 and to higher metallicities for galaxies in V3 (Table 2), suggesting that it has not fully resolved the degeneracy between the influences of dust and metallicity upon the colours of these galaxies. However, CIGALE is consistent in assigning G2 and V3 galaxies similar stellar masses and mass-weighted stellar ages to G3 and V2 galaxies (Table 2), which occupy similar regions of the $NUV - r - K_s$ plane. Stellar mass estimates are not strongly affected by an inability to resolve this degeneracy between the influences of dust and metallicity (e.g. Bell & de Jong 2001). Clusters G3 and V2, lying on the main continua in Fig. 3, seem to be intermediate between clusters G1 and G4, and V1 and V4, respectively.

The star-forming clusters along the SFMS of our GSWLC-2 sample exhibit a gradient in their star formation activity. Taking their increasing average stellar masses as a point of reference, clusters G1–4 exhibit a corresponding increase in their average $NUV - r$ colours (Table 2, Fig. 4). decrease in their average sSFRs (both SED and ind.; Table 2), and increase in their average $D(4000)$ (Fig. 6). High-mass galaxies in our GSWLC-2 sample do not form stars as readily as low-mass galaxies. This gradient is weaker for clusters V1–3 (particularly with regard to their median sSFRs; Table 2), though we note that clusters V2 and V3 have lower average stellar masses than G2 and G3. It is only in V4 that we see a rise in average stellar mass accompanied by a decrease in average *sSFR*, and an increase in $D(4000)$.

The large median sizes and low-to-intermediate median Sérsic indices of star-forming clusters from both samples indicate that they are dominated by disc galaxies (Table 3). Clusters G1–4 exhibit a rise in their median n_g to intermediate values along their SFMSs, indicating increasingly concentrated morphologies among their galaxies. In Fig. 7, these clusters form morphological sequences that are separate from the distributions of passive clusters in the same plane. The sequence of V1–4 is not as strong as that of G1–4; again, it is only in V4 that we see a significant change, with the higher stellar masses of its galaxies met with intermediate Sérsic indices.

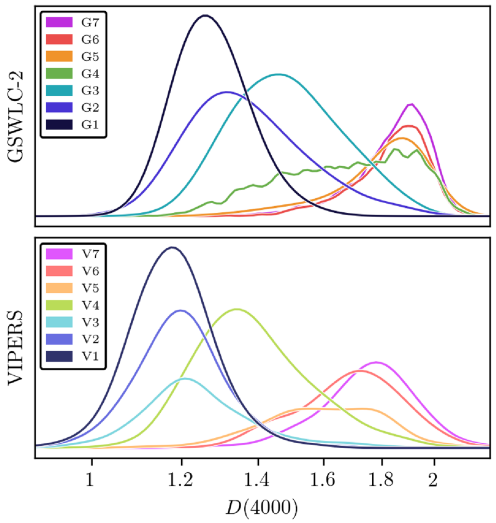

Figure 6. Smoothed kernel density estimates in $D(4000)$ (logarithmically distributed) for each of the clusters from both outcomes. Here, $D(4000)$ was measured from the spectra of galaxies (Brinchmann et al. 2004; Garilli et al. 2014) using a method introduced by Balogh et al. (1999), and is hence independent of CIGALE's estimated SEDs.

Table 3. Profiles, in terms of averages of ancillary features, of the clusters determined within each of our samples. See the main text for an explanation of the cluster naming scheme. We list the median values of the galaxies that the clusters contain for each of the features. Column '%' lists the percentage of galaxies contained within each cluster for each sample. Features listed in this table include Sérsic indices (n_g), half-light radii ($R_{1/2}$), and environmental overdensities (δ). The data are drawn from ancillary sources (see Sections 2.1.2 and 2.2.2).

Cluster	%	n_g	$\log_{10}(R_{1/2}/\mathrm{kpc})$	$\log_{10}(1+\delta)$
G1	24.0	1.04	0.57	0.40
G2	15.2	1.34	0.50	0.51
G3	17.3	1.57	0.55	0.55
G4	8.5	2.38	0.61	0.59
G5	9.7	4.09	0.40	0.85
G6	11.3	4.18	0.45	0.80
G7	14.0	4.25	0.44	0.83
V1	26.8	0.92	0.49	0.29
V2	18.4	0.95	0.48	0.29
V3	9.3	1.11	0.50	0.36
V4	18.5	1.53	0.55	0.35
V5	5.2	3.31	0.42	0.40
V6	10.3	3.29	0.40	0.40
V7	11.5	3.40	0.43	0.43

While there are slight trends in the median local environmental overdensities of the star-forming clusters in both samples (Table 3), Fig. 8 shows that their distributions thereof have very large spreads and exhibit a great deal of overlap with the distributions of other SFMS clusters from the same sample. Therefore, we cannot attribute the reduction in the star formation activity of SFMS galaxies at higher masses to mainly environmental causes for either sample.

4.4.2 Clusters of passive galaxies

Our red clusters, selected in two dimensions using the $NUV - r - K_s$ plots in Fig. 4, are: G5, G6, and G7 for our GSWLC-2 sample, and V5, V6, and V7 for our VIPERS sample. The colour that best separates the passive clusters in both samples is $FUV - NUV$. For G5–7, this separation corresponds with the higher sSFRs and lower masses of G5 galaxies, and differences in the metallicities of G6 and G7 galaxies (Table 2). V7 has been distinguished due to the high masses and low sSFRs of its galaxies. However, CIGALE's estimation of the astrophysical properties of V5 and V6 galaxies is less reliable (see below). In general, galaxies in the passive clusters are offset to redder $NUV - u$ colours than those in the SFMS clusters (Section 4.4.1).

Galaxies in clusters G6, G7, and V7 are alike with respect to most features. They share high stellar masses, low sSFRs, large $D(4000)$ (Fig. 6), and early-type morphologies (Table 2), all of which are typical of canonically passive galaxies. CIGALE attributes the difference in the $FUV - NUV$ colours of G6 and G7 galaxies (i.e. the feature that best separates these clusters) to their metallicity distributions. While G6 peaks strongly at $Z \sim -2.1$, G7 is split evenly between peaks at $Z \sim 2.1$ and $Z \sim -2.4$. The metallicities of passive GSWLC-2 galaxies are discretized by the input Bruzual & Charlot (2003) grid, and due to a lack of any input NIR photometry during their SED estimation (see Appendix B); with more precise metallicities, their distributions might overlap more. V7 also has low metallicities in comparison with other clusters determined in its sample. We note that these subsolar metallicities are unexpected for high-mass passive galaxies (e.g Gallazzi et al. 2006), indicating difficulties of breaking the age-dust-metallicity degeneracy with photometry alone, and suggesting that these metallicities are not entirely reliable. Altogether though, these clusters contain the oldest, most evolved galaxies among their respective samples: a subpopulation that is in place at the epoch of our VIPERS sample.

Galaxies in cluster G5, while also passive and early-type, have lower stellar masses than those in clusters G6 and G7. We also note a difference in the G5 median sSFRs as reported by CIGALE (SED) and by the Brinchmann et al. (2004) calibration (ind.; Table 2). G5 may contain post-starburst galaxies (PSBs; Wild et al. 2009), with this difference in sSFRs possibly arising due to the different time-scales probed by these two measures (see section 7 of Salim et al. 2016). While the fibre component of *sSFR* (ind.) is a more instantaneous measure of star formation activity (∼10 Myr, based on Hα emission), CIGALE averages star formation over a longer period of time (100 Myr, matching the time-scale traced by UV emission). Hence, even if the tail of a declining central burst of star formation activity is not captured by *sSFR* (ind.), it may still be captured by *sSFR* (SED). The spheroidal morphologies (Fig. 7, Table 3) and enhanced local environmental densities of G5 galaxies suggest an external influence upon their evolution (see Section 5.2), which is consistent with previous studies which link PSBs with mergers (Zabludoff et al. 1996; Yang et al. 2008; Almaini et al. 2017).

Clusters V5 and V6 present conflicting identities in terms of features estimated by CIGALE (Table 2). While their galaxies have very similar stellar masses and morphologies to those in V7 (Table 3), they have unusually high colour excesses, metallicities, and *sSFR* (SED). This is in contrast with the *sSFR* (ind.) and observed $D(4000)$ values of these galaxies (Table 2, Fig. 6), which show that they are indeed passive. The large spread in $D(4000)$ of V5 may be due to some minor contamination of the cluster by star-forming galaxies; its $NUV - r - K_s$ contour extends below the black line in Fig. 4, into the region containing dusty star-forming galaxies. This may also drive its median $E(B-V)$ to a higher value.

The inability of CIGALE to properly resolve the age-dust-metallicity degeneracy for V5 and V6 galaxies is due to the UV

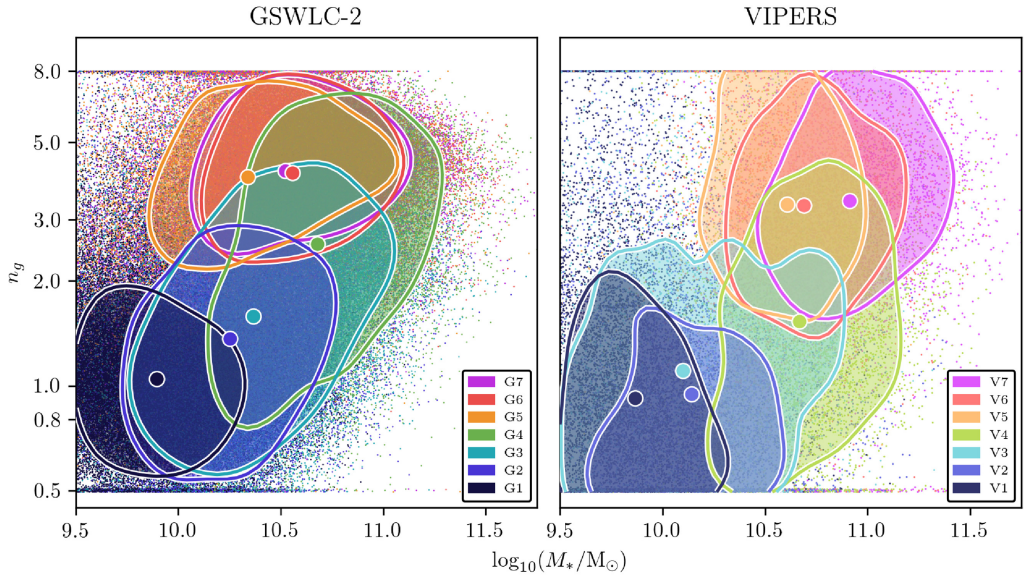

Figure 7. Sérsic index versus stellar mass for the galaxies in our samples. Sérsic indices were determined by Simard et al. (2011) for our GSWLC-2 sample, and Krywult et al. (2017) for our VIPERS sample. The distributions of clusters are shown using coloured, filled contours (drawn at a relative density of 0.4), and the coloured, circular markers show their medians. We have winsorized the Sérsic indices of the galaxies in our VIPERS sample to values of 0.5 and 8 in order to match the limits of our GSWLC-2 sample.

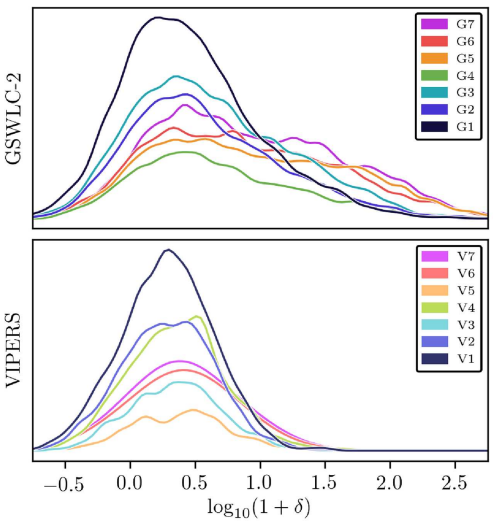

Figure 8. Smoothed kernel density estimates in local environmental overdensity (δ) for each of the clusters from both outcomes. For both samples, these overdensities are based on fifth-nearest neighbour surface densities (Baldry et al. 2006; Cucciati et al. 2017).

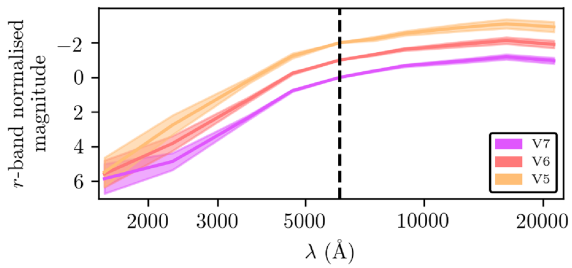

Figure 9. A comparison of the shapes of the mean (± standard deviation) estimated SEDs of galaxies in clusters V5, V6, and V6. The estimated SEDs of individual galaxies are normalized by their r-band magnitudes (the effective wavelength of which is marked by a dashed black line) before the mean estimated SEDs are calculated. The y-axis applies to the mean SED of V7; those of V5 and V6 are vertically offset by −1 and −2, respectively, to more clearly show the differences in their shapes.

regions of their SEDs. Fig. 9 shows that V5 and V6 have steeper average UV SEDs than V7. To explain the red UV colours (especially $FUV - NUV$) of their galaxies, CIGALE invokes high colour excesses and metallicities rather than low *sSFR* (SED). This appears to be a consequence of CIGALE's two-burst SFHs, which may not be a realistic description of the SFHs of most passive VIPERS galaxies. These SFHs were adjusted for the epoch of our VIPERS sample by setting the formation time of the old population to 6.5 Gyr ago instead of 10 Gyr, and including the possibility of a particularly recent burst of star formation (<50 Myr). However, a trial of the use of a gradual 1 Gyr quenching episode instead led to improvements in the quality of fit of passive SEDs (with low *sSFR*) to the photometry of the majority of V5 and V6 galaxies. Hence, it seems that further adjustments to CIGALE's SFH prescription are required when applying it at higher redshifts.[7]

Galaxies contained within the passive clusters of our VIPERS sample tend to have higher stellar masses than those contained within the passive clusters of our GSWLC-2 sample (Table 2). This is likely to be driven by stellar mass incompleteness of our VIPERS sample. Davidzon et al. (2013) show that, even at its lower redshift limit of $z = 0.5$, our VIPERS sample is incomplete[8] in passive galaxies below

[7] `LePhare` (Ilbert et al. 2006) SED estimation for the same galaxies (Moutard et al. 2016b; Siudek et al. 2018b) used single exponentials for its SFHs and reported lower colour excesses, metallicities, and *sSFR*.

[8] Our 'secure' redshift criterion (Section 2.2) may contribute slightly to this incompleteness (i.e. by selecting against faint, passive VIPERS galaxies that lack emission lines or strong absorption lines). However, Davidzon et al. (2013) used a more relaxed criterion, so we do not expect our use of this criterion to significantly influence stellar mass completeness.

$\sim 10^{10}$ $M_\odot$. Furthermore, their completeness threshold increases with redshift to $10^{10.75}$ $M_\odot$ at our upper limit of $z = 0.8$, and thus skews our clusters of passive VIPERS galaxies towards higher stellar masses.[9] Hence, where the GSWLC-2 sample has two lobes of passive galaxies in Fig. 3 (see also Appendix B), which differ in average stellar mass by ∼0.5 dex, the VIPERS sample has only one. Though our VIPERS sample does contain some passive galaxies with low stellar masses (e.g. Fig. 7), they are not substantial enough in number for FEM to model them with a dedicated cluster (i.e. like G5).

Passive clusters in both samples have high Sérsic indices and compact sizes (Table 3), indicating spheroid-dominated morphologies. They occupy separate regions of the plots in Fig. 7 to their respective SFMS clusters. Fig. 7 also shows that the n_g distributions for passive clusters are highly consistent with one another. While the passive clusters in our GSWLC-2 sample exhibit a slight offset to higher density environments in comparison with star-forming GSWLC-2 clusters, the environments of passive VIPERS clusters are consistent with those of star-forming VIPERS clusters. This difference between the two samples is, in part, expected, due to the emergence of environments of especially high densities over cosmic time (e.g. Marinoni et al. 2008; Kovač et al. 2010; Fossati et al. 2017). However, factors such as spectroscopic fibre collisions and the aforementioned incompleteness of passive VIPERS galaxies may also reduce the completeness of VIPERS at high densities. This incompleteness does not appear to have strongly affected clusters elsewhere in the feature space (Fig. 3).

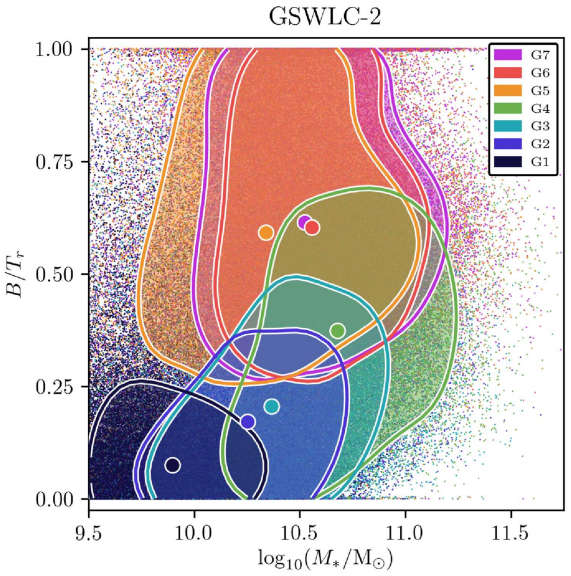

Figure 10. Bulge-to-total ratio (B/T_r) versus stellar mass for the galaxies in our GSWLC-2 sample. Here, the subscript 'r' denotes the r-band photometry from which the ratios were derived (Simard et al. 2011; based on two-component fits). The distributions of clusters are shown using coloured, filled contours (drawn at a relative density of 0.4), and the coloured, circular markers show their means.

5 DISCUSSION

Our clusters have been determined on the basis of the rest-frame colours of galaxies alone. In this section, we aim to discern what the trends of these purely colour-based clusters with other, ancillary features (see Sections 4.4.1 and 4.4.2) tell us about how strongly the SEDs of their constituent galaxies encode their evolution.

5.1 Internally driven evolution

Alongside being closely related in terms of the shapes of their SEDs (see Section 4.4.1), both sets of star-forming clusters – G1–4 and V1–4 – form clear morphological sequences in Fig. 7. In Fig. 10, we examine the bulge-to-total ratios of GSWLC-2 galaxies using two-component Simard et al. (2011) fits (no such data exist for VIPERS). The G1–4 sequence is apparent here as well, capturing the rising prominences of the bulges of their galaxies. It does *not* extend to the highest B/T_r values, despite G4 also containing some quenching and quenched galaxies. This indicates that G1–4 galaxies retain their discs as they evolve and that some G1–4 galaxies become passive without fully transforming their morphologies. The changing bulge–disc balance appears to be captured also in the large spread in $D(4000)$ of G4 galaxies in particular (Fig. 6). The overlapping environmental distributions of star-forming clusters in both samples (Fig. 8) suggest that these morphological sequences of gradual bulge growth are more likely to be due to internal processes (i.e. that act in all environments). We assume that our interpretation in this paragraph applies to galaxies in V1–4 as well.

Bar-driven inflows of star-forming gas (Sheth et al. 2005) – an internal process that acts over long time-scales – constitute a likely candidate process. These inflows are commonly invoked to explain the formation of dynamically cold 'pseudobulges' ($n_{\rm cl} \lesssim 2$) rather than the dynamically hot 'classical' ($n_{\rm cl} \gtrsim 2$) bulges that the Simard et al. (2011) two-component fits assume (Kormendy & Kennicutt 2004; Fisher & Drory 2008; Mishra, Wadadekar & Barway 2017). However, an increase in the prominence of pseudobulges would none the less be expected to be captured by the single-component fits which yield the Sérsic indices in Table 3 and Fig. 7. We do not rule out that SFMS galaxies may have undergone major and/or minor mergers or clump migration (a faster, more violent internal process; Elmegreen, Bournaud & Elmegreen 2008; Bournaud et al. 2011; Tonini et al. 2016) in their pasts; some have high total n_g values, which may capture classical bulges formed as a result of these processes. Instead, we proffer that the processes do not contribute to the gradual of the bulges of these galaxies. It has been shown, for example, that the remnant of a gas-rich merger can reform a disc and continue to form stars, thus rejoining the SFMS (Hopkins et al. 2009a, b).

The falling sSFRs of galaxies along the sequences G1–4 and V1–4 suggests that their morphologies are also linked with their quenching. This could be due to morphological quenching (i.e. the gravitational influence of the morphological components of galaxies upon star formation; Martig et al. 2009). It is more likely, though, that the prominences of the bulges among these galaxies are a marker of nuclear activity. More massive bulges host more massive black holes at their centres (Häring & Rix 2004), which supply more feedback energy to their surrounding galaxies. This feedback can inhibit further star formation by ejecting star-forming gas (Croton et al. 2006; Gabor et al. 2011; Vergani et al. 2018) or by preventing the cooling of newly accreted gas (above the 'transition mass', $\sim 10^{10.5}$ $M_\odot$ at $z \sim 0$; Kauffmann et al. 2003; Dekel & Birnboim 2006; Kereš et al. 2009; Moutard et al. 2020).

Fig. 11 shows the distributions of clusters G1–4 within the Lamareille (2010) emission-line classification diagram. This diagram is chosen with a view to its applicability to galaxies at higher redshifts as well. The equivalent widths of the relevant emission lines were determined by Brinchmann et al. (2004), and are available for

[9] Star-forming galaxies and clusters are affected to a much lesser degree.

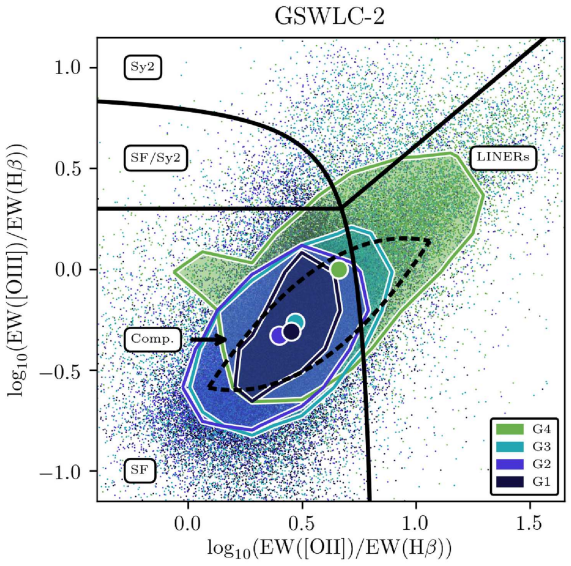

Figure 11. A diagram for the classification of emission-line galaxies (Lamareille 2010) in our GSWLC-2 sample. Different regions, labelled and demarcated by black lines, correspond to different types of galaxy: 'Sy2' to type 2 Seyfert galaxies, 'SF' to purely star-forming galaxies, 'SF/Sy2' to a mixture of type 2 Seyfert and star-forming galaxies, 'LINERs' to galaxies containing low-ionization nuclear emission-line regions, and 'Comp.' to a mixture of LINERs and star-forming galaxies. The distributions of clusters are shown using coloured, filled contours (drawn at a relative density of 0.4), and the coloured, circular markers show their medians.

94 per cent of the galaxies in G1–4. Spectroscopy of these emission lines exists for some VIPERS galaxies as well (Garilli et al. 2014), but only for 34 per cent of them, such that we would not be confident in the significance of any trend of our VIPERS clusters within the diagram. We note, however, that the few VIPERS galaxies for which this spectroscopy does exist tend to lie within the 'SF' region of the plot, above the 'Comp.' region (i.e. as in fig. 10 of Siudek et al. 2018b). Hence, we tentatively suggest a minimal influence of active galactic nuclei upon their current evolution, but reiterate that more data are needed to confirm this.

Clusters G1–4 are all centred in the 'Comp.' region of Fig. 11, indicating that galactic nuclei are prevalent throughout them. G4 in particular extends well into the 'LINERs' region of the diagram. Given the enhancement in the Sérsic indices of G4 galaxies over G1–3 galaxies (Table 3, Fig. 7), this is consistent with previous studies which find that low-ionization nuclear emission-line regions are more common in galaxies with earlier type morphologies (Heckman 1980). In addition, this increase in nuclear activity for G4 galaxies coincides with their decrease in *sSFR* in comparison with G1–3 galaxies (Table 2), supporting the suggestion that supermassive black holes are involved in their quenching.

That the sSFRs of V1–3 galaxies do not decline as strongly as those of G1–3 galaxies may be tied to their morphologies; all three also have very low median n_g. This suggests that their bulges and/or supermassive black holes have not yet grown to the extent that they can effectively inhibit star formation. This would be consistent with Fang et al. (2013) and Bluck et al. (2014), who find that bulges must exceed a threshold in mass or central density before they become associated with quenching. For V4 galaxies, the reduction in sSFR is met with a rise to intermediate median n_g, suggesting that this threshold bulge mass has been achieved in some V4 galaxies.

Altogether, G1–4 and V1–4 galaxies (which include the vast majority of green valley galaxies) appear to evolve slowly and secularly (Schawinski et al. 2014; Ilbert et al. 2015; Moutard et al. 2016b; Pacifici et al. 2016). This is reflected in the similarity of their SEDs, which all feature relatively flat UV regions that suggest a gradual reduction in their star formation over time. It is also reflected in the morphological sequences that their clusters exhibit. The rising bulge prominences and declining star formation rates of these galaxies suggests that nuclear feedback, fuelled by bar-driven inflows, is the main mechanism driving their evolution (Gabor et al. 2011; Moutard et al. 2020). While this mechanism appears to act at the epochs of both samples, the connection between morphologies and star formation is stronger at lower redshifts. This may be linked with the long time-scales over which these internal processes act, such that the gradual evolution of V1–4 galaxies may eventually lead to the more evolved distribution of galaxies given by clusters G1–4, which we assume to be their descendants. Hence, the rising prevalence of bulges grown by internal processes over cosmic time (e.g. Bruce et al. 2012; Gu et al. 2019) would appear to be linked to the cosmic decline of cosmic star formation activity. This connection between the bulges and the star formation of SFMS galaxies has previously been established (Cheung et al. 2012; Fang et al. 2013; Bluck et al. 2014; Cano-Díaz et al. 2019; McPartland et al. 2019), but in our case it emerges purely from our clustering of galaxy colours, with morphologies invoked post-clustering for interpretation. Our clustering also appears to demonstrate that the SFMS is a two-dimensional projection of this pathway which, in the full nine-dimensional colour space, extends continuously to also include high-mass passive galaxies (as revealed by G4 in particular) that retain their discs, but are degenerate with other passive galaxies in two dimensions.

5.2 Satellite quenching at low redshifts

The uniformly red $NUV - u$ colours and the uniformly high Sérsic indices of galaxies in clusters G5–7 and V5–7 imply a strong link between their passiveness and their concentrated morphologies. At high masses, this link may be obfuscated by a contribution from the internally driven evolutionary pathway that we propose in Section 5.1. We note that cluster V7 in particular, containing VIPERS galaxies with the highest masses, seems to align well with the sequence of clusters V1–4 in Fig. 7, such that it could partially be an extension of this evolutionary pathway consisting of the oldest galaxies with the most prominent bulges. This is in agreement with previous studies which find that the inner stellar density of galaxies is a successful predictor of its having been quenched (Cheung et al. 2012; Fang et al. 2013; Bluck et al. 2014).

However, other passive clusters are separated from their respective sequences of star-forming clusters in Fig. 7. Clusters G7, G6, and especially G5 (the latter containing the lowest mass passive galaxies in our GSWLC-2 sample) have high median n_g in comparison with other clusters centred at similar stellar masses (G2, G3). This separation invites the interpretation that their galaxies are subject to alternative or additional evolutionary processes. That these clusters contain those GSWLC-2 galaxies that occupy the highest density environments (Fig. 8) suggests an additional influence of external processes. Hence, we suspect that a significant proportion of galaxies among G5–7 are satellite galaxies (occupying the haloes of more massive central galaxies; Ilbert et al. 2010; Muzzin et al. 2013; Moutard et al. 2018). There is a weaker morphological separation for V5–7, and no environmental offset, which we attribute mostly to the incompleteness of low-mass passive galaxies in our VIPERS

sample; these would also be expected to trace high-density environments. Hence, our following discussion on the influence of external processes upon satellite galaxies is conducted with reference to G5–7 only. Fully establishing whether external processes influence the evolution of low-mass passive galaxies at $z \sim 0.65$ in the same way requires a more complete sample.

Major and minor mergers (Toomre 1977; Barnes 1988, 1992; Walker, Mihos & Hernquist 1996) and harassment (Moore et al. 1996; Smith et al. 2015), more common in environments of higher densities (Renzini 1999; Tonini et al. 2016), are external processes which can increase the Sérsic indices of galaxies by transforming their morphologies from disc- to spheroid-dominated (Aceves, Velázquez & Cruz 2006; Naab & Trujillo 2006; Fisher & Drory 2008). Fig. 10 shows a range of bulge-to-total ratios among galaxies in G5–7, which may be capturing the varying degrees to which these processes disrupt their morphologies. While most G5–7 galaxies are strongly spheroid-dominated, others (while still having high Sérsic indices) retain a disc component (with B/T_r values as low as ∼0.3). Whether these processes are also responsible for the quenching of G5–7 galaxies is unclear. Gravitational interactions between merging galaxies can induce central starbursts which rapidly exhaust their supplies of star-forming gas (e.g. PSBs, which we suggest comprise G5), and/or can catalyse nuclear activity which inhibits further star formation (Mihos & Hernquist 1994a, b, 1996; Di Matteo, Springel & Hernquist 2005; Springel, Di Matteo & Hernquist 2005a, b). However, a sufficiently gas-rich major merger may instead lead its remnant to form with a disc and continue forming stars (Barnes 2002; Hopkins et al. 2009a, b, 2010). In addition, a merger remnant may accrete new gas such that it can form a new disc and renew star formation (Salim & Rich 2010; Gabor et al. 2011). Generally, mergers cannot be unequivocally linked with the quenching of galaxies (see also Weigel et al. 2017), and so it is more likely that galaxies are quenched mainly by other processes.

Several external processes have been proposed to explain the quenching of star-forming galaxies as they become satellites. Examples include ram-pressure stripping (Gunn & Gott 1972; McCarthy et al. 2008), thermal evaporation (Cowie & Songaila 1977; Nipoti & Binney 2007), and viscous stripping (Nulsen 1982; Kraft et al. 2017), all of which invoke the removal of the cold interstellar medium of a galaxy via its hydrodynamical interaction with the hot intergalactic medium of high-density environments as the reason for quenching. These processes are correlated with the velocity of a galaxy as it travels through its environment, and generally quench galaxies quickly. Gas may also be removed from the extended halo of a galaxy at the outskirts of a dense environment, *by* the gravitational influence of that environment as a whole ('strangulation' or 'starvation'; Larson, Tinsley & Caldwell 1980; Peng, Maiolino & Cochrane 2015). The galaxy then quenches slowly by exhausting any remaining gas in its disc. The balance of these processes is not yet known (Bahé & McCarthy 2015; Peng et al. 2015; Smethurst et al. 2017), but recent studies advocate for a general 'delayed-then-rapid' quenching pathway (Wetzel, Tinker & Conroy 2012; Wetzel et al. 2013; Muzzin et al. 2014; Moutard et al. 2018). Galaxies initially quench slowly at the outskirts of the environment, then quickly as they approach its core, where the conditions for the aforementioned hydrodynamical interactions are expected. This delay could also explain the large spreads in the environmental distributions among all of our clusters in Fig. 8. These quenching processes are, in turn, unlikely to transform the morphologies of low-mass passive galaxies (Bekki, Couch & Shioya 2002; Boselli et al. 2009; Zinger et al. 2018).

In all, the separation of clusters G5–7 from G1–4 in terms of both their galaxies' colours (i.e. those use as an input to the clustering, in particular their $NUV - u$ and $NUV - r$ colours) and morphologies (i.e. their higher Sérsic indices), implies that their galaxies are subject to additional evolutionary processes. Hence, we suggest that the strong overlap between the passivity and the morphologies of G5–7 galaxies appears to be a product of different sets of environmental processes, which drive their quenching and morphological transformation separately (Poggianti et al. 1999; Kelkar et al. 2019). In addition, it implies that the quenching of galaxies precedes, or at least be simultaneous to, their morphological transformation (Schawinski et al. 2014; Woo et al. 2017). While the merger of two gas-rich, star-forming galaxies may produce a rejuvenated remnant, mergers between passive progenitors will invariably produce passive remnants with increasingly spheroidal morphologies, ranging from lenticular galaxies with classical bulges (Mishra et al. 2017; Mishra, Wadadekar & Barway 2018, 2019) through to pure spheroids.

5.3 Clusters in the size-mass plane

Fig. 12 shows the size-mass distribution of the clusters in each of our samples. The stellar masses originate from the same CIGALE SEDs that were used to generate the colours with which we represent the galaxies for the clustering, and the half-light radii from fits of single Sérsic profiles (see Sections 2.1.2 and 2.2.2). The size of a galaxy, in the context of its stellar mass and its morphology, is another important record of its assembly history. The positions and distributions of both sets of clusters in these plots match well with broader blue versus red, and early- versus late-type distinctions made in the same (or similar) plane(s) by other studies (Shen et al. 2003; van der Wel et al. 2014; Lange et al. 2015). This result again demonstrates that FEM, via just the nine input colours, is able to identify subpopulations that are degenerate in two dimensions and that are ordinarily distinguished using a combination of photometric and morphological information.

The most significant difference between the two plots in Fig. 12 is the absence of compact massive galaxies in our GSWLC-2 sample in comparison with our VIPERS sample. The canonical explanation for the growth of these galaxies is ongoing minor merger activity and accretion (Naab, Johansson & Ostriker 2009; Hopkins et al. 2010). The resultant shift between the passive VIPERS clusters and the passive GSWLC-2 clusters is approximately in accordance with the expected redshift evolution of the size–mass relation for early-type, passive galaxies (van Dokkum et al. 2015), though the mass-incompleteness of passive VIPERS galaxies means that we are unlikely to have precisely captured this shift in this paper. The large overlap of G4 and V4 with their respective passive clusters in Fig. 12 seems to support the additional 'late-track' (late with respect to cosmic time rather than to morphology) of galaxy evolution proposed by Barro et al. (2013) to yield disc-dominated passive galaxies (Ilbert et al. 2010; Carollo et al. 2013; Schawinski et al. 2014). Both sets of SFMS clusters are similarly distributed, capturing the minimal evolution of the sizes of star-forming galaxies between their two redshifts (Lilly et al. 1998; van der Wel et al. 2014).

6 SUMMARY AND CONCLUSIONS

We present results from the application of the FEM clustering algorithm to samples of galaxies at low ($z \sim 0.06$, from GSWLC-2) and intermediate ($z \sim 0.65$, from VIPERS) redshifts. Galaxies are represented using nine UV-through-NIR broadband rest-frame colours, derived from fits of ensembles of synthetic spectra to observed photometry with CIGALE. Using unsupervised machine

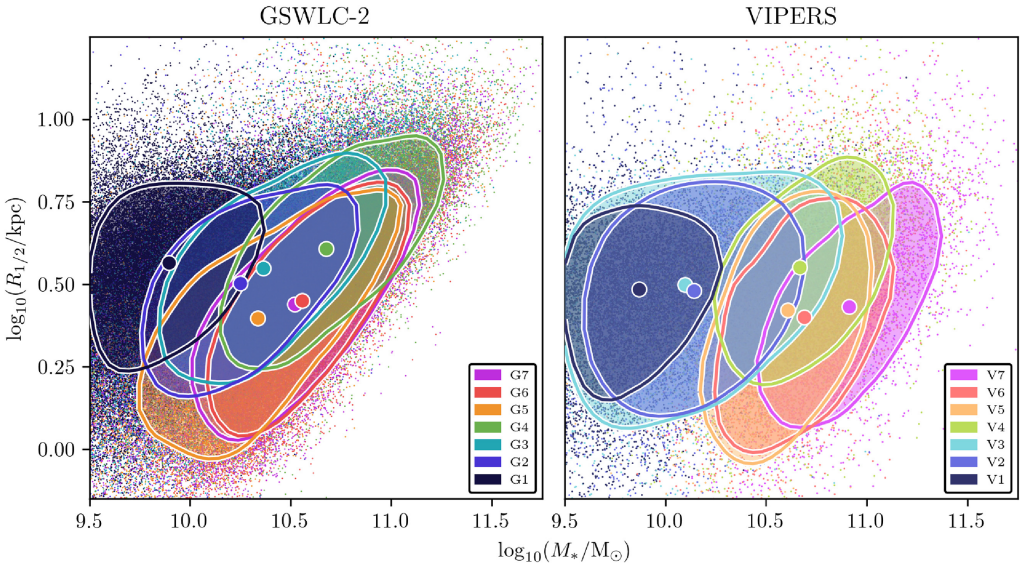

Figure 12. Half-light radius versus stellar mass for the galaxies in our samples. Circularized half-light radii are calculated from single Sérsic fits by Simard et al. (2011) for our GSWLC-2 sample, and Krywult et al. (2017) for our VIPERS sample. The distributions of clusters are shown using coloured, filled contours (drawn at a relative density of 0.4), and the coloured, circular markers show their medians.

learning to characterize the structures of our samples in this nine-dimensional feature space, our aims (following Siudek et al. 2018b) were to understand the evolution of subpopulations of galaxies in terms of these colours over cosmic time, and to establish how strongly these colours alone encode the assembly histories of galaxies. An advantage of FEM is its incorporation of dimensionality reduction on the fly, which ensures that it determines clusters using only the most important and discriminative information available among the input features. We summarize our results as follows:

(i) Our cluster evaluation search reveals that both of our samples are best partitioned into seven clusters (Table 1). In addition, the best-fitting submodels to each of our samples, identified independently, are closely related, both allowing variation in the shapes of clusters and differing only in their treatment of 'noise' among the input features. For both samples, these seven clusters break down into four 'blue' clusters containing mostly star-forming galaxies (and the vast majority of green valley galaxies), and three 'red' clusters containing mostly passive galaxies (Fig. 4). These two families of clusters are clearly separable, both in terms of the input colours to the clustering as well as in terms of ancillary features, which suggests differences in the evolution of their galaxies. Clustering outcomes in general are highly robust and reproducible.

(ii) Overall, FEM uses the nine rest-frame colours similarly to determine the partitions (Fig. 2), reducing the dimensionality of the feature space to 6 in both cases. Altogether, optical colours are most important to the clustering; individually, UV colours are. The availability of photometry with which to constrain the SEDs of galaxies is advantageous to the clustering. UV colours are slightly more important to the clustering in our GSWLC-2 sample, which has more *GALEX* coverage than our VIPERS sample. Similarly, the lack of any NIR coverage for our GSWLC-2 sample means that NIR colours are less important to its clustering. However, given the broader overall similarity between the clustering structures of the samples (Fig. 3), it appears that clustering (a statistical method) combined with SED estimation (which can infer rest-frame magnitudes from incomplete photometry) has enabled us to partially 'fill the gaps' of missing data in our samples.

(iii) Blue clusters (containing mostly star-forming galaxies and the vast majority of green valley galaxies) in both samples form clear morphological sequences (Fig. 7). The correlation between their median Sérsic indices and their median stellar masses captures the growth of the bulges of their galaxies along the SFMS (Fig. 10). At the highest masses, this growth corresponds with a drop in sSFRs. Hence, the quenching of high-mass galaxies is influenced by their inner stellar densities, above a certain threshold, which appears to be linked with nuclear activity (Fig. 11). The retention of discs by the highest mass galaxies along this morphological sequence indicates that some galaxies quench without fully transforming their morphologies. The lack of a strong trend of these clusters with local environmental overdensity (Fig. 8) suggests that this evolutionary pathway is dominated by internal processes. This pathway, prominent at the epochs of both samples, appears consistent with 'mass quenching', as proposed by Peng et al. (2010). In addition, the SFMS appears to be a two-dimensional projection of this pathway which, in nine dimensions, extends all of the way to high-mass passive galaxies that retain their discs. We expect that the long time-scales involved would ultimately lead the VIPERS star-forming clusters to resemble the GSWLC-2 star-forming clusters by the present day.

(iv) Red clusters (containing mostly passive galaxies) are clearly separate from their corresponding sequences of blue clusters. Galaxies in red clusters in both samples have uniformly high Sérsic indices, indicating a fundamental link between centrally concentrated morphologies and passiveness (Fig. 7). Passive clusters in our low-redshift sample are separated from their respective sequence of star-forming clusters, particularly towards lower stellar masses (Figs 7 and 10). We assume that this separation originates from the influence of alternative or additional processes to those that dictate the evolution of actively star-forming galaxies. Invoking the offset of these low-redshift passive clusters to high local environmental overdensities (Fig. 8), we suggest that some of their galaxies are satellites, and subject to external processes. The homogeneity of their early-type morphologies implies that their quenching precedes, or is at least simultaneous to, their morphological transformation. In all, this pathway appears consistent with 'environment quenching' (Peng

et al. 2010). This morphological separation is not as apparent for the passive clusters in our VIPERS sample (Fig. 7), which is mainly due to incompleteness of low-mass passive galaxies (which would also be expected to trace high-density environments). Hence, we are prohibited from commenting on the prevalence of this evolutionary pathway at intermediate redshifts.

Our study appears to confirm the existence of two distinct evolutionary pathways of galaxies through the green valley (Poggianti et al. 1999; Faber et al. 2007; Peng et al. 2010; Barro et al. 2013; Fritz et al. 2014; Schawinski et al. 2014; Moutard et al. 2016b). We re-emphasize that while much of our interpretation involves the use of ancillary features (and especially morphological information), the separation of the clusters into two main families of blue/green and red clusters originates in the colours used as inputs to the clustering. Hence, these pathways appear to be strongly encoded within the SEDs of galaxies. Our results invite further investigation into the extent to which a galaxy's assembly history may be discerned purely from its SED.

The use of further ancillary features would be instrumental in further substantiating and constraining these pathways. A wealth of such features are available for our GSWLC-2 sample, due to its basis in SDSS. Examples include Galaxy Zoo 2 morphologies (Willett et al. 2013) which include bar and merger classifications, and Yang et al. (2007) group memberships to enable a distinction between central and satellite galaxies. A more detailed analysis of our low-redshift sample in this manner is reserved for a future study. We note that the Galaxy And Mass Assembly project (Driver et al. 2009) could provide an alternative low-redshift sample, given its panchromatic data release (Driver et al. 2016) and its rich library of value-added catalogues (Baldry et al. 2018). The upcoming Deep Extragalactic VIsible Legacy Survey (DEVILS; Davies et al. 2018), which aims to improve completeness at $0.3 < z < 1.0$, could be the basis for an improved intermediate-redshift sample upon its completion. Furthermore, the Legacy Survey of Space and Time (Ivezić et al. 2019), which will provide galaxy colours and morphologies together, constitutes a particularly promising foundation for a future follow-up study.

The incompleteness of low-mass passive galaxies at intermediate redshifts would be alleviated by moving to deeper surveys such as G10-COSMOS (Andrews et al. 2017) and 3D-*HST* (Momcheva et al. 2016), both of which also have panchromatic photometric data releases. This would enable an examination of environment quenching at earlier epochs, and of its proposed increase in prevalence at lower redshifts (Fossati et al. 2017; Moutard et al. 2018; Papovich et al. 2018). Surveys like this could also extend our comparison to redshifts as high as $z \sim 2$, thus facilitating the constraint of the changing balance of evolutionary pathways, informed by clustering of rest-frame colours, over a greater extent of cosmic time.

ACKNOWLEDGEMENTS

We thank Steven Bamford, Olga Cucciati, Marie Martig, Tsutomu T. Takeuchi, and our referee for feedback which improved the quality of our manuscript.

The work presented in this paper was conducted using the following softwares: the SCIKIT-LEARN (Pedregosa et al. 2011), MATPLOTLIB (Hunter 2007), SCIPY (Jones et al. 2001), and NUMPY (Oliphant 2006; Harris et al. 2020) packages for the `Python 3` programming language (https://www.python.org); the FISHER-EM subspace clustering package (Bouveyron & Brunet 2012; known in this paper as `FEM`) for the `R` statistical computing environment (R Core Team 2019); and the Starlink Tool for OPerations on Catalogues And Tables (TOPCAT; Taylor 2005).

The construction of the GALEX-SDSS-WISE Legacy Catalogue (GSWLC) was funded through NASA award NNX12AE06G.

The VIMOS Public Extragalactic Redshift Survey (VIPERS) was performed using the ESO Very Large Telescope, under the 'Large Programme' 182.A-0886. The participating institutions and funding agencies are listed at http://vipers.inaf.it.

Funding for SDSS-III was provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III website is http://www.sdss3.org.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

ST has been supported by a United Kingdom Science and Technology Facilities Council postgraduate studentship. MS has been supported by the European Union's Horizon 2020 Research and Innovation programme under the Maria Sklodowska-Curie grant agreement (No. 754510), the Polish National Science Centre (UMO-2016/23/N/ST9/02963), and the Spanish Ministry of Science and Innovation through the Juan de la Cierva-formacion programme (FJC2018-038792-I). KM has been supported by the Polish National Science Centre (UMO-2018/30/E/ST9/00082). This research has been supported by the Polish National Science Centre (UMO-2018/30/M/ST9/00757) and the Polish Ministry of Science and Higher Education (DIR/WK/2018/12).

DATA AVAILABILITY

The data underlying this article will be made available upon reasonable request to the corresponding authors.

REFERENCES

Aceves H., Velázquez H., Cruz F., 2006, MNRAS, 373, 632
Ahn C. P. et al., 2014, ApJS, 211, 17
Almaini O. et al., 2017, MNRAS, 472, 1401
Andrews S. K., Driver S. P., Davies L. J. M., Kafle P. R., Robotham A. S. G., Wright A. H., 2017, MNRAS, 464, 1569
Arnouts S. et al., 2013, A&A, 558, A67
Bahé Y. M., McCarthy I. G., 2015, MNRAS, 447, 969
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ż., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469
Baldry I. K. et al., 2018, MNRAS, 474, 3875
Ball N. M., Brunner R. J., 2010, Int. J. Mod. Phys. D, 19, 1049
Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, ApJ, 527, 54
Bamford S. P. et al., 2009, MNRAS, 393, 1324

Barchi P. H., da Costa F. G., Sautter R., Moura T. C., Stalder D. H., Rosa R. R., de Carvalho R. R., 2016, J. Comput. Interdiscip. Sci., 7, 114
Barnes J. E., 1988, ApJ, 331, 699
Barnes J. E., 1992, ApJ, 393, 484
Barnes J. E., 2002, MNRAS, 333, 481
Baron D., 2019, Machine Learning in Astronomy: A Practical Overview
Barro G. et al., 2013, ApJ, 765, 104
Bekki K., Couch W. J., Shioya Y., 2002, ApJ, 577, 651
Bell E. F., de Jong R. S., 2001, ApJ, 550, 212
Biernacki C., Celeux G., Govaert G., 2000, IEEE Trans. Pattern Anal. Mach. Intell., 22, 719
Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014, MNRAS, 441, 599
Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103
Boselli A., Gavazzi G., 2006, PASP, 118, 517
Boselli A., Boissier S., Cortese L., Gavazzi G., 2009, Astron. Nachr., 330, 904
Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33
Bouveyron C., Brunet C., 2012, Stat. Comput., 22, 301
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bruce V. A. et al., 2012, MNRAS, 427, 1666
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cano-Díaz M., Ávila-Reese V., Sánchez S. F., Hernández-Toledo H. M., Rodríguez-Puebla A., Boquien M., Ibarra-Medel H., 2019, MNRAS, 488, 3929
Carollo C. M. et al., 2013, ApJ, 773, 112
Chabrier G., 2003, PASP, 115, 763
Charlot S., Longhetti M., 2001, MNRAS, 323, 887
Chary R., Elbaz D., 2001, ApJ, 556, 562
Cheung E. et al., 2012, ApJ, 760, 131
Conroy C., 2013, ARA&A, 51, 393
Conseil S., Amouts S., Milliard B., Zamojski M., Liebaria A., Guillaume M., 2011, in Evans I., Accomazzi A., Mink D., Rots A., eds, Astronomical Data Analysis Software and Systems XX, EMphot – Photometric Software with Bayesian Priors: Application to GALEX. San Francisco, p. 107
Cowie L. L., Songaila A., 1977, Nature, 266, 501
Croton D. J. et al., 2006, MNRAS, 365, 11
Cucciati O. et al., 2017, A&A, 602, A15
Da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595
Davidzon I. et al., 2013, A&A, 558, A23
Davidzon I. et al., 2016, A&A, 586, A23
Davidzon I. et al., 2019, MNRAS, 489, 4817
Davies L. J. M. et al., 2018, MNRAS, 480, 768
de Souza R. S. et al., 2017, MNRAS, 472, 2808
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Driver S. P. et al., 2009, Astron. Geophys., 50, 5.12
Driver S. P. et al., 2016, MNRAS, 455, 3911
Driver S. P. et al., 2018, MNRAS, 475, 2891
Elmegreen B. G., Bournaud F., Elmegreen D. M., 2008, ApJ, 688, 67
Faber S. M., 1972, A&A, 20, 361
Faber S. M. et al., 2007, ApJ, 665, 265
Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, ApJ, 776, 63
Fisher D. B., Drory N., 2008, AJ, 136, 773
Fossati M. et al., 2017, ApJ, 835, 153
Fritz A. et al., 2014, A&A, 563, A92
Gabor J. M., Davé R., Oppenheimer B. D., Finlator K., 2011, MNRAS, 417, 2676
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., 2006, MNRAS, 370, 1106
Garilli B. et al., 2014, A&A, 562, A23
Gilbank D. G., Baldry I. K., Balogh M. L., Glazebrook K., Bower R. G., 2010, MNRAS, 405, 2594
Gilbank D. G., Baldry I. K., Balogh M. L., Glazebrook K., Bower R. G., 2011a, MNRAS, 412, 2111
Gilbank D. G. et al., 2011b, MNRAS, 414, 304
Gu Y., Fang G., Yuan Q., Cai Z., Wang T., 2018, ApJ, 855, 10
Gu Y., Fang G., Yuan Q., Lu S., Li F., Cai Z.-Y., Kong X., Wang T., 2019, ApJ, 884, 172
Gunn J. E., Gott J. R., III, 1972, ApJ, 176, 1
Guzzo L. et al., 2014, A&A, 566, A108
Häring N., Rix H.-W., 2004, ApJ, 604, L89
Harris C. R. et al., 2020, Nature, 585, 357
Heckman T. M., 1980, A&A, 500, 187
Hemmati S. et al., 2019, ApJ, 881, L14
Hocking A., Geach J. E., Davey N., Sun Y., 2017, in 2017 International Joint Conference on Neural Networks. Curran Associates, Inc., New York, p. 4179
Hocking A., Geach J. E., Sun Y., Davey N., 2018, MNRAS, 473, 1108
Hopkins P. F. et al., 2009a, MNRAS, 397, 802
Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009b, ApJ, 691, 1168
Hopkins P. F. et al., 2010, ApJ, 715, 202
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ilbert O. et al., 2006, A&A, 457, 841
Ilbert O. et al., 2010, ApJ, 709, 644
Ilbert O. et al., 2013, A&A, 556, A55
Ilbert O. et al., 2015, A&A, 579, A2
Ivezić Ž. et al., 2019, ApJ, 873, 111
Jarvis M. J. et al., 2013, MNRAS, 428, 1281
Jones E., Oliphant T., Peterson P., 2001, SciPy: Open Source Scientific Tools for Python, available at http://www.scipy.org/
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Kelkar K., Gray M. E., Aragón-Salamanca A., Rudnick G., Jaffé Y. L., Jablonka P., Moustakas J., Milvang-Jensen B., 2019, MNRAS, 486, 868
Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160
Kormendy J., Kennicutt Robert C. J., 2004, ARA&A, 42, 603
Kovač K. et al., 2010, ApJ, 708, 505
Kraft R. P. et al., 2017, ApJ, 848, 27
Krywult J. et al., 2017, A&A, 598, A120
Kullback S., Leibler R. A., 1951, Ann. Math. Statist., 22, 79
Lamareille F., 2010, A&A, 509, A53
Lang D., Hogg D. W., Schlegel D. J., 2016, AJ, 151, 36
Lange R. et al., 2015, MNRAS, 447, 2603
Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692
Le Fèvre O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham, p. 1670
Lilly S. et al., 1998, ApJ, 500, 75
Lisboa P., Ellis I., Green A., Ambrogi F., Dias M., 2008, Pattern Recognit. Lett., 29, 1814
Lloyd S., 1982, IEEE Trans. Inf. Theory, 28, 129
MacQueen J., 1967, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1: Statistics. Univ. California Press, Berkeley, p. 281
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388
Manzoni G. et al., 2021, New Astron., 84, 101515
Maraston C., 2005, MNRAS, 362, 799
Marchetti A. et al., 2013, MNRAS, 428, 1424
Marinoni C. et al., 2008, A&A, 487, 7
Martig M., Bournaud F., Teyssier R., Dekel A., 2009, ApJ, 707, 250
Martin G., Kaviraj S., Hocking A., Read S. C., Geach J. E., 2020, MNRAS, 491, 1408
Martin D. C. et al., 2005, ApJ, 619, L1
Martin D. C. et al., 2007, ApJS, 173, 342
Masters K. L. et al., 2010, MNRAS, 405, 783

McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593
McPartland C., Sanders D. B., Kewley L. J., Leslie S. K., 2019, MNRAS, 482, L129
Mihos J. C., Hernquist L., 1994a, ApJ, 425, L13
Mihos J. C., Hernquist L., 1994b, ApJ, 431, L9
Mihos J. C., Hernquist L., 1996, ApJ, 464, 641
Mishra P. K., Wadadekar Y., Barway S., 2017, MNRAS, 467, 2384
Mishra P. K., Wadadekar Y., Barway S., 2018, MNRAS, 478, 351
Mishra P. K., Wadadekar Y., Barway S., 2019, MNRAS, 487, 5572
Momcheva I. G. et al., 2016, ApJS, 225, 27
Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613
Morrissey P. et al., 2007, ApJS, 173, 682
Moutard T. et al., 2016a, A&A, 590, A102
Moutard T. et al., 2016b, A&A, 590, A103
Moutard T., Sawicki M., Arnouts S., Golob A., Malavasi N., Adami C., Coupon J., Ilbert O., 2018, MNRAS, 479, 2147
Moutard T., Malavasi N., Sawicki M., Arnouts S., Tripathi S., 2020, MNRAS, 495, 4237
Muzzin A. et al., 2013, ApJ, 777, 18
Muzzin A. et al., 2014, ApJ, 796, 65
Naab T., Trujillo I., 2006, MNRAS, 369, 625
Naab T., Johansson P. H., Ostriker J. P., 2009, ApJ, 699, L178
Nipoti C., Binney J., 2007, MNRAS, 382, 1481
Noeske K. G. et al., 2007, ApJ, 660, L43
Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793
Nulsen P. E. J., 1982, MNRAS, 198, 1007
Okada T., Tomita S., 1985, Pattern Recognit., 18, 139
Oliphant T., 2006, A Guide to NumPy. Trelgol Publishing, USA
Pacifici C. et al., 2016, ApJ, 832, 79
Papovich C. et al., 2018, ApJ, 854, 30
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Peek J. E. G., Schiminovich D., 2013, ApJ, 771, 68
Peng Y.-j. et al., 2010, ApJ, 721, 193
Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192
Poggianti B. M., Smail I., Dressler A., Couch W. J., Barger A. J., Butcher H., Ellis R. S., Oemler Augustus J., 1999, ApJ, 518, 576
Puget P. et al., 2004, in Moorwood A. F. M., Iye M., eds, Proc. SPIE Conf. Ser. Vol. 5492, WIRCam: the Infrared Wide-Field Camera for the Canada-France-Hawaii Telescope. SPIE, Bellingham, p. 978
R Core Team, 2019, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria https://www.R-project.org
Renzini A., 1999, in Carollo C. M., Ferguson H. C., Wyse R. F. G., eds, The Formation of Galactic Bulges. p. 9
Roberts M. S., 1963, ARA&A, 1, 149
Salim S., 2014, Serb. Astron. J., 189, 1
Salim S., Rich R. M., 2010, ApJ, 714, L290
Salim S. et al., 2007, ApJS, 173, 267
Salim S. et al., 2016, ApJS, 227, 2
Salim S., Boquien M., Lee J. C., 2018, ApJ, 859, 11
Sánchez Almeida J., Aguerri J. A. L., Muñoz-Tuñón C., de Vicente A., 2010, ApJ, 714, 487
Schawinski K. et al., 2009, MNRAS, 396, 818
Schawinski K. et al., 2014, MNRAS, 440, 889
Schwarz G., 1978, Ann. Statist., 6, 461
Scodeggio M. et al., 2018, A&A, 609, A84
Scott D., Thompson J., 1983, Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface, 528. North-Holland, Amsterdam, p. 173
Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41
Sérsic J. L., 1968, Atlas de Galaxias Australes. Observatorio Astronomico, Cordoba
Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
Sheth K., Vogel S. N., Regan M. W., Thornley M. D., Teuben P. J., 2005, ApJ, 632, 217
Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11
Siudek M. et al., 2017, A&A, 597, A107
Siudek M. et al., 2018a, preprint (arXiv:1805.09905)
Siudek M. et al., 2018b, A&A, 617, A70
Skrutskie M. F. et al., 2006, AJ, 131, 1163
Smethurst R. J. et al., 2015, MNRAS, 450, 435
Smethurst R. J., Lintott C. J., Bamford S. P., Hart R. E., Kruk S. J., Masters K. L., Nichol R. C., Simmons B. D., 2017, MNRAS, 469, 3670
Smith R. et al., 2015, MNRAS, 454, 2502
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Spinrad H., 1962, ApJ, 135, 715
Spinrad H., Taylor B. J., 1971, ApJS, 22, 445
Springel V., Di Matteo T., Hernquist L., 2005a, MNRAS, 361, 776
Springel V., Di Matteo T., Hernquist L., 2005b, ApJ, 620, L79
Steinhardt C. L., Weaver J. R., Maxfield J., Davidzon I., Faisst A. L., Masters D., Schemel M., Toft S., 2020, ApJ, 891, 136
Strateva I. et al., 2001, AJ, 122, 1861
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29
Tonini C., Mutch S. J., Croton D. J., Wyithe J. S. B., 2016, MNRAS, 459, 4109
Toomre A., 1977, in Tinsley B. M., Larson Richard B., Gehret D. C., eds, Evolution of Galaxies and Stellar Populations. Yale University Observatory, New Haven, p. 401
Turner S. et al., 2019, MNRAS, 482, 126
Valiante E. et al., 2016, MNRAS, 462, 3146
van der Wel A. et al., 2014, ApJ, 788, 28
van Dokkum P. G. et al., 2015, ApJ, 813, 23
Vergani D. et al., 2018, A&A, 620, A193
Walker I. R., Mihos J. C., Hernquist L., 1996, ApJ, 460, 121
Weigel A. K. et al., 2017, ApJ, 845, 145
Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336
Wild V., Walcher C. J., Johansson P. H., Tresse L., Charlot S., Pollo A., Le Fèvre O., de Ravel L., 2009, MNRAS, 395, 144
Willett K. W. et al., 2013, MNRAS, 435, 2835
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879
Woo J., Carollo C. M., Faber S. M., Dekel A., Tacchella S., 2017, MNRAS, 464, 1077
Wright E. L. et al., 2010, AJ, 140, 1868
Wuyts S. et al., 2007, ApJ, 655, 51
Wyder T. K. et al., 2007, ApJS, 173, 293
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yang Y., Zabludoff A. I., Zaritsky D., Mihos J. C., 2008, ApJ, 688, 945
Yip C. W. et al., 2004, AJ, 128, 585
York D. G. et al., 2000, AJ, 120, 1579
Yuan H. B., Liu X. W., Xiang M. S., 2013, MNRAS, 430, 2188
Zabludoff A. I., Zaritsky D., Lin H., Tucker D., Hashimoto Y., Shectman S. A., Oemler A., Kirshner R. P., 1996, ApJ, 466, 104
Zinger E., Dekel A., Kravtsov A. V., Nagai D., 2018, MNRAS, 475, 3654

APPENDIX A: ITERATIONS OF FEM

In Fig. A1, we show ICL scores reported at each of up to 25 iterations by various combinations of submodel and k for our GSWLC-2 sample. These 'iteration profiles' are mostly quite flat; hence, 25 iterations are more than sufficient for allowing FEM to stabilize to an outcome. In addition, the bulk of the clustering structure

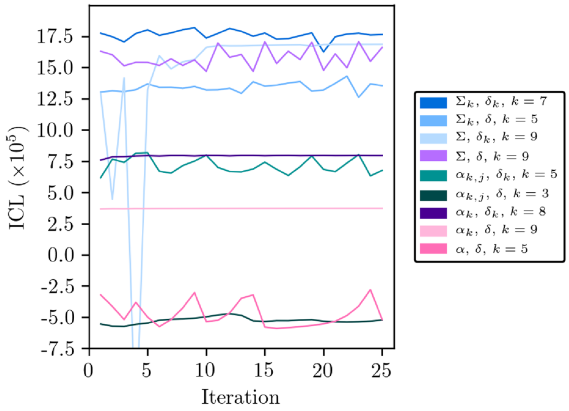

Figure A1. ICL scores reported at iterations 1 through 25 by various combinations of submodel and k for our GSWLC-2 sample. For each submodel, we show the value of k which yields the highest ICL score. These iteration profiles are generally quite flat, indicating that FEM quickly converges to a stable outcome. The large changes exhibited by Σ, δ_k, $k = 9$ are due to the emptying of clusters as it iterates.

appears to be determined during the k-means initialization step, which spreads the cluster centres out ahead of the first iteration. The ICL criterion rewards separated clusters, so k-means initializations are particularly well suited to yielding useful clustering outcomes. Trials of the use of uniform random initializations resulted in more combinations of submodels and k failing to converge.

Variations in the ICL values reported by individual combinations of submodel and k over successive iterations arise due to the Fisher step of FEM, in which the subspace within which the clusters are to be modelled is found. Hence, the updating of the model parameters during the Maximization step is *indirectly* related to the probabilities calculated in the Expectation step. For traditional EM algorithms, these steps are directly related and thereby guarantee convergence. The large changes between successive iterations exhibited by some combinations (e.g. Σ, δ_k, $k = 9$) are most often due to the emptying of clusters; a reduction in the number of clusters used by FEM leads, in these cases, to a sudden increase in ICL.

APPENDIX B: SMOOTHING OF FEATURE DATA FOR OUR GSWLC-2 SAMPLE

Preliminary tests revealed that a truncated, bimodal substructure among passive galaxies within the nine-dimensional colour space representing our GSWLC-2 sample (see the left-hand plot of Fig. B1; also visible in Fig. 3) led to an inability of FEM to converge for the majority of submodels and values of k. This truncated bimodal substructure is due to the lack of input NIR photometry to the CIGALE SED estimation of GSWLC-2 galaxies, such that their NIR SEDs must be inferred from UV and optical photometry. This, in turn, leads to poorly constrained, discretized metallicities: galaxies at $r - K_s \lesssim 0.67$ peak strongly at $\log_{10}(Z) \sim -2.4$, and those at $r - K_s \gtrsim 0.67$ at $\log_{10}(Z) \sim -2.1$. The NIR SEDs of VIPERS galaxies, on the other hand, are constrained by K_s-band photometry and hence have slightly more freedom to vary. This smooths their colour and metallicity distributions.

We hence opt to apply a small level of Gaussian smoothing to the GSWLC-2 distributions of the rest-frame absolute magnitudes reported by CIGALE. The smoothing scale for the rest-frame absolute magnitude of a given galaxy is given by its Bayesian error. These errors are winsorized at the mean value of the logarithmic distribution

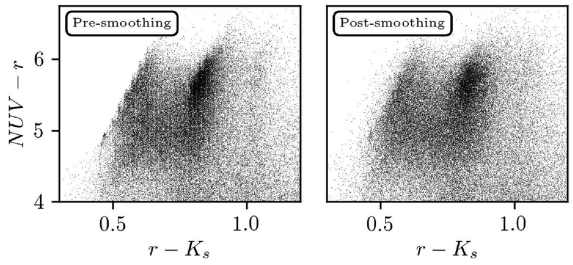

Figure B1. The effect of our smoothing on the distribution of GSWLC-2 galaxies in the passive region of the $NUV - r - K_s$ colour–colour plane. Substructures in the distribution of galaxies within this region are preserved post-smoothing.

of errors (i.e. errors larger than the mean value are set to the mean value). This winsorization ensures that the smoothing scale is kept small enough to avoid the potential loss of astrophysically meaningful substructures, while still enabling FEM to converge more readily. The absolute rest-frame magnitude most affected by this smoothing is *FUV*, whose errors are winsorized at a maximum value of 0.25 (all other magnitudes have a maximum error <0.1 after winsorization). The right-hand plot of Fig. B1 demonstrates the effect of our smoothing, showing that the bimodality in the colours of passive galaxies is retained post-smoothing. While this bimodality is likely to be an artefact, trends in the astrophysical features of galaxies between its peaks are still likely to be genuine (see also Section 4.4.2).

APPENDIX C: BEHAVIOUR OF THE VARIOUS SUBMODELS OF FEM FOR OUR SAMPLES

Our model selection approach considers ICL scores for 72 different combinations of submodel and k for *each* of our samples. The comparison of these 72 combinations is simplified greatly by the realization that several submodels exhibit consistent patterns of behaviour across all values of k.

FEM is unable to converge to an outcome for several combinations of submodel and k. The most common diagnosis made by FEM in the case of non-convergence is that a cluster has become empty (i.e. that it no longer contains galaxies). Table 1 shows that several submodels are unable to converge beyond a maximum value of k, suggesting a limit to their ability to properly partition the samples. Alternatively, submodels that converge at k, but fail to converge at $k - 1$ and $k + 1$ appear to be striking a 'sweet spot' in terms of this ability. Different combinations are generally very consistent with respect to convergence, converging for either all or none of our 100 initializations.

Given their flexibility and their high levels of parametrization, the Σ_k, δ_k and Σ_k, δ submodels offer the greatest promise among all of the FEM submodels for yielding detailed and astrophysically meaningful partitions of our samples. The outcomes they produce are similar; they exhibit near-identical trends in their ICL scores for $k = 2$ through $k = 5$ for our GSWLC-2 sample in Table 1. They differ only in their treatment of the noise terms, which appears to be a minor detail in comparison with their shared use of full, unique covariance matrices. Outcomes at higher values of k generally consist of splits of clusters present in outcomes at lower values of k.

Submodels featuring non-unique covariance matrices for the Gaussian density functions representing the clusters (i.e. submodels with Σ and α, such that they all have the same shape) consistently produce clusters with highly disparate sizes. Some clusters are large,

containing 30 to 60 per cent of the galaxies in our samples each (and each often spanning both blue and red galaxies); others are empty or nearly empty, containing $\lesssim$1 per cent of the galaxies in our samples each. Nearly empty clusters appear to capture small, undesirable artefacts in the structure of our samples within their input feature spaces. While it is unclear why FEM registers a valid ICL score for these outcomes when they include empty clusters (often cited as a cause for the failure of FEM; see above), it is clear that these submodels are too crude to return more than a very broad partition of our samples, and that their outcomes are limited in their capacity for astrophysical interpretation. All of this is also true for the Σ, δ clustering outcome at $k = 9$ for our GSWLC-2 sample, which achieved the highest ICL score in our model selection search despite including empty and nearly empty clusters. For these reasons, we reject this outcome for analysis.

A general property of clustering outcomes reported by submodels which assume diagonal covariance matrices ($\alpha_{k,j}$, α_k) for the Gaussian density functions within the discriminative latent subspace is that they segment our samples principally along a single dimension. Several representative examples of their clustering structures are shown in Fig. C1, revealing that this single dimension is most strongly associated with the UV colours among our nine input features, with little-to-no distinction made between galaxies based on their NIR colours. We note that these submodels scored highest when we tested clustering of our samples using *i*-band magnitudes of galaxies as a reference point for defining colours (as in Siudek et al. 2018b; see also Section 3.4), producing the same striping pattern within the $NUV - r - K_s$ plane. While this simple segmentation does correspond broadly with incremental changes in the star formation activity of galaxies within our samples, other submodels (with Σ_k) return more detailed partitions and achieve higher ICL scores anyway.

The large spread in the ICL scores reported in Table 1 arises directly from a large spread in the log-likelihood values of the fits. This large spread in the log-likelihood values arises, in turn, primarily from a $1/\delta_k$ coefficient in the log-likelihood function of DLM model (which may be seen in full in appendix 2 of Bouveyron & Brunet 2012). Submodels which yield very large but negative log-likelihood (and hence, ICL) values tend to have very small δ_k values for most (if not all) of their clusters; usually 0.001, which is the floor that FEM imposes upon the value of δ_k. Very small values of δ_k produce very large, positive values of $1/\delta_k$, and (via a $-1/2$ coefficient of the log-likelihood function) very large, negative values of the log-likelihood and, thus, of the ICL criterion. The addition of this especially low-variance noise to subspace Gaussians leads to highly peaked full space Gaussians which are unlikely to reflect the more continuous distributions of both samples (see Fig. 3).

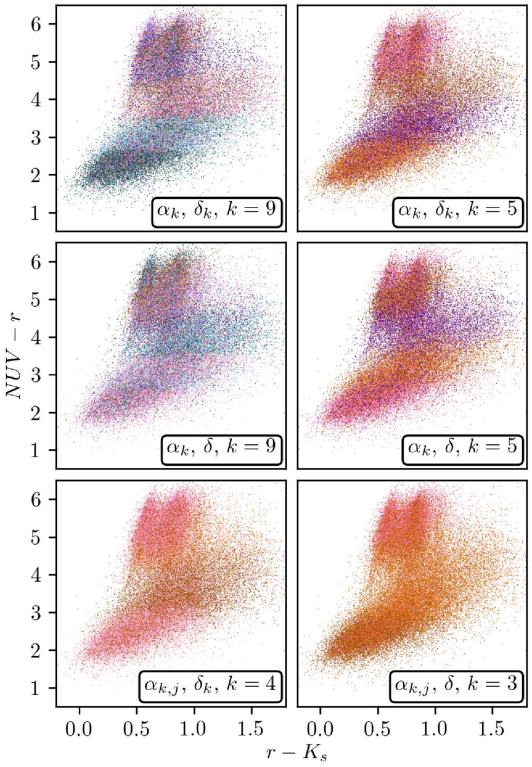

Figure C1. Examples of the clustering structures determined by $\alpha_{k,j}$ and α_k submodels for our GSWLC-2 sample, shown in the $NUV - r - K_s$ colour–colour plane. The combination of submodel and k for each outcome is shown to the lower right of each plot. Individual galaxies are coloured in accordance with the cluster to which they belong. The choice of colours in this figure is not intended to imply any trends within or between plots. The horizontal striping pattern exhibited by these examples in these plots, which is a general property of $\alpha_{k,j}$- and α_k-based outcomes, indicates segmentation mainly along a single axis.

This paper has been typeset from a TEX/LATEX file prepared by the author.