

Reconsidering the dynamical states of galaxy clusters using PCA and UMAP

Roan Haggar^{1,2*}, Federico De Luca^{3,4}, Marco De Petris⁵, Elizaveta Sazonova^{1,2}, James E. Taylor^{1,2}, Alexander Knebe^{6,7,8}, Meghan E. Gray⁹, Frazer R. Pearce⁹, Ana Contreras-Santos⁶, Weiguang Cui^{6,7,10}, Ulrike Kuchner⁹, Robert A. Mostoghiu Paun^{11,12} and Chris Power^{8,13}

¹Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

²Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada

³Dipartimento di Fisica, Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Roma, Italy

⁴INFN, Sezione di Roma 2, Università degli studi di Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Roma, Italy

⁵Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy

⁶Departamento de Física Teórica, Módulo 15 Universidad Autónoma de Madrid, E-28049 Madrid, Spain

⁷Centro de Investigación Avanzada en Física Fundamental (CIAFF), Universidad Autónoma de Madrid, E-28049 Madrid, Spain

⁸International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

⁹School and Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD, UK

¹⁰Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

¹¹Centre for Astrophysics & Supercomputing, Swinburne University of Technology, 1 Alfred St, Hawthorn, VIC 3122, Australia

¹²ARC Centre of Excellence for Dark Matter Particle Physics (CDM), Australia

¹³ARC Centre of Excellence for All-Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

Accepted 2024 June 21. Received 2024 June 20; in original form 2023 September 30

ABSTRACT

Numerous metrics exist to quantify the dynamical state of galaxy clusters, both observationally and within simulations. Many of these correlate strongly with one another, but it is not clear whether all of these measures probe the same intrinsic properties. In this work, we use two different statistical approaches – principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) – to investigate which dynamical properties of a cluster are in fact the best descriptors of its dynamical state. We use measurements taken directly from THE THREE HUNDRED suite of galaxy cluster simulations, as well as morphological properties calculated using mock X-ray and SZ maps of the same simulated clusters. We find that four descriptions of dynamical state naturally arise, and although correlations exist between these, a given cluster can be ‘dynamically relaxed’ according to all, none, or some of these four descriptions. These results demonstrate that it is highly important for future observational and theoretical studies to consider in which sense clusters are dynamically relaxed. Cluster dynamical states are complex and multidimensional, and so it is not meaningful to classify them simply as ‘relaxed’ and ‘unrelaxed’ based on a single linear scale.

Key words: methods: numerical – galaxies: clusters: general – galaxies: kinematics and dynamics – dark matter – X-rays: galaxies: clusters.

1 INTRODUCTION

The structure of a galaxy cluster consists of a large dark matter halo, typically with a mass in the range 10^{14} – 10^{15} M_{\odot} , filled with hot intracluster gas, and populated with hundreds or thousands of galaxies. A wide range of physical processes take place within clusters, meaning they play a crucial role in many areas of astrophysics and cosmology.

The evolution of galaxies is strongly dependent on their cosmic environment, and clusters represent one of the most extreme environments for a galaxy. Strong tidal forces due to a cluster’s gravitational potential can disrupt the structure of a galaxy, dramatically changing

its morphology (Moore et al. 1996; Mihos et al. 2017; Knebe et al. 2020; Haggar et al. 2023). Additionally, processes such as ram-pressure stripping (Gunn & Gott 1972) by the intracluster gas can remove both the cold gas in a galactic disc and the surrounding halo gas, quenching star formation in galaxies (Larson, Tinsley & Caldwell 1980; Zabel et al. 2019). Galaxy clusters are also an important tool in constraining cosmology – for example, studying the shapes of cluster halo density profiles can provide information about the nature of dark matter (Eckert et al. 2022; Limousin, Beauchesne & Jullo 2022). Galaxy clusters can also be used as a proxy for measuring cosmological parameters, either through cluster number counts from large surveys (Evrard 1989; de Haan et al. 2016; Abdullah, Klypin & Wilson 2020, for example), or using cluster properties. For example, Amoura et al. (2021) measure the formation times of clusters in a suite of simulations with varying values of Ω_M , the matter density

* E-mail: rhaggar@uwaterloo.ca

parameter, and σ_8 , the amplitude of the matter power spectrum at a scale of $8 h^{-1}\text{Mpc}$, and show that the formation times of galaxy clusters depend on the values of Ω_M and σ_8 .

One of the core properties of a galaxy cluster is its dynamical state – that is, whether or not it is dynamically relaxed. Numerous areas of galaxy cluster physics, including many of those described above, are dependent on an understanding of cluster dynamical states. It has been shown that various phenomena in the intracluster medium (ICM), such as turbulence, differ between relaxed and unrelaxed clusters (Lau, Kravtsov & Nagai 2009; Vallés-Pérez, Planelles & Quilis 2021; Simonte et al. 2022). Furthermore, Nagai & Lau (2011) and Vazza et al. (2013) both show that, in dynamically disturbed clusters, the hot X-ray-emitting ICM contains dense, cooler clumps of gas, associated with infalling galaxies and galaxy groups. The ICM drives galaxy evolution mechanisms such as ram-pressure stripping; this has also been shown to be enhanced in dynamically disturbed groups and clusters (Mauduit & Mamon 2007; Roberts & Parker 2017; Lourenço et al. 2023).

Additionally, many of the astronomical and cosmological measurements described above are indirectly related to the cluster dynamical states, as they rely on accurate measurements of the masses of clusters, which are in turn dependent on their dynamical state. Cluster masses are often calculated under the assumption that a cluster is in hydrostatic equilibrium. As a result of this, clusters that are dynamically unrelaxed (and so are not in hydrostatic equilibrium) can have their masses underestimated by up to 20 per cent (Nagai, Vikhlinin & Kravtsov 2007; Kravtsov & Borgani 2012). Relaxed clusters can also have their masses underestimated, albeit to a lesser extent (Lau, Nagai & Nelson 2013; Gianfagna et al. 2021), and the strength of this bias is dependent on the redshift of the cluster in question (Bennett & Sijacki 2022). Cluster scaling relations also differ between clusters that are relaxed, and those that are rapidly accreting material (Planelles & Quilis 2009; Lau et al. 2015; Chen et al. 2019). Additionally, properties such as location of a cluster’s splashback radius (Adhikari, Dalal & Chamberlain 2014; More, Diemer & Kravtsov 2015) are dependent on its dynamical state. This is the radius within which the cluster material dominates over the surrounding infalling material, and so this implies that the region in which a galaxy is impacted by a cluster is also dependent on cluster dynamical state. Because of this, a thorough understanding of cluster dynamical states is vital if we are to use clusters as an astronomical and cosmological tool.

In its simplest form, a system of collisionless particles can be described as dynamically relaxed and virialized once the velocities of particles in the system are uncorrelated with their initial velocities. Equivalently, this means the average magnitude of the velocity of each particle, v , is equal to the change in velocity of the particle due to interactions with other particles, δv . The typical time required for a system to reach this stage is given by the relaxation time, t_{relax} ,

$$t_{\text{relax}} \approx \frac{0.1N}{\ln(N)} t_{\text{cross}}, \quad (1)$$

where N is the number of particles in a system, and t_{cross} is the average crossing time for a particle in the system (Binney & Tremaine 1987). For a galaxy cluster, the number of particles (galaxies) is $\sim 10^3$, and the crossing time is ~ 1 Gyr, leading to a typical relaxation time of ~ 10 Gyr, comparable to the age of the Universe. However, in the context of galaxy clusters, this definition of dynamical state is not particularly useful for several reasons. First, clusters are not collisionless systems. Galaxies in clusters frequently experience near-misses or tidal interactions (Knebe et al. 2006; Muldrew, Pearce & Power 2011; Bahé et al. 2019), and the intracluster gas, which

makes up a significant portion of a cluster’s mass, is not a collisionless fluid. Secondly, this definition describes a closed system, which galaxy clusters are not; clusters are continuously accreting material from their surrounding environment. As such, the effects from their ‘boundary’ need to be included to quantify their dynamical state. Finally, according to this definition, only material that has been in a cluster for greater than 10 Gyr can be dynamically relaxed. This means that a $z = 0$ cluster can only be truly relaxed if it accreted all, or most, of its material before $z = 1.5$. While this is technically possible, it is an exceptionally rare scenario in a typical Lambda cold dark matter cosmology.

Consequently, throughout the literature, numerous properties of clusters are treated as indicators of ‘dynamical state’, each of which is used to quantify how relaxed is a galaxy cluster. Many of these are theory-based metrics, taken from simulations of galaxy groups or clusters. In Cui et al. (2017), the dynamical state of a cluster is described by three properties: dynamically unrelaxed clusters are those with a centre of mass that is offset from the cluster halo density peak, those with large amounts of substructure, and those that are not virialized.

This combination of observables has been widely used for some time. For example, Neto et al. (2007) classify dark matter haloes based on these three parameters, and place a limit on each of these, describing relaxed haloes as those that satisfy all three of these conditions, and unrelaxed haloes as those that do not. Cui et al. (2018) place similar constraints on simulated galaxy clusters, and use this to classify clusters as relaxed and unrelaxed. Haggard et al. (2020), by contrast, combine the three into a single parameter, χ_{DS} , giving a single continuous parameter describing how relaxed is a cluster. Other work also uses these measures, such as Wen & Han (2013), who use the amount of substructure as an indicator of dynamical state. They describe how a large amount of structure is indicative of a recent merger event. The results of Kimmig et al. (2023) also demonstrate this – they show that the amount of substructure, and in particular the size of the eighth most massive substructure, is indicative of the merger history of a cluster over the last 2 Gyr. Further works have shown that numerous other cluster properties correlate with recent merger activity, such as the virialization and centre-of-mass offset of a cluster (Power, Knebe & Knollmann 2012), its concentration (Wang et al. 2020), or similarly its sparsity (Richardson & Corasaniti 2022). Contreras-Santos et al. (2022) also use the time since a major merger as an measure of cluster relaxation, and show how this correlates with the χ_{DS} measure from Haggard et al. (2020).

Already, it is apparent that different measures of a cluster’s dynamical state are probing different core properties. Measurements such as the substructure fraction and centre-of-mass offset are quantities that can be measured observationally at a single epoch, although in observations they are projected into two dimensions. This is in contrast to the time since the last major merger, which is a property of the history of a cluster. Other studies also take this approach to dynamical state, connecting it to the total history of a cluster. For example, Diemer & Kravtsov (2014) introduced the accretion rate proxy Γ , variants of which have since become widely used as measures of halo relaxation (e.g. Vallés-Pérez, Planelles & Quilis 2020). An alternative approach is to use a redshift-dependent definition of dynamical state, based on the fact that different cluster properties evolve over different time-scales (e.g. Mendoza et al. 2023; Vallés-Pérez et al. 2023).

Gouin, Bonnaire & Aghanim (2021) probe the dynamical state of clusters from the IllustrisTNG simulation (Nelson et al. 2019) using three separate measures of cluster growth history: the $z = 0$ halo growth rate, the $z = 0.5$ mass accretion rate, and the cluster formation

time ($z_{0.5}$, see Section 2.2.1). They also show that dynamically disturbed clusters are more strongly connected to the cosmic web – that is, they have more cosmic filaments attached to the cluster (see Santoni et al. 2024, for a similar study with THE300 clusters). Similarly, Darragh Ford et al. (2019) find that simulated clusters that have recently experienced a major merger have a higher connectivity. This is yet another interpretation of dynamical state, as a property of the surrounding region of the Universe. This cosmic environment can also impact global properties of the cluster and its dark matter halo, such as its shape (Gouin et al. 2021; Smith et al. 2023) and concentration (Neto et al. 2007).

Additional complexity comes from the large number of cluster dynamical state metrics and morphological parameters that are used in observational astronomy. For example, cluster shapes can be mapped using X-ray and Sunyaev–Zel’dovich (SZ) effect observations. From these, many quantifiable properties can be measured, such as the amount of substructure (Ge et al. 2016), offset of the brightest cluster galaxy (BCG) from the X-ray/SZ peak (Zenteno et al. 2020), or the power spectrum of the hot gas distribution (Cerini, Cappelluti & Natarajan 2023). Combinations of multiple X-ray or SZ morphological parameters have been shown to provide even more robust measures of dynamical state than these individual parameters (Parekh et al. 2015; Yuan & Han 2020; Campitiello et al. 2022; Yuan, Han & Wen 2022; see Rasia, Meneghetti & Ettori 2013 and Zhang et al. 2024 for more comprehensive discussions). Other observational studies include Wen & Han (2013), who quantify cluster substructure using optical data from the Sloan Digital Sky Survey (Aihara et al. 2011).

The result of having so many different properties in common use is that the ‘dynamical state’ of a cluster is not clearly or consistently defined in the literature. Consequently, when the dynamical states of clusters are inferred from observations, it is not entirely apparent which fundamental property of a cluster these are probing. It is also not clear if they are all probing the same intrinsic cluster property, or if the ‘dynamical state’ of a cluster is actually made up of several properties.

In this study, we aim to gain a deeper understanding of cluster dynamical state measurements by investigating the connections between theory-based and observable properties of dynamical state, the degeneracy between different measures of dynamical state, and the core, intrinsic properties of clusters that these measurements are actually probing. Our primary approach to this is through dimensionality reduction – reducing a large number of dynamical state metrics to a smaller set of variables will make the nature of dynamical state clearer and easier to interpret. Previous studies have taken a similar approach to ours, although mostly to consider the accretion histories of dark matter haloes, rather than the dynamical states of galaxy clusters. One notable example is Wong & Taylor (2012), who use principal component analysis (PCA; Jolliffe & Cadima 2016) to determine the principal components of 10 input properties of dark matter haloes, and study the mass accretion histories of these haloes. They show that splitting these clusters by their first and second principal components naturally displays two separate modes of accretion history: the halo formation time, and the acceleration/deceleration of a halo’s accretion. For further discussion, we also refer the reader to Jeesson-Daniel et al. (2011) and Skibba & Macciò (2011), who perform similar analyses on simulations of dark matter haloes.

In this work we develop the methods used in previous studies, applying PCA to 17 indicators of cluster dynamical state, based on 3D data (as opposed to directly observable quantities) from simulations of $z = 0$ galaxy clusters. This analysis uses THE THREE HUNDRED

project, a suite of hydrodynamical zoom simulations of massive galaxy clusters, taken from a large $1 h^{-1}\text{Gpc}$ cosmological volume. We physically interpret the principal components that naturally arise from this analysis, and show how they correspond to the mass accretion histories of these clusters. Despite this work focusing on a theoretical approach to dynamical states, we also compare our PCA to X-ray and SZ properties measured from mock observations of the same simulated clusters, originally calculated in De Luca et al. (2021). Finally, we expand further on previous studies in this area by analysing the same simulated clusters using uniform manifold approximation and projection (UMAP; McInnes, Healy & Melville 2018), an alternative approach to dimensionality reduction. While we do not develop a quantitative means to classify clusters, we demonstrate qualitatively the different dynamical states of clusters that exist in our simulations.

We interpret our results as showing that the dynamical state of a cluster is not a single property. Instead, a single cluster has multiple ‘dynamical states’, and can be relaxed in all of these dimensions, or none of them, or *some* of them.

The paper is structured as follows: in Section 2 we introduce the simulation data used throughout this work, as well as the dynamical state indicators we use directly from the simulations (Section 2.2.1) and from mock X-ray and SZ-effect maps generated from the simulation data (Section 2.2.2). In Section 3 we study these parameters using both PCA and UMAP, and discuss how the results connect to the mock observations of clusters, and their mass accretion histories. Finally, in Section 4 we summarize our findings, and the implications this has for how the dynamical state of galaxy clusters should be interpreted.

2 SIMULATIONS AND METHODOLOGY

This work utilizes simulation data from THE THREE HUNDRED project. We study these galaxy clusters in the final snapshot of the simulations ($z = 0$). In Section 2.1 we provide an overview of the simulation data, but for a detailed description we refer the reader to Cui et al. (2018).

2.1 Simulation data

THE THREE HUNDRED project (hereafter THE300) is a suite of hydrodynamical resimulations of large galaxy clusters. The simulations are based on the MDPL2 MultiDark simulation (Klypin et al. 2016).¹ MDPL2 is a large dark matter-only simulation, with a comoving box size of $1 h^{-1}\text{Gpc}$, which uses *Planck* cosmology ($\Omega_M = 0.307$, $\Omega_B = 0.048$, $\Omega_\Lambda = 0.693$, $h = 0.678$, $\sigma_8 = 0.823$, $n_s = 0.96$) (Planck Collaboration XIII 2016).²

From this simulation, the 324 most massive dark matter haloes were selected, and resimulated from their initial conditions (at an initial redshift of $z_{\text{init}} = 120$) with baryonic physics included. For each cluster, the $1 h^{-1}\text{Gpc}$ dark matter-only simulation box was re-centred on the cluster, and the particles within $15 h^{-1}\text{Mpc}$ of the cluster centre at $z = 0$ were traced back to their initial conditions. Each of these particles was then split into a dark matter particle and a gas particle, with masses of $m_{\text{DM}} = 1.27 \times 10^9 h^{-1}M_\odot$ and $m_{\text{gas}} = 2.36 \times 10^8 h^{-1}M_\odot$, respectively, set according to the cosmic

¹The MultiDark simulations are publicly available from the CosmoSim data base, <https://www.cosmosim.org>.

²The reduced Hubble constant, h , is defined such that the Hubble constant, $H_0 = h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

baryon fraction. The resolution of dark matter particles outside of this radius was degraded, allowing the large-scale tidal forces acting on the cluster to be maintained with a reduced computational cost.

In this work, we use clusters from THE300 data set simulated using the GADGETX code. GADGETX is an updated version of the GADGET3 code (Springel 2005; Beck et al. 2016; Sembolini et al. 2016), and uses a smoothed-particle hydrodynamics scheme to simultaneously evolve the baryonic and dark matter components of a simulation. In addition to gas, the stochastic star formation in GADGETX produces stellar particles of varying mass, typically around $m_{\text{star}} \sim 4 \times 10^7 h^{-1} M_{\odot}$ (Tornatore et al. 2007; Murante et al. 2010; Rasia et al. 2015). Type II supernova feedback and active galactic nucleus feedback are included, as described in Springel & Hernquist (2003) and Steinborn et al. (2015), respectively. The final data set consists of 324 galaxy clusters, with masses ranging from $M_{200} = 5 \times 10^{14} h^{-1} M_{\odot}$ to $M_{200} = 2.6 \times 10^{15} h^{-1} M_{\odot}$. Here, M_{200} is the mass enclosed within a sphere of radius R_{200} , defined such that the average density within this sphere is equal to 200 times the critical density of the Universe at that redshift. For the cluster masses in THE300, the $15 h^{-1} \text{Mpc}$ high-resolution region corresponds approximately to $7\text{--}10 R_{200}$. In addition to the extensive description in Cui et al. (2018), further description of THE300 data set is available in other previous studies that have used these data. We particularly refer the reader to studies that have investigated cluster dynamical states using THE300 (Haggar et al. 2020; Capalbo et al. 2021; De Luca et al. 2021; Contreras-Santos et al. 2022; Li et al. 2022, for example).

In this work, we specifically use the data from the final snapshot of THE300 simulations ($z = 0$), as the focus of this work is on present-day galaxy clusters. However, some of the cluster properties we calculate also rely on cluster properties at $z > 0$ (see Section 2.2.1 for details). This information requires the construction of halo merger trees, which are described in the following section.

2.1.1 Galaxy identification and tree-building

In each snapshot of THE300 data, the haloes and subhaloes are identified using the Amiga Halo Finder, AHF³ (see Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009, for a detailed description of AHF). AHF is a density peak halo finder, and is used to identify the particles in the main cluster halo, subhaloes of the cluster, and haloes in the surrounding region. For each of these (sub)haloes, properties such as the position, velocity, and mass are given as outputs. The halo catalogues in each snapshot were linked together using the MERGERTREE tree-builder, which is part of the AHF package. This tree-builder uses a merit function, M_1 , given in table B1 of Knebe et al. (2013), to identify the main progenitor of each halo by searching for particles that are common between the two haloes. MERGERTREE is also able to ‘skip’ snapshots, meaning that if AHF is unable to resolve a halo in one snapshot, the tree-builder can instead find an appropriate progenitor in an earlier snapshot (see Knebe et al. 2011; Srisawat et al. 2013, for additional details on MERGERTREE).

For most of this work, we include all of these subhaloes (and thus all of the particles) in the calculations of dynamical state parameters. For those measures that rely on galaxy properties (e.g. cluster richness, N_{200} , and magnitude difference between galaxies, m_{12} , detailed in Section 2.2.1), we include all galaxies with a total mass greater than $10^{10.5} h^{-1} M_{\odot}$, and a stellar mass of greater than $10^{9.5} h^{-1} M_{\odot}$. These limits have an extremely minor impact on our

results, as most dynamical state indicators based on galaxy properties are dependent on the largest, brightest galaxies in a cluster.

2.2 Measures of dynamical state

2.2.1 3D measures

From the simulations, we utilize 17 different properties, each of which is associated with cluster dynamical state. These are detailed below, along with relevant information on how they were calculated in our simulations. Many of these quantities are calculated using R_{200} as an outer boundary, as this is the characteristic radius used by AHF in calculating halo properties; we state explicitly when this is not the case. These properties are also calculated using all particles within the relevant radius (typically R_{200} unless stated otherwise), apart from c_{200} which only uses dark matter particles.

(i) f_s : the fraction of cluster mass inside a given radius that is contained within subhaloes. Two values of this are used, $f_s(R_{200})$ and $f_s(R_{500})$, equal to the substructure mass fraction inside the radii R_{200} and R_{500} , respectively. Using these two radii allows us to probe both the outer and inner regions of a cluster: as Cui et al. (2017) show, dynamical state indicators are dependent on cluster-centric distance.

(ii) Δ_r : the offset of the centre of mass of the cluster from the density peak of the cluster halo, as a fraction of the cluster radius R_{200} . Similarly to f_s , $\Delta_r(R_{200})$ and $\Delta_r(R_{500})$ are both used, which are each calculated using the centre of mass of all material inside the radii R_{200} and R_{500} , respectively.

(iii) η : the virial ratio, a measure of how well a cluster obeys the virial theorem, based on its total kinetic energy, T , its total potential energy, W , and its energy from surface pressure, E_s . It is typically defined in the literature as $\tilde{\eta} = (2T - E_s)/|W|$, so that $\tilde{\eta} = 1$ for virialized haloes. However, PCA is designed to capture linear, monotonic relationships between variables. This variable, where the ‘extreme’ cases (most virialized haloes) correspond to an intermediate value ($\tilde{\eta} = 1$) is therefore not well suited to PCA. Consequently, we perform a transformation, defining a new η such that

$$\eta = \left| \frac{(2T - E_s)}{|W|} - 1 \right|. \quad (2)$$

Increasing values of this quantity correspond to a greater deviation from virialization, and thus to less virialized haloes, making this quantity better suited for use in PCA. Note also that this virial ratio differs from the classic definition of virialization, due to the additional surface pressure term which accounts for clusters’ ongoing accretion of material (Poole et al. 2006; Shaw et al. 2006). The surface pressure is calculated as the energy from surface pressure integrated over the halo boundary – a detailed mathematical description is given in section 3 of Cui et al. (2017). The virial ratio is also calculated twice, for all material inside R_{200} and R_{500} .

(iv) $z_{0.5}$: the formation time of a cluster: the redshift at which the cluster mass, M_{200} , is equal to half its value at $z = 0$.

(v) λ : a dimensionless spin parameter, used to describe the bulk rotation of a cluster. It is defined in the same way as Bullock et al. (2001),

$$\lambda = \frac{J_{200}}{\sqrt{GM_{200}^3 R_{200}}}, \quad (3)$$

where J_{200} is the total angular momentum of material inside R_{200} . We note that other definitions of the cluster spin also exist (e.g. Peebles

³<http://popia.ft.uam.es/AHF>

1969), but we use this value as it is calculated using only mass within a well-defined radius, R_{200} .

(vi) c/a : sphericity, the ratio of the minor and major axes of the cluster halo's moment of inertia tensor. As this calculation of the moment of inertia includes all particles (dark matter, gas, and stars), it accounts for these particles' varying masses accordingly. We note that several alternative measures of cluster shape are also used throughout the literature, including triaxiality, ellipticity, and prolativity (Lau et al. 2021), but we only use sphericity in this work.

(vii) c_{200} : the concentration of the dark matter halo. This is equal to the ratio between R_{200} and R_s , the scale radius of a halo, as defined by an NFW profile (Navarro, Frenk & White 1996). Here, $\rho(r)$ is the dark matter density of a halo as a function of halo-centric distance, and ρ_0 is some characteristic density:

$$\rho(r) = \frac{\rho_0}{\left(\frac{r}{R_s}\right) \left(1 + \frac{r}{R_s}\right)^2}. \quad (4)$$

This is an output from AHF, which selects cluster centres based on a density peak finder. The concentration of a halo is not calculated directly, but is instead a numerical solution to equation (9) in Prada et al. (2012), based on the maximum circular velocity of a cluster.

(viii) $z_{\text{merge},50}$: the redshift at which a cluster last experienced a merger that increased its mass by more than 50 per cent. These are equal to the values of z_{start} calculated in Contreras-Santos et al. (2022), which describe the jump in the mass accretion history of the cluster and therefore the onset of a merger phase.

(ix) γ : the mass accretion rate of the cluster, defined as the fractional increase in M_{200} within the last dynamical time, t_d . We use the crossing time of a cluster as its dynamical time (see Binney & Tremaine 1987; Contreras-Santos et al. 2022, for example), equal to

$$t_d \simeq \frac{R_{200}}{v_{\text{circ}}} = \sqrt{\frac{R_{200}^3}{GM_{200}}}. \quad (5)$$

Using the definition of M_{200} , and the fact that the critical density $\rho_{\text{crit}} = 3H^2/8\pi G$, we find that the dynamical time at $z = 0$ is given by $t_d \simeq 1/(10H_0)$. For the cosmology used in THE300 simulations, this dynamical time is equal to approximately 1.4 Gyr, or a redshift of $z = 0.1$. Hence, the value of γ is the fractional increase in mass between $z = 0.1$ and $z = 0$.

(x) N_{fil} : the number of filaments, or connectivity, of a cluster. The connectivity of each cluster was calculated using the DIScrete PERsistent Structure Extractor (DisPerSE) filament finding algorithm (Sousbie 2011). In our specific case, the cosmic filaments were identified based on the number density of gas particles around a cluster. Multiple definitions for the connectivity of clusters exist – we use it to refer to the number of these filaments beginning at the cluster centre (node) and passing through a sphere of radius R_{200} surrounding the cluster. Cosmic filaments around the clusters in THE300 have been studied extensively in other works via galaxies (Cornwell et al. 2022; Kuchner et al. 2022) and their gas component (Santoni et al. 2024).

(xi) $D_{n,f}$: this environment parameter is defined as the distance to the n th nearest halo whose mass is greater than fM_{200} , in units of R_{200} . In our case, we use $n = 1$, $f = 0.1$, and so $D_{1,0.1}$ is the distance to the nearest halo whose mass is greater than one-tenth of the cluster's mass (see also Jeason-Daniel et al. 2011; Wong & Taylor 2012). This is consequently a measure of how isolated a galaxy cluster is from other clusters with masses of the same order of magnitude.

(xii) N_{200} : richness, the number of cluster members whose absolute magnitude is between m_3 and $m_3 + 2$, where m_3 is the magnitude of the third-brightest cluster member (this definition is given by Abell 1958, and while multiple definitions of cluster richness exist, this is the one we use throughout this work). We use R -band luminosities, calculated using the stellar population synthesis code STARDUST (see Devriendt, Guiderdoni & Sadat 1999).

(xiii) m_{12} : difference in magnitude between the brightest and second-brightest cluster member galaxies, or 'magnitude gap'. Multiple names are used throughout the literature to refer to this property, or similar properties (e.g. the 'fossil parameter' in Ragagnin et al. 2019, also used as a measure of dynamical state). As in our calculations of cluster richness, we calculate this using R -band magnitudes.

(xiv) σ_{BCG} : the velocity dispersion of stars in the BCG of our simulated clusters. This was calculated using the stellar particles located within a spherical aperture of radius $200 h^{-1} \text{kpc}$ around the centre of the cluster halo, a radius which should include most of the material associated with the BCG (Lin & Mohr 2004; Contreras-Santos et al. 2022).

It should be noted that the order in which we present these 17 dynamical state indicators is arbitrary. In principle, one could group these parameters together – for example, $z_{0.5}$, $z_{\text{merge},50}$, and γ all depend on temporal information about a cluster. However, we have not chosen to do this here, as these properties are taken by our analysis in a random order, without any further information relating them to one another. Later, we will deliberately reorder these 17 parameters based on our PCA (see Section 3.1).

These measurements are all dependent on information that is only available from the full 3D version of our simulations. While the strength of this dependence varies, it means that none of these are directly measurable from observations. Some of these properties can be measured quite well in observations, such as the velocity dispersion of the BCG, σ_{BCG} , which can be measured using spectroscopy, and the magnitude difference between the two brightest galaxies, m_{12} . Conversely, some are much more sensitive to 2D projection effects. For example, the centre-of-mass offset, Δ_r , is strongly dependent on the viewing angle, particularly if the offset is caused by a single major merger event. If the merger takes place along the line of sight, the apparent offset will be minimal. If instead it occurs in the plane of the sky, the calculated offset of the centre of mass will be apparent and measurable (see Zenteno et al. 2020). A similar property is the sphericity, c/a – very different values of this will be measured depending on whether a cluster's major axis is aligned along the line of sight or not. Other properties such as the substructure fraction, f_s , and cluster richness, N_{200} , are likely to be measured quite well in observations, but will still suffer somewhat due to the presence of interloper galaxies along the line of sight. Additionally, some of these cluster properties cannot be measured at all. The cluster formation time, $z_{0.5}$, requires knowledge of the growth histories of a cluster over several gigayears, which can be inferred from other properties but not measured directly.

The fact that some of these properties are difficult, or impossible, to measure observationally is a topic we plan to address in a future study (see also Section 4). However, in this work our focus is on the actual properties of a cluster, rather than the limitations of what can be measured. The exception is the mock X-ray and SZ data that we use, which we describe in the following subsection.

2.2.2 X-ray and SZ morphological parameters

In De Luca et al. (2021), mock observations were created for all the clusters of THE300 sample. Maps were generated for each of these clusters in X-rays, and also as they would be seen through the SZ effect, in terms of the Compton parameter. From these maps, six morphological parameters are calculated, as described below. More rigorous descriptions of these parameters, as well as details on the production of the mock maps and a more thorough review of the literature, can be found in De Luca et al. (2021).

(i) *A*: asymmetry (Schade et al. 1995), the normalized difference in flux between the original map, and a rotated map. The value of *A* chosen is the maximum calculated from four different rotations/reflections (90°, 180°, and reflection along the main cluster axes).

(ii) *K*: light concentration ratio (Santos et al. 2008), the ratio of surface brightness computed within two concentric apertures. For the X-ray maps these are $0.025 R_{500}$ and $0.25 R_{500}$, and for the SZ maps these radii are $0.05 R_{500}$ and $0.25 R_{500}$.

(iii) *W*: centroid shift (Mohr, Fabricant & Geller 1993), a measure of how much the centroid of a map shifts as different apertures are used to calculate the centre.

(iv) *P*: power ratio (Buote & Tsai 1995) is based on a multipole decomposition applied to the maps of the ICM. Specifically, *P* is the third-order power ratio.

(v) *G*: Gaussianity (Cialone et al. 2018), the ratio of the two values for standard deviation required to describe a 2D Gaussian fit to the map. This can distinguish elongated and circular clusters, and so is analogous to the cluster sphericity, c/a .

(vi) *S*: strip variation (Cialone et al. 2018), the normalized difference between four light profiles, inclined by 45° to one another, passing through the centroid.

Additionally, De Luca et al. (2021) calculate *M*, a normalized, linear sum of these six morphological parameters, each weighted such that the difference in *M* between relaxed and unrelaxed clusters is maximized. Each of these seven total parameters (six parameters, plus the combined measure) is calculated for both X-ray and SZ-effect maps, and we use subscripts to distinguish between these. For an additional study using similar methods, we also refer the reader to Campitiello et al. (2022).

3 RESULTS

3.1 PCA of dynamical state indicators

Principal component analysis (PCA) is a commonly used dimensionality reduction technique, which defines new variables (principal components) in a multidimensional parameter space. These principal components are linear sums of the input parameters, defined by the eigenvectors of the covariance matrix of the data, meaning they are orthogonal and uncorrelated to one another. The principal components can also be ordered based on the variance in the data for which they account, thus allowing one to only consider the ‘most important’ components. These components can be interpreted physically, and can be used to identify correlations, trends, and degeneracies in high-dimensional data. PCA requires the input data to be standardized; we have applied this to each of our 17 parameters, and also for our subsequent UMAP analysis (Section 3.4). Some discussion of the non-standardized distributions can be found in Appendix B.

Fig. 1 shows the contributions of each of the 17 dynamical state

indicators to the four major principal components, PC1, PC2, PC3, and PC4, as determined by PCA. For each dynamical state indicator, its contribution (i.e. the coordinate value in that dimension) to each of these four components is shown by the height of the bars, coloured based on the component number. Horizontal lines are marked at ± 0.24 . We take component coordinates that have an absolute value of greater than 0.24 to be ‘important’ contributing parameters to a principal component; as these principal components are normalized, the root-mean-square contribution is $N_{\text{par}}^{-0.5} \approx 0.24$, where N_{par} is the dimensionality of the data (17 in our case). While this distinction is still somewhat arbitrary, it is in line with previous similar work, such as Wong & Taylor (2012), who use a boundary of 0.3 for $N_{\text{par}} = 13$. These data are also shown in Table 1, and component coordinates greater than 0.24 are shown in bold.

Throughout the remainder of this work, we consider only these four dominant principal components. Between them, these four explain 64 per cent of the variance of the 17 dynamical state indicators (38 per cent, 14 per cent, 6 per cent, and 6 per cent for PC1–PC4, respectively), as shown in Fig. A1. We choose to consider only four parameters as this is the minimum number such that every dynamical state indicator contributes strongly to at least one principal component, and one of the aims of this work is to group these indicators in as simple a way as possible. While there is not a sharp decrease in the importance of components after PC4, we do not believe that including these additional components allows for significantly more scientific interpretation; this is discussed further in Appendix A.

The major contributors to each principal component can be summarized as follows:

- (i) PC1: substructure fraction, centre-of-mass offset, formation time, accretion rate, cluster richness, and dominance of BCG.
- (ii) PC2: virial ratio, concentration, accretion rate, and BCG velocity dispersion.
- (iii) PC3: cluster spin, connectivity, and distance to nearest large halo.
- (iv) PC4: sphericity, time since last major merger, and BCG velocity dispersion.

Throughout the remainder of this work, we have reordered our 17 indicators from the order in which they were presented in Section 2.2.1, based on the principal component to which they strongly contribute (see Fig. 2, for example).

We can interpret these principal components physically, as four different forms of dynamical state.

PC1 describes the formation time of a cluster, and the properties of galaxies in this cluster – that is, whether a cluster is a recently forming rich cluster with many bright galaxies, or an old, poor cluster dominated by a single BCG. This component primarily describes the history of substructure accretion by this cluster, and is the ‘most important’ principal component, explaining more than one-third of the total variance in the data set.

PC2 describes the relaxation and virialization of the dark matter halo of the cluster. It is dependent on whether the halo is virialized or not, and how concentrated it is. In turn, the BCG velocity dispersion is also included in this. Highly concentrated haloes have a greater BCG velocity dispersion as their central dark matter density is greater, while low-concentration haloes (those with more of a central core) have a lower BCG velocity dispersion.

PC3 represents the local environment of a cluster: how connected it is to cosmic filaments, and whether it is in an isolated region of the Universe. The local environment will impact the shear forces on a cluster, potentially explaining the inclusion of spin in this component.

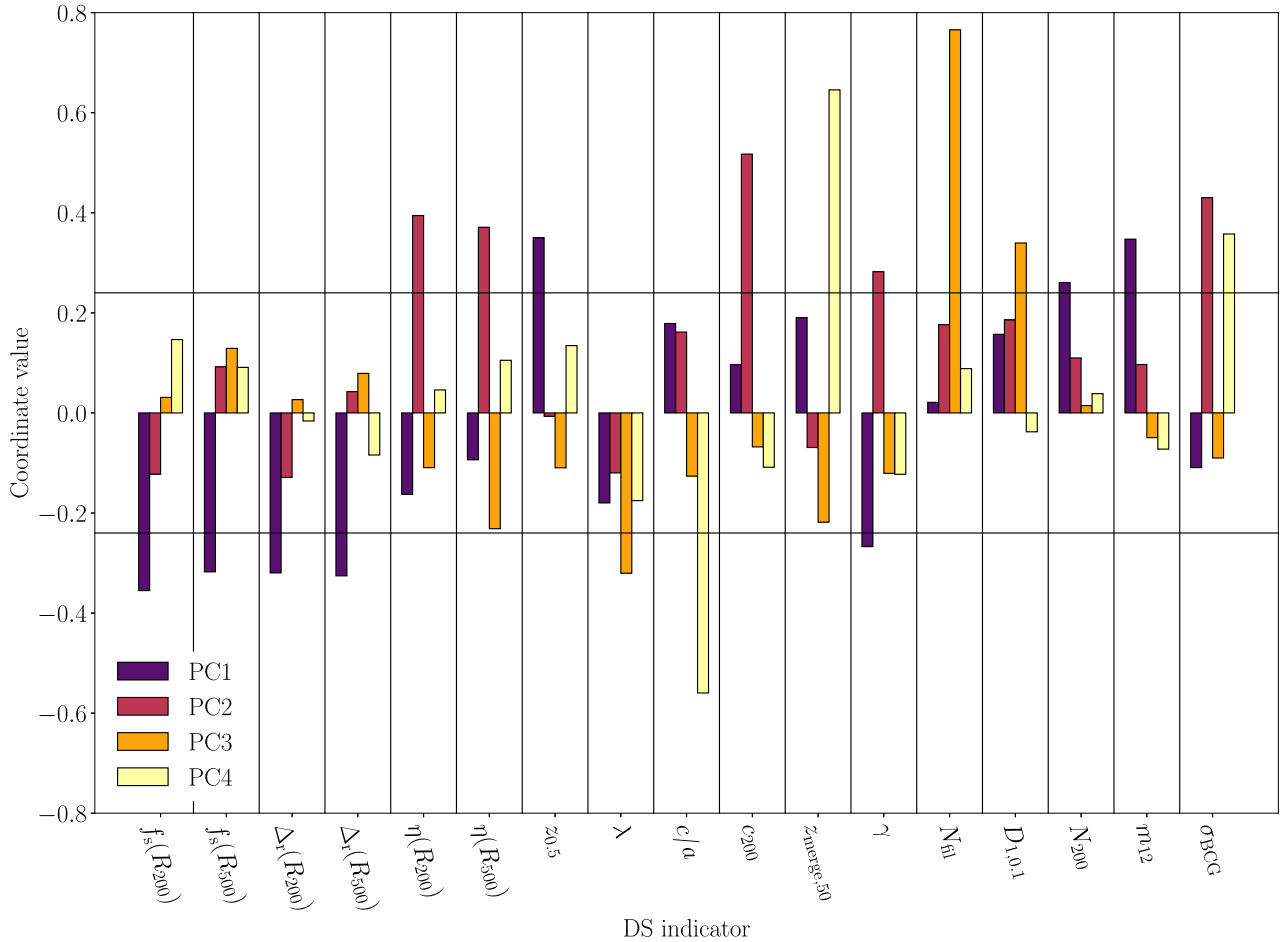


Figure 1. Contribution of each of the 17 dynamical state parameters to the four main principal components, as determined by PCA. Dynamical state indicators are listed horizontally, and the value of the bars shows the coordinate value of that dimension for PC1, PC2, PC3, and PC4, coloured based on the component number. Horizontal lines are marked at ± 0.24 , the root-mean-square contribution of any parameter to a given principal component – this is the boundary at which we consider a parameter to be an ‘important’ contributor to that principal component. These data are tabulated in Table 1.

Table 1. Coordinate values for the four major dynamical state principal components (these data are also shown in Fig. 1). Component coordinates with an absolute value greater than 0.24 are highlighted in bold.

Parameter	PC1	PC2	PC3	PC4
$f_s(R_{200})$	-0.36	-0.12	0.03	0.15
$f_s(R_{500})$	-0.32	0.09	0.13	0.09
$\Delta_r(R_{200})$	-0.32	-0.13	0.03	-0.02
$\Delta_r(R_{500})$	-0.33	0.04	0.08	-0.08
$\eta(R_{200})$	-0.16	0.39	-0.11	0.05
$\eta(R_{500})$	-0.09	0.37	-0.23	0.11
$z_{0.5}$	0.35	-0.01	-0.11	0.13
λ	-0.18	-0.12	-0.32	-0.18
c/a	0.18	0.16	-0.13	-0.56
c_{200}	0.10	0.52	-0.07	-0.11
$z_{\text{merge},50}$	0.19	-0.07	-0.22	0.65
γ	-0.27	0.28	-0.12	-0.12
N_{fil}	0.02	0.18	0.77	0.09
$D_{1,0.1}$	0.16	0.19	0.34	-0.04
N_{200}	0.26	0.11	0.01	0.04
m_{12}	0.35	0.10	-0.05	-0.07
σ_{BCG}	-0.11	0.43	-0.09	0.36

PC4 includes the information on whether this cluster is a recent merger. Currently merging clusters are unlikely to be spherical as they will consist of a superposition of two approximately spherical haloes. Moreover, Contreras-Santos et al. (2022) use THE300 data to show that BCG properties are strongly impacted by major mergers, explaining the inclusion of BCG velocity dispersion in this component.

Interestingly, we note that PC4 actually consists of a positive $z_{\text{merge},50}$ component and a negative c/a component, implying that the value of this component is greater for elongated clusters, and those that last merged long ago. While this is not necessarily expected, we believe that this behaviour is a consequence of the definition of $z_{\text{merge},50}$, which only includes mergers that have already finished – ongoing mergers will not be counted as a ‘recent merger’, even though their measure of c/a will be strongly impacted (this is the same as the ‘reduced’ merger sample in Contreras-Santos et al. 2022). Additionally, we note that both of these parameters also appear in PC1 (albeit with lower contributions than in PC4) and vary in the same direction. PC1 explains more of the total variance than PC4, and so some of the merger behaviour is also encompassed within PC1, as well as in PC4. This is supported by the fact that there is no significant overall correlation between $z_{\text{merge},50}$ and c/a (Spearman’s rank, $\rho_s = 0.06$, $p = 0.30$, as shown later in Fig. 3).

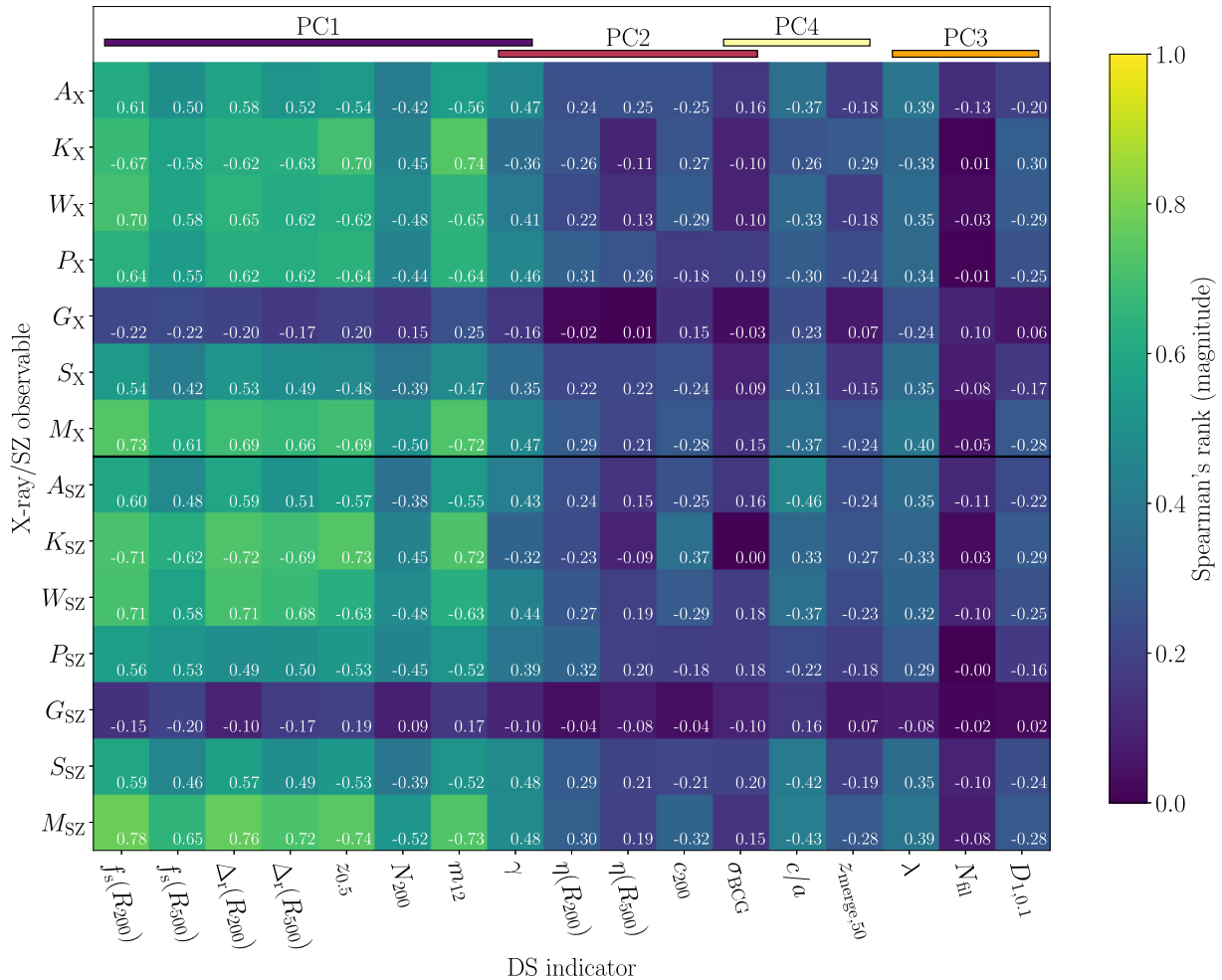


Figure 2. Spearman's rank correlation coefficient, ρ_s , between the 17 dynamical state parameters described in Section 2.2.1, and the seven observable morphological parameters described in Section 2.2.2, for both X-ray and SZ mock observations (separated by a horizontal black line, with X-ray measurements in the top half). The colour of each cell represents the (absolute) value of the Spearman's rank; the lighter colours represent a stronger correlation. The value of the Spearman's rank is also written in each cell. Note that the 17 dynamical state indicators have been reordered, and grouped together based on the principal component to which they contribute (see Section 3.1).

Similarly, we note that PC2 includes positive contributions from η , c_{200} , γ , and σ_{BCG} , indicating that a greater value of this component corresponds to clusters that are non-virialized, rapidly accreting, have a disturbed BCG, and are highly concentrated. Mergers are generally known to make cluster haloes less concentrated, and so this result is also unexpected. It can potentially be explained by the fact that some of the components of PC2 (such as γ) also contribute to PC1, similarly to how some elements of PC4 contribute to PC1. Also, as shown in Wang et al. (2020), the relationship between mergers and halo concentration can be somewhat complex and non-monotonic; we discuss this further in Section 3.4.

Each of these four principal components are, by definition, orthogonal and uncorrelated, and they are each driven by specific properties of a cluster. However, it is important to note that these four different forms of the dynamical state of a galaxy cluster are not independent of one another; multiple dynamical state indicators contribute in a non-negligible way to several principal components. We do not perform a detailed, qualitative analysis of the relationship between these different forms in this work, but instead focus on exploring these four different dimensions of dynamical state.

3.2 X-ray and SZ morphological indicators

This paper is primarily a theoretical study, focusing on properties that are taken directly from simulations, rather than from mock observations of simulated clusters. In this section we draw some connections to observational astronomy, by comparing these parameters to mock X-ray and SZ observations of the same galaxy clusters.

To test which of the different interpretations of 'dynamical state' described in Section 3.1 are actually probed by X-ray and SZ morphological parameters, in Fig. 2 we show the Spearman's rank correlation coefficient between each of these 12 observables (plus the composite parameters, M_X and M_{SZ}) and the 17 dynamical state indicators that we use. Here we only show the magnitude of the Spearman's rank – that is, we do not distinguish between positive and negative correlations, we are only studying the strength of the correlation.

Overall, the morphological parameters correlate well with dynamical state indicators that contribute strongly to the primary principal component, PC1, with a median correlation coefficient of 0.53 (median of the absolute values of the Spearman's rank coefficients). The exception to this is the Gaussianity of the cluster, which does

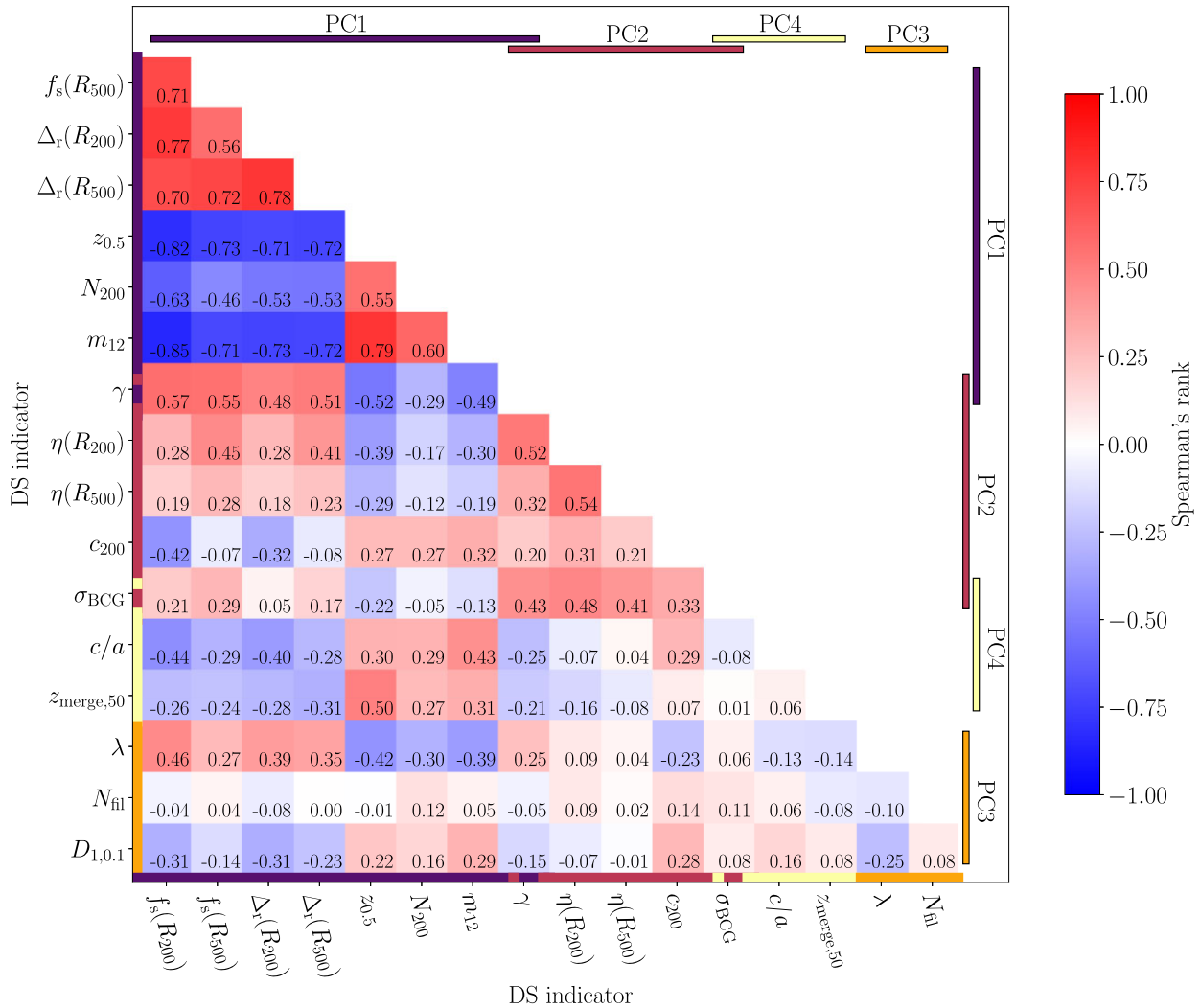


Figure 3. Spearman's rank correlation coefficient, ρ_s , between all 3D dynamical state parameters. The darker coloured cells represent stronger correlations; the value of the Spearman's rank is also written in each cell. A corner plot showing the distributions of each of these parameters can be found in Appendix B. The principal component to which each dynamical state indicator contributes is indicated at the top of the figure, and by the coloured bars along the axes.

not correlate as well with the PC1 parameters, regardless of whether it is measured from X-ray or SZ maps. This result is not unexpected – previous studies have also found that Gaussianity is generally a less informative parameter than other X-ray/SZ morphological parameters. For example, De Luca et al. (2021) show that G is weighted three to four times weaker than the other parameters in the calculation of the composite parameter M . They attribute this to several factors: Gaussianity is mostly dependent on the global shape of a cluster, not on cluster substructure, and is also highly dependent on projection effects. Furthermore, even virialized dark matter haloes can have an ellipsoidal (non-spherical) shape, and so it is not necessarily a good metric for distinguishing relaxed and unrelaxed clusters. Similarly, Cialone et al. (2018) show that Gaussianity calculated from SZ maps is less effective at separating clusters by dynamical state, and give similar reasons for this finding.

In both of these sets of mock observations, the weighted sums of morphological parameters (M_X and M_{SZ}) correlate most strongly with the elements of PC1. This is predictable, given that M is defined in such a way that the difference in M between relaxed and unrelaxed clusters is maximized. These results are in general agreement with

De Luca et al. (2021). It is important to note that the apertures for calculating these morphological parameters were chosen such that clusters could be separated based on $f_s(R_{500})$ and $\Delta_r(R_{500})$. However, the correlations of the morphological parameters with $f_s(R_{200})$ and $\Delta_r(R_{200})$ are similarly strong, and so we do not consider this to be an issue in our analysis.

The X-ray and SZ morphological parameters correlate far less well with the parameters that contribute to PC2, PC3, and PC4 – the median (absolute) Spearman's rank coefficient is equal to 0.24 for these combinations. The exception to this is, once again, the cluster Gaussianity, particularly in X-rays. For parameters in PC1, the correlation coefficients associated with G_X were lower than those for the other morphological indicators, but for several parameters in PC2, PC3, and PC4, G_X is comparable to the other morphological indicators. For example, G_X has a moderate correlation with the cluster sphericity c/a ($\rho_s = 0.23$, $p = 2 \times 10^{-5}$), due to the fact that G_X is a direct measure of a cluster's shape.

These results are summarized in Table 2, which shows how each of these 14 morphological measurements of each cluster correlate with the four main principal component values of that cluster. Indeed,

Table 2. Spearman’s rank correlation coefficient between X-ray and SZ morphological parameters, and each of the four principal components. These data are similar to that shown in Fig. 2, but showing the correlation coefficient for each full principal component, not for their contributing dynamical state indicators.

Observable	PC1	PC2	PC3	PC4
A_X	−0.65	−0.11	−0.08	0.08
K_X	0.72	0.13	−0.07	0.04
W_X	−0.71	−0.17	0.04	0.06
P_X	−0.70	−0.03	0.01	0.02
G_X	0.25	0.11	0.02	−0.07
S_X	−0.57	−0.12	−0.04	0.06
M_X	−0.78	−0.13	0.01	0.05
A_{SZ}	−0.64	−0.11	−0.04	0.11
K_{SZ}	0.75	0.23	−0.11	0.02
W_{SZ}	−0.73	−0.11	−0.01	0.07
P_{SZ}	−0.60	−0.01	0.02	0.03
G_{SZ}	0.18	−0.05	−0.05	−0.05
S_{SZ}	−0.63	−0.06	−0.08	0.12
M_{SZ}	−0.82	−0.14	0.02	0.06

we see that (apart from the Gaussianity), each of these X-ray/SZ measures correlates well with PC1, which we interpret as a measure of the time since much of the cluster’s galaxy population was accreted and built up (see Section 3.1). However, the halo virialization (PC2), local environment (PC3), and recent merger history (PC4) do not correlate as well with any of the morphological parameters.

Despite the fact that the observable morphological properties of a cluster only strongly correlate with the dynamical state indicators that make up PC1, there are still correlations between other indicators. Fig. 3 shows the Spearman’s rank correlation coefficient between all 17 of these 3D (non-observational) parameters. For example, the time since the last major merger, $z_{\text{merge},50}$, and the cluster formation time, $z_{0.5}$, have a positive correlation ($\rho_s = 0.50$). This result is quite intuitive, as clusters that have recently merged will not have a high formation redshift. Other dynamical state indicators also correlate well – for instance, all five of the parameters that contribute to PC2 show weak to moderate correlations with each other ($|\rho_s| \geq 0.20$, $p \leq 2 \times 10^{-4}$). These correlations are weaker for the dynamical state indicators in PC3 and PC4, but this is not unexpected given that these components explain less of the variance of the total data set (see Fig. A1).

This figure also displays some counter-intuitive results. For example, one might consider clusters with lots of substructure to be ‘rich’, but there is actually a negative correlation between f_s and N_{200} (for $f_s(R_{200})$, $\rho_s = -0.63$, and for $f_s(R_{500})$, $\rho_s = -0.46$). This is due to the fact that N_{200} is defined by the number of galaxies of similar magnitude to the third brightest, and clusters with lots of substructure are more likely to have several bright galaxies, meaning that the threshold for galaxies to be counted in N_{200} is higher. Nevertheless, N_{200} is a useful dynamical state indicator, although using the term ‘richness’ for this is somewhat ambiguous. While we do not explicitly show the correlations between these parameters in Fig. 3 (for the sake of clarity), we have included a corner plot showing these correlations in Appendix B.

3.3 Mass accretion histories of clusters

In order to try and learn more about the dynamical histories of these clusters, we separate our sample of 324 clusters based on the coordinate values of their principal components, and study their

mass accretion histories. A similar analysis was performed by Wong & Taylor (2012), who also split simulated dark matter haloes into different classes based on PCA analysis. They show that the mass accretion histories of their classes of clusters differ – their first principal component separates the clusters into early-forming and late-forming, and their second component describes whether a halo’s growth is accelerating or decelerating.

For our similar analysis, we find the upper and lower quartiles of the values of each of our principal components. For example, based on PC1, we have a group of clusters that have low values of PC1, and a group with high values. As the contribution of $z_{0.5}$ to PC1 is positive (0.35, see Table 1), this also implies that selecting clusters with low values of PC1 is equivalent to choosing a group with low values of $z_{0.5}$, making them a ‘late-forming’ quartile. Similarly, we describe the group with high values of PC1 (and so higher values of $z_{0.5}$) as the ‘early-forming’ quartile. We then look at the median mass accretion history of the clusters in each of these two extreme groups, ignoring the intermediate clusters. The median mass accretion histories for each principal component are shown in Fig. 4.

When splitting clusters by their PC1 coordinate values, we see a clear difference between the mass accretion histories. As expected, the mass of the early-forming clusters is built up far earlier. On average, the early-forming clusters have built up half of their mass by a redshift of 0.7, while the late-forming clusters do not do this until a redshift of 0.2. This is very similar to the first principal component found by Wong & Taylor (2012), which also splits their haloes into early-forming and late-forming.

We interpret the PC2 coordinate values as separating the clusters based on how virialized are their dark matter haloes. Separating the clusters based on their PC2 coordinate values, we find that the difference in their mass accretion histories is less pronounced than for PC1, and that there is a significant overlap in the spread of the data. However, by ‘bootstrapping’ the data, we can study the uncertainty in the median mass accretion histories. From this, we find that there is a significant difference between the clusters split by PC2, particularly at low redshifts. The ‘more virialized’ clusters have experienced only a very small change in mass since $z = 0.1$, having an average mass at $z = 0.1$ of $(0.97 \pm 0.01)M_{200}(z = 0)$ (median and uncertainty). Meanwhile, the ‘less virialized’ clusters have an average mass of $(0.80^{+0.02}_{-0.07})M_{200}(z = 0)$ at $z = 0.1$. This indicates that this ‘virialization’ component is strongly dependent on the very recent growth history of a cluster, although we emphasize that there is a large overlap in the spread of these data. We also note that the shapes of these mass accretion histories are similar to the mass accretion histories of haloes that Wong & Taylor (2012) separate by PC2. They describe these as accelerating or decelerating growth rates, although their PC2 component is interpreted as describing the shape and spin of their haloes, not the ‘virialization’ as we find. Some slight variations between our results and those of Wong & Taylor (2012) are to be expected, as their work uses dark-matter-only simulations, not hydrodynamical simulations. The baryonic effects in THE300 are particularly strong in the cluster centres (Haggard et al. 2021), a region on which PC2 appears to be strongly dependent (see also Section 4).

The component values of the third principal component, PC3, do not appear to have a strong impact on the mass accretion histories of clusters. Splitting the clusters by their coordinate values in this component (into ‘symmetric environments’ with many filaments and a large distance to the nearest large halo, and ‘asymmetric environments’ with few filaments and a nearby large halo) does not show a large difference in mass accretion histories. This indicates that the present-day local environment of a cluster is not closely tied to its mass accretion history, although Fig. 3 does show a weak correlation

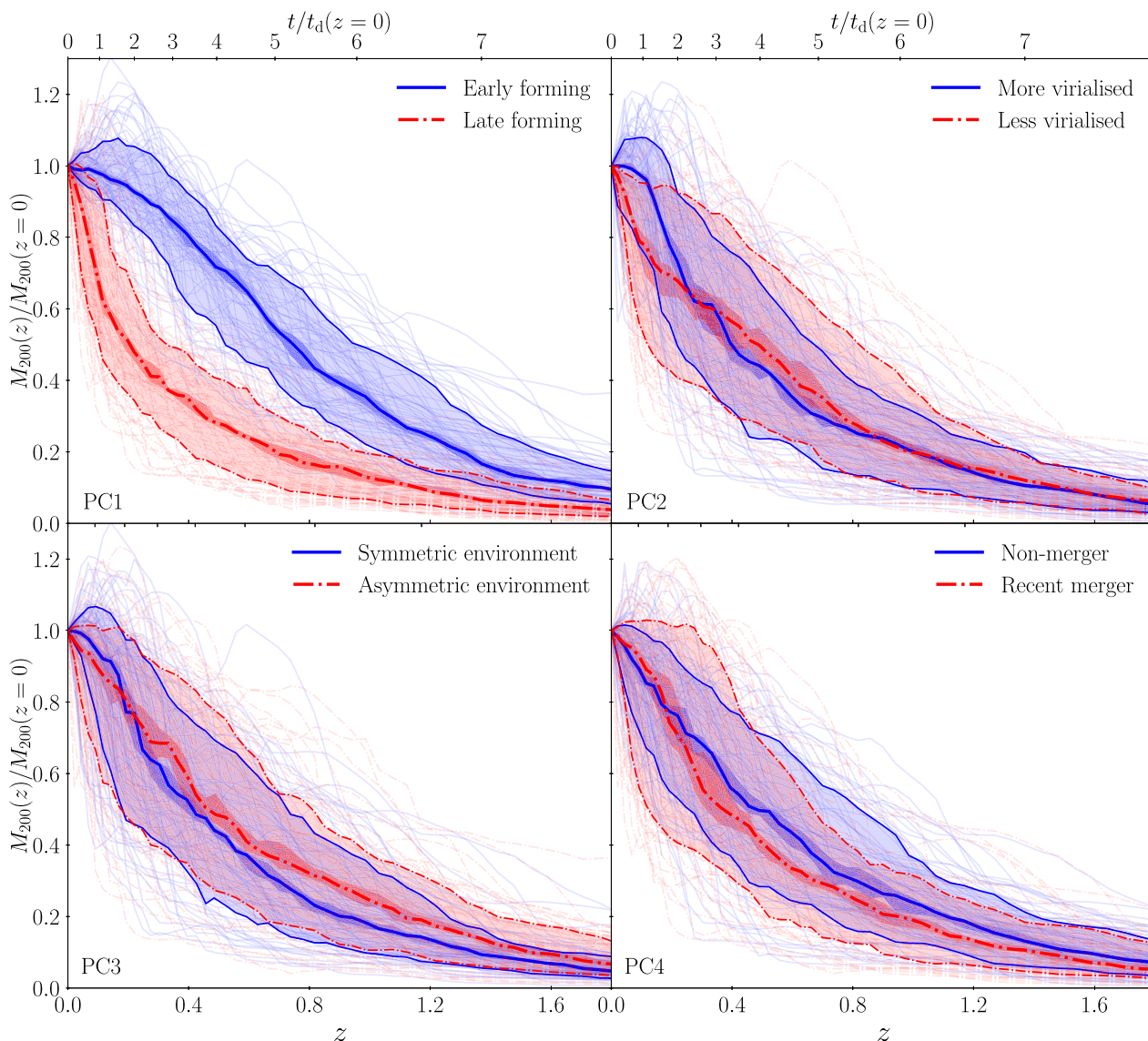


Figure 4. Median mass accretion histories for clusters, split based on the values of their principal components, with a different principal component considered in each panel (PC1 in top-left, PC2 in top-right, PC3 in bottom-left, PC4 in bottom-right). The upper and lower quartiles of each principal component are shown. The horizontal axis shows redshift (bottom axis), and lookback time in units of the dynamical time at $z = 0$ (top axis), equal to approximately 1.4 Gyr as defined in equation (5). The vertical axis shows the fraction of the $z = 0$ mass that has been built up by a given time. The faint lines show the mass accretion histories of individual galaxy clusters. Some of these have temporary decreases in mass, which can be explained by several well-established mechanisms, such as the decrease in M_{200} as a halo virializes after a period of rapid growth (e.g. Tormen, Bouchet & White 1997), or through mismatches in halo tracking (Fakhouri & Ma 2009; Behroozi et al. 2015). The thick lines show the median accretion history for that group, the light shaded regions and thin lines show the 1σ (16–84 percentile) spread in the data, and the dark shaded regions show the uncertainty in the median.

between $z_{0.5}$ and $D_{1,0.1}$ ($\rho_s = 0.22$, $p = 6 \times 10^{-5}$). We also note that a physical interpretation of this component is challenging. Based on past work, one would expect that relaxed clusters have both fewer filaments and a large distance to their nearest neighbour, but splitting clusters based on their PC3 values puts clusters with many filaments in a group with those that have a large distance to the nearest neighbour. However, this is likely due to the weak correlations between the components of PC3 that we find throughout this work, particularly N_{fil} (also shown in Fig. 3). We discuss the strength of these correlations further in Section 3.4.

For the fourth principal component, PC4, the accretion histories are similar, although there is a slight difference in the shapes of the

profiles; similarly to in PC1, the ‘relaxed’ (non-merging) clusters build up their mass at slightly earlier redshifts. However, the main difference between the recent mergers and non-mergers is that the spread in mass accretion histories is far greater in recently merged clusters. For example, at $z = 0.2$, the non-merging clusters have an average mass of $0.77^{+0.17}_{-0.23}$ times their present-day mass (median and 1σ spread). In contrast, the recently merging clusters have an average mass of $0.73^{+0.28}_{-0.32}$ times their present-day mass, corresponding to a 50 per cent greater spread. This increased spread in accretion histories for recently merging clusters is likely due to the stochastic nature of mergers – these objects will have experienced a large jump in their mass at some recent time, but the exact time of this jump varies

between clusters. Again, this is similar to the findings of Wong & Taylor (2012), who explain that mass accretion histories are not well modelled by smooth curves due to the stochasticity of merger events.

3.4 Cluster dynamical state with UMAP

Uniform manifold approximation and projection (UMAP; McInnes et al. 2018) is an alternative dimensionality reduction technique to PCA. UMAP is commonly used as part of more complex machine learning studies, but is also a useful data visualization method in its own right. UMAP involves first constructing a graph of data points (in our case, 324 galaxy clusters), each connected to their nearest N neighbours in a high-dimensional space (the 17 dimensions defined by the dynamical state indicators). Some additional connections between points are also added, with a decreasing likelihood for points separated by greater distances. Next, a low-dimensional space, the ‘embedding space’, is populated randomly with (324) corresponding data points. These are iteratively repositioned according to some loss function. This function is designed to match the pairwise distances between connected points in the new low-dimensional space to the pairwise distances in the original high-dimensional space.

This process preserves local structure in the data set, meaning that objects with similar properties remain close together in the low-dimensional ‘embedding space’. It also preserves global properties of the data set, so that groups of objects that are far apart in the high-dimensional space remain separated from one another. Similarly to PCA, UMAP is able to reduce a high-dimensional data set into a smaller number of parameters (in our case, two). Unlike PCA however, the stochastic, iterative nature of UMAP allows for complex, non-linear relationships between parameters to also be captured. This means that the outputs of UMAP are harder to physically interpret – whereas PCA produces mathematically well-defined axes, the output ‘axes’ of UMAP are determined numerically. Similarly, a simple measure of the fractional variance that is captured by UMAP, analogous to the ‘explained variance’ of PCA, does not exist (McInnes et al. 2018). We use UMAP as a complementary method to PCA, approaching the same problem but in a different way.

Fig. 5 shows the results of our UMAP analysis for the 17 dynamical state indicators also used in the PCA. Each point represents one galaxy cluster, with the horizontal and vertical axes showing the two combined parameters produced by the two-dimensional UMAP analysis, analogous to the principal components that we summarized in Table 1. Due to the definition and optimization processes in UMAP, points that are close together on a UMAP plot are close together in the 17-dimensional space that we started with. Consequently, these are likely to be dynamically similar clusters. As the UMAP axes are a dimensionless (due to the standardization of the data), complex combination of many parameters, these axes are left unlabelled for clarity.

Each panel in this figure shows data for one of our 17 3D dynamical state indicators, and the colour of each point in this panel represents the value of this dynamical state indicator for each cluster. Linear colour scales are used for each, but the colour scale is flipped for some quantities, such that the clusters we interpret as more dynamically relaxed (according to this individual parameter) are shown by darker colours. This allows us to see which regions of this new UMAP embedding space contain dynamically relaxed clusters, according to each of the different definitions. Similarly to in Fig. 2, the panels are grouped by the principal component to which they most strongly contribute, according to our PCA – otherwise, this approach is fully independent of our PCA.

For each of these panels, we calculate the direction in which the z -axis (i.e. the value of the dynamical state indicator) varies most quickly, and thus the direction in this space in which the clusters become more dynamically unrelaxed. We do so using the partial correlation coefficients (PCCs) approach described in Lawrance (1976); see also Baker et al. (2022) for a clear explanation of this method. The PCC between two quantities, A and B , while controlling for a third quantity, C , is given by $\rho_{AB|C}$, and is calculated using the equation below:

$$\rho_{AB|C} = \frac{\rho_{AB} - \rho_{AC}\rho_{BC}}{\sqrt{1 - \rho_{AC}^2}\sqrt{1 - \rho_{BC}^2}}. \quad (6)$$

Here, ρ_{XY} is the Spearman’s rank correlation coefficient between two quantities, X and Y . Equation (6) allows the partial correlation coefficient between the colour (the value of the dynamical state indicator) and the vertical UMAP axis to be found, while controlling for the horizontal axis. Similarly, we find the PCC between the colour and the horizontal UMAP axis, while controlling for the vertical axis. The ratio of these two PCCs can then be used to calculate the maximal variation direction of the colour in the UMAP embedding space (see also Bluck et al. 2020, for further details of this method). We plot each of these directions of maximal dynamical state variation in the top-left panel of Fig. 5, coloured by the principal component to which they contribute most strongly. The length of each arrow is equal to the Spearman’s rank correlation coefficient between the dynamical state indicator (colour) and the position of clusters along this direction in embedding space. Each arrow is also shown on the bottom-left corner of the panel to which it corresponds. This top-left panel therefore shows which direction in this embedding space contains the unrelaxed clusters, according to each of the 3D dynamical state measures.

Fig. 5 shows that each of the parameters making up the first principal component of dynamical state ($f_s(R_{200})$, $f_s(R_{500})$, $\Delta_r(R_{200})$, $\Delta_r(R_{500})$, $z_{0.5}$, m_{12} , and N_{200}) vary in approximately the same direction across the UMAP plots, from bottom to top. According to these measures, the relaxed clusters are generally found at the bottom of these plots, and the unrelaxed clusters at the top. This indicates that the UMAP has also grouped clusters based on this interpretation of dynamical state, separating clusters that have formed and developed their substructure recently, from those that formed long ago. The top-left panel of Fig. 5 confirms that the increasingly disturbed clusters, according to PC1 parameters, are at the top of the UMAP plots.

However, the dynamical state indicators that contribute to the second, third, and fourth principal components do not all vary in the same way. The PC4 parameters (c/a and $z_{\text{merge},50}$) do follow a similar trend to those in PC1, with more dynamically disturbed clusters (low- $z_{\text{merge},50}$ and low- c/a) being found near the top of the plot – this is consistent with our interpretation of PC4 in Section 3.1, where we described how some of the merger history behaviour is also included in PC1. The PC2 indicators, representing the virialization and concentration of the cluster haloes ($\eta(R_{200})$, $\eta(R_{500})$, and c_{200}), instead vary more strongly from right to left, although not all in the same direction; γ follows a similar trend. The UMAP can also separate clusters by their environmental dynamical state indicators (PC3: λ , N_{fil} , and $D_{1,0,1}$), but not very well, indicated by the weak trends in these panels. The high- λ and low- $D_{1,0,1}$ clusters, usually interpreted as unrelaxed, are located towards the top of the plot – these trends are not strong, but are significant, and are along the same axis as the PC1 components’ trends. For spin, λ , $|\rho_s| = 0.44$. For the environment parameter, $D_{1,0,1}$, $|\rho_s| = 0.24$ ($p = 10^{-5}$). The number of filaments and UMAP position correlate in the opposite

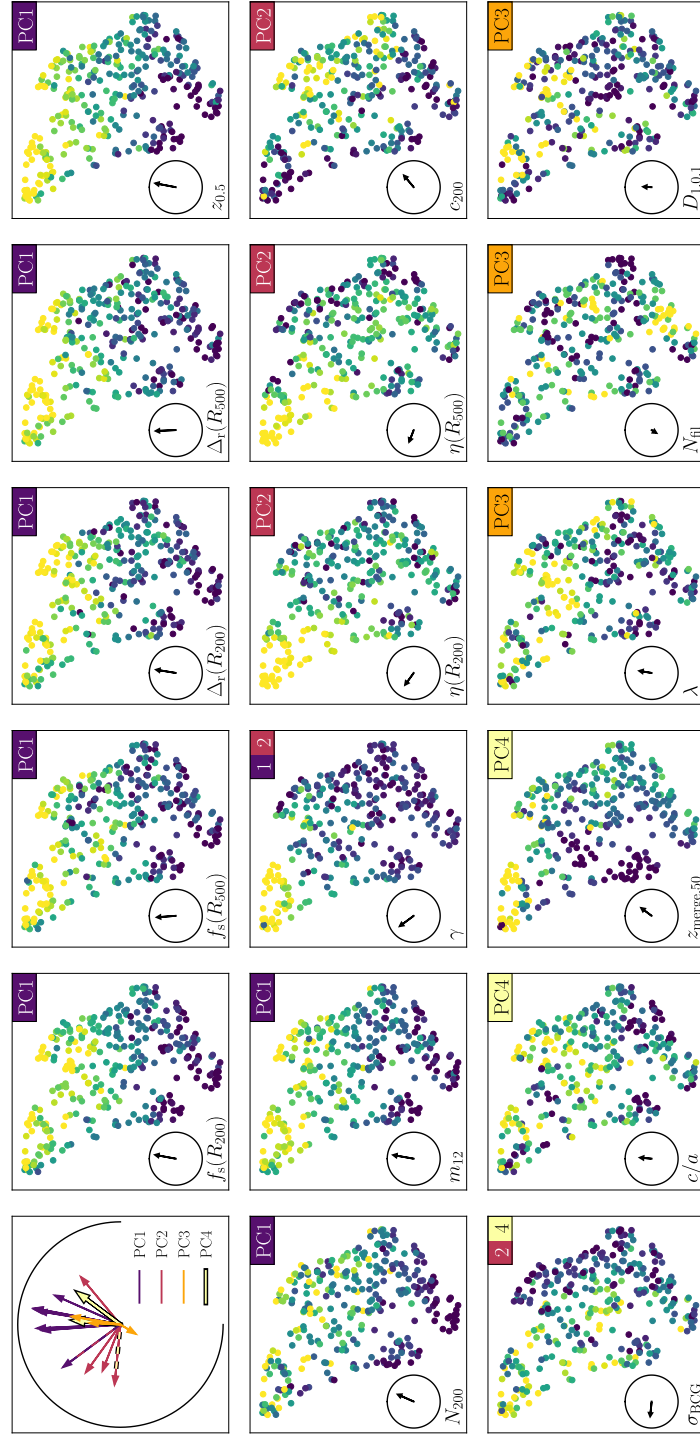


Figure 5. UMAP of 17 cluster properties, also used in PCA. For all panels apart from top-left, each point represents one galaxy cluster, with the horizontal and vertical axes representing the first and second UMAP axes, respectively. The colour of each point shows the value of one dynamical state indicator for that cluster – the specific indicator is written in the bottom-left of each panel. The colour scale has been implemented using our interpretation of the dynamical state parameters, such that a darker colour represents clusters that are dynamically relaxed according to that single parameter. For $f_s(R_{200})$, $f_s(R_{500})$, $\Delta_r(R_{200})$, $\Delta_r(R_{500})$, γ , $\eta(R_{200})$, $\eta(R_{500})$, σ_{BCG} , λ , and N_{fil} , lighter colours represent greater values of these parameters (as defined in Section 2.2.1). For $z_{0.5}$, N_{200} , m_{12} , c_{200} , c/a , $z_{merge,50}$, and $D_{1,0.1}$, lighter colours represent lower values. For simplicity, quantitative colour bars are not shown here. In the top-right of each panel, the principal component (or components, in the case of γ and σ_{BCG}) to which that parameter belongs is indicated. The dynamical state indicators are grouped in the same way as in Fig. 2. For each of these panels, we calculate the direction in which the clusters become more disturbed (less relaxed) using partial correlation coefficients (PCC) analysis. These directions, coloured by the principal component to which the parameter belongs, are shown in the top-left panel. For γ and σ_{BCG} , these arrows are dashed with colours corresponding to the two components to which they contribute. The length of each arrow represents the strength of the correlation, quantified by the Spearman’s rank correlation coefficient, ρ_s , with the black circle showing where $\rho_s = 1$. The magnitude and directions of each arrow is also included in the bottom-left of the panel to which it corresponds. With the exception of N_{fil} , all of these correlations are significant ($p \leq 10^{-5}$).

direction along the same axis, but this correlation is not significant ($|\rho_s| = 0.10$, $p = 0.06$).

This UMAP analysis gives many subtle results that are not immediately straightforward to interpret, but overall it confirms the indications of our PCA – the different groupings of clusters in the plots correspond to different classes of dynamically relaxed objects. For example, the early-forming, low-substructure clusters (PC1) are found at the bottom of the UMAP plots. These clusters have highly concentrated haloes, as would typically be expected of a relaxed cluster (Yuan & Han 2020). At the top of the UMAP plots, we instead find recently formed clusters. However, there is also a third major population of clusters in the top-left of the plots – these are dynamically unrelaxed haloes that have formed recently, have much substructure, and have non-virialized haloes, yet are highly concentrated. Clusters in this population have recently experienced major mergers ($z_{\text{merge},50}$), and have very high accretion rates, γ , indicating that they are currently accreting large amounts of material. This is in contrast to the other clusters at the top of these plots, which have recently merged but have not accreted lots of material in the last dynamical time (since $z = 0.1$). This difference is equivalent to the mass accretion histories in the top-right panel of Fig. 4, where we group clusters based on PC2 and show that their recent ($z < 0.2$) growth histories are different.

These three populations can be interpreted as follows: high-concentration galaxy clusters are likely to have formed long ago, while low-concentration clusters are likely recently formed, often after a major merger. However, if a merger is still ongoing, or if large amounts of galaxies are pulled into the cluster immediately afterwards (potentially accompanying the merging cluster), this can make a cluster appear highly concentrated, and therefore dynamically relaxed. This explanation is strongly supported by Wang et al. (2020), who show that major mergers can produce oscillations in concentration, driving concentration up significantly when merging material reaches the first pericentre of its orbit, before concentration quickly decreases again. Alternatively, if a cluster is continuing to accrete large amounts of diffuse material into its halo (rather than a single large object), this could also result in a high mass accretion rate without a corresponding decrease in halo concentration.

As a result of this, a simple measurement of the concentration of a cluster is insufficient to draw conclusions about the virialization of its halo, its formation time, or its merger history. Additionally, the dynamical state of a cluster’s BCG (σ_{BCG}) also seems to only probe some of these unrelaxed clusters. This complex behaviour cannot be fully accounted for by a 1D linear fit, which explains our previous counter-intuitive result that highly concentrated clusters appear to be less virialized and rapidly accreting (Fig. 1 and Fig. 3); simplifying this to a single correlation coefficient does not capture the complete behaviour.

Additionally, the UMAP analysis shows that the early-forming clusters can be split into two groups based on their merger histories. The group in the bottom-left have not experienced major mergers for a very long time (~ 10 Gyr), or not at all throughout their history. Meanwhile, those in the bottom-right have experienced a major merger more recently, although still not for several gigayears. There seems to be a (weak) correlation with the connectivity of these clusters, N_{fil} : early-forming clusters that have still experienced a major merger are more strongly connected than early-forming clusters that have never experienced a major merger. This may be indicative of the extremely long time-scales (~ 10 Gyr) over which cosmic filaments are persistent, but a more detailed analysis would be needed to investigate this connection.

4 DISCUSSION AND CONCLUSIONS

Using hydrodynamical simulations of galaxy clusters, we show that the wide variety of parameters used in the literature to quantify the ‘dynamical state’ of a cluster are actually probing multiple properties of a cluster. Consequently, by applying PCA, we conclude that the dynamical state of a cluster as described in the literature is actually made up of approximately four different properties, summarized in Section 3.1. To use any measurement (or set of measurements) to describe a cluster as simply ‘dynamically relaxed’ or ‘dynamically unrelaxed’ is an incomplete description. Different classes of dynamically relaxed clusters exist, and so instead one must specify in which sense is a cluster relaxed.

The main component of a cluster’s dynamical state is its ‘formation dynamical state’, describing the formation time of a cluster – that is, whether a cluster built up most of its mass and galaxy population recently or long ago. This formation state is indicated by galaxies and substructure within the cluster. Recently forming clusters have more substructure and a greater offset between their centre of mass and the position of their BCG. Additionally these clusters are typically less dominated by a single, bright galaxy, and are accreting material at a faster rate. This ‘formation state’ is similar to the dynamical state used in many previous studies (Cui et al. 2017; Gouin et al. 2021, for example). It should be noted that several such studies include the virial ratio, η , in this measure of dynamical state, making it analogous to a combination of our PC1 and PC2; however it has been shown that the virial ratio is not an important contributor to this measure of dynamical state (Haggar et al. 2020; De Luca et al. 2021). As we see in Fig. 4, separating clusters based on their first principal component from our PCA is equivalent to separating them based on their formation times, in agreement with Wong & Taylor (2012). The ‘late-forming’ clusters have a median formation time of $z_{0.5} = 0.2 \pm 0.1$, compared to $z_{0.5} = 0.7 \pm 0.2$ for the relaxed ‘early-forming’ clusters.

In addition to this dynamical description of the subhaloes and galaxies in a cluster, the diffuse material in the cluster halo itself can be virialized and dynamically relaxed. This ‘virialization dynamical state’ describes how well the material in a cluster obeys the modified virial theorem (accounting for surface pressure, see Poole et al. 2006; Shaw et al. 2006). It is particularly well described by the properties of the central regions of a cluster – the NFW concentration of the halo, and the velocity dispersion of the BCG – as well as the present-day accretion rate of the cluster, γ . The UMAP analysis in Section 3.4 also showed that, in this space, the ‘virialization state’ varies along a different axis to the ‘formation state’, described by PC1. These complex relationships can be seen through the apparent inconsistencies when only considering monotonic relationships between quantities. For example, Fig. 3 shows that highly concentrated haloes (typically considered ‘relaxed’) correlate with having less substructure (a sign of being relaxed), but also with a less virialized halo according to η . The complex relationships between these dynamical state indicators are much more apparent in Fig. 5.

Clusters can also be dynamically relaxed in terms of their local environment, or in terms of their merger history. These components of dynamical state are driven by global properties of a cluster, such as their shape and spin. These components are typically noisier, and the cluster properties that constitute these components do not correlate as well with one another. However, it is still meaningful to separate clusters based on this – for instance, Fig. 4 shows an increased spread in the mass accretion histories of recently merged clusters.

Of these four forms of dynamical state, X-ray and SZ measurements of morphological properties of clusters, such as their

asymmetry, concentration, and centroid shift, overwhelmingly probe the formation dynamical state (PC1). Generally, X-ray and SZ morphological parameters are very well suited to describing the formation of clusters and their accretion of substructures, but not well suited to describing the virialization of cluster haloes, their merger history, or their cosmic environment. This is not unexpected: as we discussed in Section 3.4, mergers appear to have a somewhat chaotic effect on other dynamical state parameters, and can lead to a cluster halo appearing either relaxed or unrelaxed. Additionally, these morphological measurements, particularly the X-ray measurements, are heavily weighted towards the central regions of clusters (R_{500} or smaller). We would therefore not necessarily expect them to be good indicators of clusters' larger scale cosmic environments. There are also some slight exceptions to this rule – for example, the light concentration ratio calculated from the SZ-effect, K_{SZ} , does indeed correlate with the NFW concentration, c_{200} ($\rho_s - 0.37$).

As stated previously, this work is very much a theoretical, simulation-focused study of dynamical state, and any further detailed analysis of observable cluster properties is beyond the scope of this paper. However, numerous other observable properties of clusters exist, and in future work we hope to examine how these observable quantities correspond to dynamical state. For example, although the mock X-ray observations used in this work are focused on the cluster centres, X-rays can also be used to map out substructures in cluster outskirts (Zhang et al. 2020), and even cosmic filaments (Walker et al. 2019; Biffi et al. 2022). Other measures of cluster X-ray and SZ maps also exist, such as decomposition into Zernike polynomials (Capalbo et al. 2021). Beyond this, optical data can also be used to quantify the dynamical states of clusters, by determining properties such as their substructure (Wen & Han 2013), richness, and brightest galaxies magnitude difference, m_{12} (Zhooldideh Haghighi et al. 2020). A future study (Cornwell et al. in preparation) will use THE300 simulations to investigate how cluster dynamical states can be determined using spectroscopic measurements.

The results from this paper confirm that, although a description of dynamical state (see Binney & Tremaine 1987) is theoretically quite simple, the dynamical state of galaxy clusters in practice is more complex. This has implications for wider work on galaxy clusters. Many studies split clusters into two samples, ‘relaxed’ and ‘unrelaxed’, based on a small number of properties, but this has the potential to combine clusters with very different dynamical histories into a single group. Instead, it is important to describe in which sense a cluster is known to be dynamically relaxed or unrelaxed – for example, in terms of their substructure accretion history, their recent merger history, or how virialized they are at the present day.

In future we plan to carry out a more quantifiable analysis of this multidimensional dynamical state, rather than the qualitative description presented in this work. Instead of a linear scale of dynamical state, it may be more natural to quantify the dynamical state of clusters in two or more dimensions; the UMAP analysis in this paper is similar to this, but it not simple to interpret physically. Such a study could also be extended to study the dynamical state in different ‘apertures’, looking at the kinematics of the cluster centre, or of its outskirts. Additionally, we plan to investigate whether clusters can be naturally separated into groups based on their dynamical states. Previous work (e.g. Zhang et al. 2022) has investigated defining a bimodal function to describe dynamical state, to allow clusters to be split into ‘relaxed’ and ‘unrelaxed’ groups. In a future study, we will investigate whether a multimodal description of dynamical state exists that would allow clusters to be split into more than two groups. This would provide a means to compare the properties of clusters that do actually have similar dynamical states, and to select samples of

truly relaxed galaxy clusters in order to reduce biases in cosmological studies.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referee for their comments on this work, which have helped to improve the paper by adding robustness and clarity.

This work has been made possible by THE THREE HUNDRED collaboration.⁴ This work has received financial support from the European Union’s Horizon 2020 Research and Innovation programme under the Marie Skłodowska-Curie grant agreement number 734374, i.e. the LACEGAL project.⁵ THE THREE HUNDRED simulations used in this paper have been performed in the MareNostrum Supercomputer at the Barcelona Supercomputing Center, due to CPU time granted by the Red Española de Supercomputación.

WC is supported by the STFC AGP Grant ST/V000594/1 and the Atracción de Talento Contract no. 2020-T1/TIC-19882 granted by the Comunidad de Madrid in Spain. He also thanks the Ministerio de Ciencia e Innovación (Spain) for financial support under Project grant PID2021-122603NB-C21 and ERC: HORIZON-TMA-MSCA-SE for supporting the LACEGAL-III project with grant number 101086388. MDP acknowledges financial support from PRIN 2022 (Mass and selection biases of galaxy clusters: a multiprobe approach – n. 20228B938N) and from Sapienza Università di Roma – Progetti di Ricerca Medi 2022, RM1221816758ED4E. JET acknowledges support from an NSERC Discovery Grant.

This work uses the SCIPY (Virtanen et al. 2020), NUMPY (van der Walt, Colbert & Varoquaux 2011), MATPLOTLIB (Hunter 2007), PANDAS (McKinney 2010), and SCIKIT-LEARN (Pedregosa et al. 2011) packages for PYTHON.⁶

The authors contributed to the paper in the following ways: RH carried out the analysis, wrote the paper, and produced the figures. FDL and MDP calculated the morphological parameters from the X-ray/SZ mock observations, which were produced by WC. ES carried out the UMAP analysis. AK produced the simulation halo catalogues and merger trees, from which several of the 3D dynamical state measures were calculated. AK also assisted with the general direction of the work, along with JET, MEG, and FRP. ACS calculated the times since last major merger. WC calculated the dynamical state indicators f_s , Δ_r , and η . UK calculated the cluster connectivities. RAMP calculated the cluster formation times, $z_{0.5}$. CP calculated the BCG velocity dispersions. All authors had the opportunity to comment on the paper.

DATA AVAILABILITY

The simulations underlying this work have been provided by THE THREE HUNDRED collaboration. The data may be shared on reasonable request to the corresponding author, with the permission of the collaboration.

REFERENCES

- Abdullah M. H., Klypin A., Wilson G., 2020, *ApJ*, 901, 90
Abell G. O., 1958, *ApJS*, 3, 211

⁴<https://www.the300-project.org>

⁵https://cordis.europa.eu/project/rcn/207630_en.html

⁶<https://www.python.org>

- Adhikari S., Dalal N., Chamberlain R. T., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 019
- Aihara H. et al., 2011, *ApJS*, 193, 29
- Amoura Y., Drakos N. E., Berrouet A., Taylor J. E., 2021, *MNRAS*, 508, 100
- Bahé Y. M. et al., 2019, *MNRAS*, 485, 2287
- Baker W. M., Maiolino R., Bluck A. F. L., Lin L., Ellison S. L., Belfiore F., Pan H.-A., Thorp M., 2022, *MNRAS*, 510, 3622
- Beck A. M. et al., 2016, *MNRAS*, 455, 2110
- Behroozi P. et al., 2015, *MNRAS*, 454, 3020
- Bennett J. S., Sijacki D., 2022, *MNRAS*, 514, 313
- Biffi V., Dolag K., Reiprich T. H., Veronica A., Ramos-Ceja M. E., Bulbul E., Ota N., Ghirardini V., 2022, *A&A*, 661, A17
- Binney J., Tremaine S., 1987, *Galactic Dynamics*. Princeton Univ. Press, Princeton, NJ
- Bluck A. F. L., Maiolino R., Sánchez S. F., Ellison S. L., Thorp M. D., Piotrowska J. M., Teimoorinia H., Bundy K. A., 2020, *MNRAS*, 492, 96
- Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, *ApJ*, 555, 240
- Buote D. A., Tsai J. C., 1995, *ApJ*, 452, 522
- Campitiello M. G. et al., 2022, *A&A*, 665, A117
- Capalvo V., De Petris M., De Luca F., Cui W., Yepes G., Knebe A., Rasia E., 2021, *MNRAS*, 503, 6155
- Cerini G., Cappelluti N., Natarajan P., 2023, *ApJ*, 945, 152
- Chen H., Avestruz C., Kravtsov A. V., Lau E. T., Nagai D., 2019, *MNRAS*, 490, 2380
- Cialone G., De Petris M., Sembolini F., Yepes G., Baldi A. S., Rasia E., 2018, *MNRAS*, 477, 139
- Contreras-Santos A. et al., 2022, *MNRAS*, 511, 2897
- Cornwell D. J. et al., 2022, *MNRAS*, 517, 1678
- Cui W., Power C., Borgani S., Knebe A., Lewis G. F., Murante G., Poole G. B., 2017, *MNRAS*, 464, 2502
- Cui W. et al., 2018, *MNRAS*, 480, 2898
- Darragh Ford E. et al., 2019, *MNRAS*, 489, 5695
- De Luca F., De Petris M., Yepes G., Cui W., Knebe A., Rasia E., 2021, *MNRAS*, 504, 5383
- de Haan T. et al., 2016, *ApJ*, 832, 95
- Devriendt J. E. G., Guiderdoni B., Sadat R., 1999, *A&A*, 350, 381
- Diemer B., Kravtsov A. V., 2014, *ApJ*, 789, 1
- Eckert D., Ettori S., Robertson A., Massey R., Pointecouteau E., Harvey D., McCarthy I. G., 2022, *A&A*, 666, A41
- Evrard A. E., 1989, *ApJ*, 341, L71
- Fakhouri O., Ma C.-P., 2009, *MNRAS*, 394, 1825
- Ge C., Wang Q. D., Tripp T. M., Li Z., Gu Q., Ji L., 2016, *MNRAS*, 459, 366
- Gianfagna G. et al., 2021, *MNRAS*, 502, 5115
- Gill S. P. D., Knebe A., Gibson B. K., 2004, *MNRAS*, 351, 399
- Gouin C., Bonnaire T., Aghanim N., 2021, *A&A*, 651, A56
- Gunn J. E., Gott J. Richard L., 1972, *ApJ*, 176, 1
- Haggar R., Gray M. E., Pearce F. R., Knebe A., Cui W., Mostoghiu R., Yepes G., 2020, *MNRAS*, 492, 6074
- Haggar R., Pearce F. R., Gray M. E., Knebe A., Yepes G., 2021, *MNRAS*, 502, 1191
- Haggar R., Kuchner U., Gray M. E., Pearce F. R., Knebe A., Yepes G., Cui W., 2023, *MNRAS*, 518, 1316
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jeason-Daniel A., Dalla Vecchia C., Haas M. R., Schaye J., 2011, *MNRAS*, 415, L69
- Jolliffe I. T., Cadima J., 2016, *Phil. Trans. R. Soc. A*, 374, 20150202
- Kimmig L. C., Remus R.-S., Dolag K., Biffi V., 2023, *ApJ*, 949, 92
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Knebe A., Power C., Gill S. P. D., Gibson B. K., 2006, *MNRAS*, 368, 741
- Knebe A. et al., 2011, *MNRAS*, 415, 2293
- Knebe A. et al., 2013, *MNRAS*, 435, 1618
- Knebe A. et al., 2020, *MNRAS*, 495, 3002
- Knollmann S. R., Knebe A., 2009, *ApJS*, 182, 608
- Kravtsov A. V., Borgani S., 2012, *ARA&A*, 50, 353
- Kuchner U. et al., 2022, *MNRAS*, 510, 581
- Larson R. B., Tinsley B. M., Caldwell C. N., 1980, *ApJ*, 237, 692
- Lau E. T., Kravtsov A. V., Nagai D., 2009, *ApJ*, 705, 1129
- Lau E. T., Nagai D., Nelson K., 2013, *ApJ*, 777, 151
- Lau E. T., Nagai D., Avestruz C., Nelson K., Vikhlinin A., 2015, *ApJ*, 806, 68
- Lau E. T., Hearin A. P., Nagai D., Cappelluti N., 2021, *MNRAS*, 500, 1029
- Lawrance A. J., 1976, *Am. Stat.*, 30, 146
- Li Q., Han J., Wang W., Cui W., De Luca F., Yang X., Zhou Y., Shi R., 2022, *MNRAS*, 514, 5890
- Limousin M., Beauchesne B., Jullo E., 2022, *A&A*, 664, A90
- Lin Y.-T., Mohr J. J., 2004, *ApJ*, 617, 879
- Lourenço A. C. C. et al., 2023, *MNRAS*, 526, 4831
- Mauduit J. C., Mamon G. A., 2007, *A&A*, 475, 169
- McInnes L., Healy J., Melville J., 2018, preprint (arXiv:1802.03426)
- McKinney W., 2010, *Proceedings of the 9th Python in Science Conference*, 445, 51
- Mendoza I., Mansfield P., Wang K., Avestruz C., 2023, *MNRAS*, 523, 6386
- Mihos J. C., Harding P., Feldmeier J. J., Rudick C., Janowiecki S., Morrison H., Slater C., Watkins A., 2017, *ApJ*, 834, 16
- Mohr J. J., Fabricant D. G., Geller M. J., 1993, *ApJ*, 413, 492
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, *Nature*, 379, 613
- More S., Diemer B., Kravtsov A. V., 2015, *ApJ*, 810, 36
- Muldrew S. I., Pearce F. R., Power C., 2011, *MNRAS*, 410, 2617
- Murante G., Monaco P., Giovalini M., Borgani S., Diaferio A., 2010, *MNRAS*, 405, 1491
- Nagai D., Lau E. T., 2011, *ApJ*, 731, L10
- Nagai D., Vikhlinin A., Kravtsov A. V., 2007, *ApJ*, 655, 98
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Nelson D. et al., 2019, *Comput. Astrophys. Cosmol.*, 6, 2
- Neto A. F. et al., 2007, *MNRAS*, 381, 1450
- Parekh V., van der Heyden K., Ferrari C., Angus G., Holwerda B., 2015, *A&A*, 575, A127
- Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
- Peebles P. J. E., 1969, *ApJ*, 155, 393
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Planelles S., Quilis V., 2009, *MNRAS*, 399, 410
- Poole G. B., Fardal M. A., Babul A., McCarthy I. G., Quinn T., Wadsley J., 2006, *MNRAS*, 373, 881
- Power C., Knebe A., Knollmann S. R., 2012, *MNRAS*, 419, 1576
- Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, *MNRAS*, 423, 3018
- Ragagnin A., Dolag K., Moscardini L., Biviano A., D'Onofrio M., 2019, *MNRAS*, 486, 4001
- Rasia E., Meneghetti M., Ettori S., 2013, *Astron. Rev.*, 8, 40
- Rasia E. et al., 2015, *ApJ*, 813, L17
- Richardson T. R. G., Corasaniti P. S., 2022, *MNRAS*, 513, 4951
- Roberts I. D., Parker L. C., 2017, *MNRAS*, 467, 3268
- Santoni S., De Petris M., Ferragamo A., Yepes G., Cui W., 2024, *EPJ Web Conf.*, 293, 00048
- Santos J. S., Rosati P., Tozzi P., Böhringer H., Ettori S., Bignamini A., 2008, *A&A*, 483, 35
- Schade D., Lilly S. J., Crampton D., Hammer F., Le Fevre O., Tresse L., 1995, *ApJ*, 451, L1
- Sembolini F. et al., 2016, *MNRAS*, 457, 4063
- Shaw L. D., Weller J., Ostriker J. P., Bode P., 2006, *ApJ*, 646, 815
- Simonte M., Vazza F., Brighenti F., Brügggen M., Jones T. W., Angelinelli M., 2022, *A&A*, 658, A149
- Skibba R. A., Macciò A. V., 2011, *MNRAS*, 416, 2388
- Smith R. et al., 2023, *MNRAS*, 525, 4685
- Sohn J., Geller M. J., Diaferio A., Rines K. J., 2020, *ApJ*, 891, 129
- Sousbie T., 2011, *MNRAS*, 414, 350
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Srisawat C. et al., 2013, *MNRAS*, 436, 150
- Steinborn L. K., Dolag K., Hirschmann M., Prieto M. A., Remus R.-S., 2015, *MNRAS*, 448, 1504
- Tormen G., Bouchet F. R., White S. D. M., 1997, *MNRAS*, 286, 865
- Tornatore L., Borgani S., Dolag K., Matteucci F., 2007, *MNRAS*, 382, 1050
- Vallés-Pérez D., Planelles S., Quilis V., 2020, *MNRAS*, 499, 2303

- Vallés-Pérez D., Planelles S., Quilis V., 2021, *MNRAS*, 504, 510
 Vallés-Pérez D., Planelles S., Monllor-Berbegal Ó., Quilis V., 2023, *MNRAS*, 519, 6111
 van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
 Vazza F., Eckert D., Simionescu A., Brüggén M., Etori S., 2013, *MNRAS*, 429, 799
 Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
 Walker S. et al., 2019, *BAAS*, 51, 218
 Wang K., Mao Y.-Y., Zentner A. R., Lange J. U., van den Bosch F. C., Wechsler R. H., 2020, *MNRAS*, 498, 4450
 Wen Z. L., Han J. L., 2013, *MNRAS*, 436, 275
 Wong A. W. C., Taylor J. E., 2012, *ApJ*, 757, 102
 Yuan Z. S., Han J. L., 2020, *MNRAS*, 497, 5485
 Yuan Z. S., Han J. L., Wen Z. L., 2022, *MNRAS*, 513, 3013
 Zabel N. et al., 2019, *MNRAS*, 483, 2251
 Zenteno A. et al., 2020, *MNRAS*, 495, 705
 Zhang X., Simionescu A., Akamatsu H., Kaastra J. S., de Plaa J., van Weeren R. J., 2020, *A&A*, 642, A89
 Zhang B., Cui W., Wang Y., Dave R., De Petris M., 2022, *MNRAS*, 516, 26
 Zhang C. et al., 2024, *MNRAS*, 530, 4234
 Zhoolideh Haghighi M. H., Raouf M., Khosroshahi H. G., Farhang A., Gozaliasl G., 2020, *ApJ*, 904, 36

APPENDIX A: EXPLAINED VARIANCE OF PRINCIPAL COMPONENTS

The four principal components selected and analysed throughout this work collectively describe 64 per cent of the total variance of the 17 dynamical state indicators described in Section 2.2.1. In Fig. A1, we show the proportion of the variance explained by all 17 of the principal components, and the cumulative explained variance.

Clearly, PC1 is the dominant component, explaining 38 per cent of the total variance alone. Other than this, there is not a clearly visible ‘cut-off’ point, beyond which parameters are far less important. Consequently, the decision to keep only the first four principal components is not obviously mathematically justifiable. This choice was made in order to select the minimum number of components such that each of the dynamical state indicators contributed strongly to at least one of them, to provide an idea of the number of dimensions along which dynamical state varies. For example, PC5 (which explains 5 per cent of the total variance) is dominated by

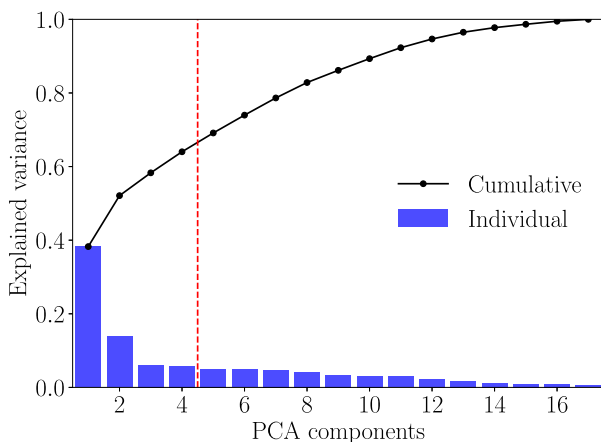


Figure A1. Cumulative proportion of variance of the dynamical state indicators data set explained by the ordered principal components (black line and points), and by the individual components (blue bars). Throughout our analysis, we only consider the first four components, indicated by the vertical red dashed line.

Table A1. Similar to Table 1, but showing coordinate values for PC5, which is excluded throughout this work. Component coordinates with an absolute value greater than 0.24 are highlighted in bold; the three parameters that contribute strongly to PC5 (λ , N_{fil} , and $D_{1,0.1}$) also contribute strongly to PC3, albeit with different weights.

Parameter	Contribution to PC5
$f_s(R_{200})$	-0.04
$f_s(R_{500})$	0.05
$\Delta_r(R_{200})$	-0.02
$\Delta_r(R_{500})$	0.02
$\eta(R_{200})$	-0.14
$\eta(R_{500})$	-0.13
$z_{0.5}$	-0.04
λ	-0.40
c/a	-0.23
c_{200}	0.08
$z_{\text{merge},50}$	-0.14
γ	0.09
N_{fil}	-0.55
$D_{1,0.1}$	0.59
N_{200}	-0.24
m_{12}	-0.05
σ_{BCG}	0.05

λ , N_{fil} , and $D_{1,0.1}$, which all also strongly contribute to PC3 (see Table A1). We interpret this as PC5 providing a ‘second-order’ correction to the environment of a cluster, as described by PC3; we believe that including such a component would not meaningfully add to this work.

APPENDIX B: CORRELATIONS BETWEEN DYNAMICAL STATE INDICATORS

In Fig. 3 we show the Spearman’s rank correlation coefficient between each of the 3D dynamical state indicators used in this study (described in Section 2.2.1). In Fig. B1, we explicitly show the correlations between these measures, from which these correlation coefficients were calculated. As in Fig. 3, we also group these parameters based on the principal component to which they most strongly contribute. Finally, we also include the cluster mass as an additional parameter here, to explicitly show the mass dependence of these quantities.

It should be noted that several of these parameters have a highly skewed distribution (for example, η). These parameters were all standardized for the PCA and UMAP analysis in this work to approximate a normal distribution with a mean of zero and a unit standard deviation, but here we show the raw distributions. Most of these parameters also display either no dependence, or a weak dependence, on cluster mass ($|\rho_s| \leq 0.15$, $p \geq 0.005$). The exception is the velocity dispersion of the BCG ($\rho_s = 0.74$), which has a fairly strong positive correlation with the cluster mass. This result is expected, as numerous previous studies have found that the BCG velocity dispersion scales with the mass of its host cluster (e.g. Sohn et al. 2020). One could account for this mass dependence by normalizing the BCG velocity dispersion by the maximum circular orbital speed around the cluster, v_{circ} . For THE300, the Spearman’s rank correlation coefficient between M_{200} and $\sigma_{\text{BCG}}/v_{\text{circ}}$ is far reduced ($\rho_s = 0.19$, $p = 6 \times 10^{-4}$); although we have chosen to just use σ_{BCG} in this study, this ratio is a potential alternative choice that one could use instead.

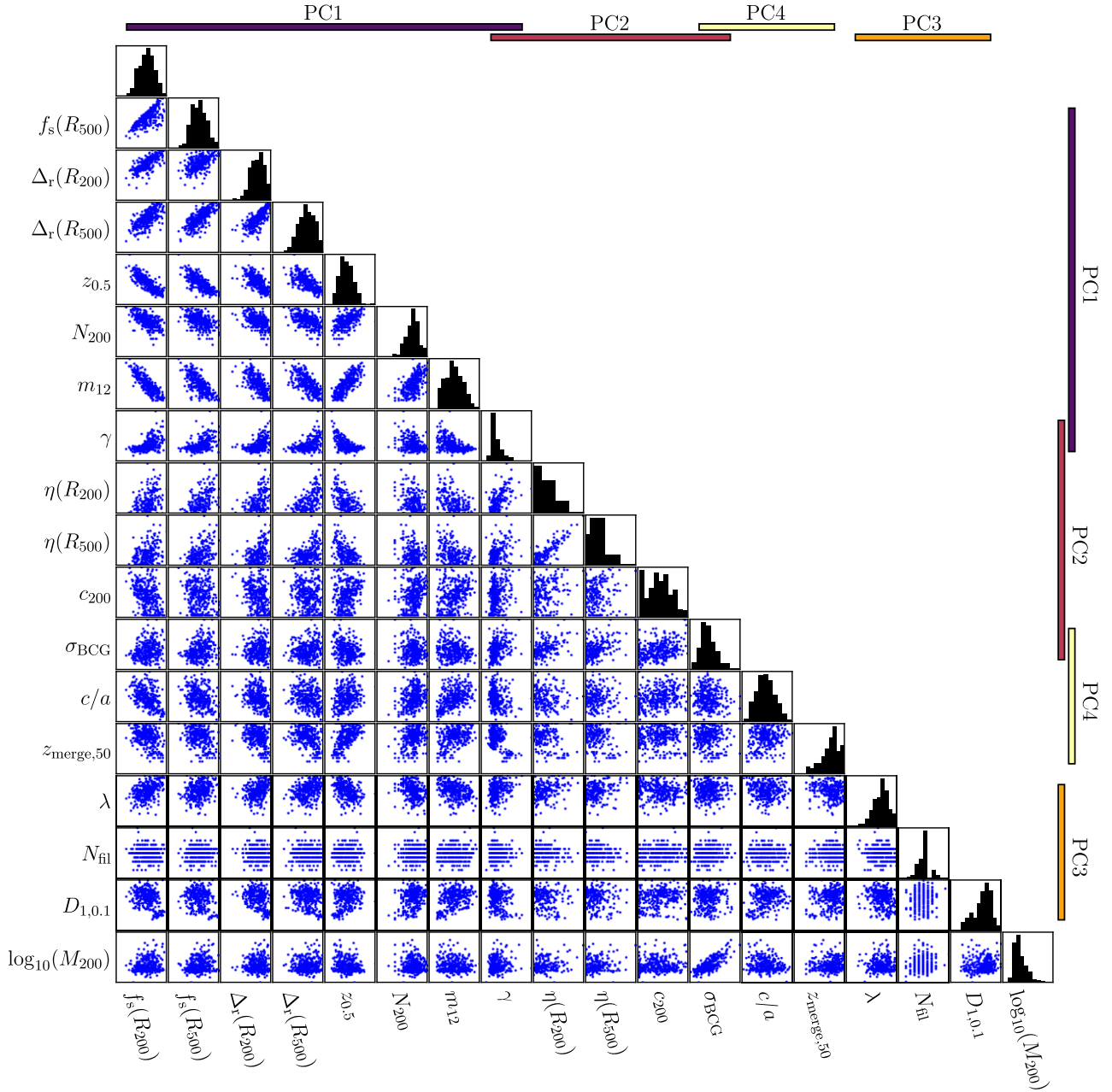


Figure B1. Corner plot showing the correlations between the 17 dynamical state indicators described in Section 2.2.1. The Spearman's rank correlation coefficients displayed in Fig. 3 are calculated from these scatter plots. This figure also includes histograms showing the distribution of these parameters. We also include the logarithmic mass of each galaxy cluster as an additional quantity here, to show any mass dependence of these 17 parameters.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.