

Satellite quenching time-scales in clusters from projected phase space measurements matched to simulated orbits

Kyle A. Oman^{1,2★} and Michael J. Hudson^{2,3}

¹*Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada*

²*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada*

³*Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada*

Accepted 2016 August 29. Received 2016 July 26; in original form 2015 December 22

ABSTRACT

We measure the star formation quenching efficiency and time-scale in cluster environments. Our method uses N -body simulations to estimate the probability distribution of possible orbits for a sample of observed Sloan Digital Sky Survey galaxies in and around clusters based on their position and velocity offsets from their host cluster. We study the relationship between their star formation rates and their likely orbital histories via a simple model in which star formation is quenched once a delay time after infall has elapsed. Our orbit library method is designed to isolate the environmental effect on the star formation rate due to a galaxy's present-day host cluster from 'pre-processing' in previous group hosts. We find that quenching of satellite galaxies of all stellar masses in our sample (10^9 – $10^{11.5} M_{\odot}$) by massive ($>10^{13} M_{\odot}$) clusters is essentially 100 per cent efficient. Our fits show that all galaxies quench on their first infall, approximately at or within a Gyr of their first pericentric passage. There is little variation in the onset of quenching from galaxy-to-galaxy: the spread in this time is at most ~ 2 Gyr at fixed M_* . Higher mass satellites quench earlier, with very little dependence on host cluster mass in the range probed by our sample.

Key words: galaxies: clusters: general – galaxies: evolution.

1 INTRODUCTION

A detailed understanding of the mechanisms that quench star formation in galaxies remains elusive. It now seems clear that quenching is strongly correlated with an 'internal' parameter that is closely related to galaxy mass: stellar mass (Kauffmann et al. 2003; Baldry et al. 2004), velocity dispersion (Graves, Faber & Schiavon 2009; Smith, Lucey & Hudson 2009), or structural properties, such as the central stellar surface mass density (Cheung et al. 2012; Fang et al. 2013) or the bulge fraction (Bluck et al. 2014; Omand, Balogh & Poggianti 2014). The physical cause of this quenching is still not known, although AGN (Granato et al. 2004; Bower et al. 2006; Croton et al. 2006) and/or mergers (Hopkins et al. 2006) and/or disc instabilities (Dekel, Sari & Ceverino 2009) are often cited. It has become clear that environment also plays a role: once a galaxy falls into a more massive halo (such as a group or cluster) and becomes a satellite, there is an additional probability of quenching over and above the stronger 'mass-related' quenching (Balogh et al. 2004; van den Bosch et al. 2008; Peng et al. 2010). It is this latter 'satellite quenching' that is the subject of this paper.

An infalling, actively star-forming satellite galaxy might be affected by its host halo in several ways, as reviewed by Boselli & Gavazzi (2006). Ram pressure stripping may remove cold gas from the disc (Gunn & Gott 1972). It has long been known that cluster galaxies are H I-deficient (e.g. Giovanelli & Haynes 1985). Furthermore, ram pressure stripping has been observed in individual infalling galaxies in a number of nearby clusters. In particular, in the Coma cluster at least 40 per cent of blue galaxies within 500 kpc of the centre have young stars formed from stripped material visible at ultraviolet wavelengths (Smith et al. 2010), which suggests that ram pressure stripping is ubiquitous. However, from these snapshots it is difficult to tell how rapid this process is, and how effective it is overall. A closely related physical process is 'strangulation' (Larson, Tinsley & Caldwell 1980; Balogh, Navarro & Morris 2000) in which the hot gas halo is stripped by ram pressure, thus removing the source that would otherwise have replenished the cold gas in the disc. Because the cold gas is not immediately affected, the time-scale for strangulation should be longer than ram pressure stripping of the cold gas disc.

One way to constrain the quenching mechanism(s) is by measuring how effective quenching is, where or when it first occurs in the satellite's orbit, and how long a satellite takes to quench. Environment has a role in regulating star formation on a range of host mass scales (e.g. Peng et al. 2010). In this paper we focus

* E-mail: koman@uvic.ca

on galaxy clusters. These lend themselves well to our methodology for two reasons. First, because their centres and extents, in the sense of both position and velocity, are much better defined than in poorer galaxy groups (though centring becomes easier again for systems such as the Milky Way and its satellites). Secondly, our choice to study relatively large satellites around massive clusters means we are able to draw our sample of observed galaxies from a large volume, yielding a statistically powerful data set. While clusters host a relatively small fraction of the passive galaxy population – only about 15 per cent (2.5 per cent) of red galaxies with $M_* \geq 10^9 M_\odot$ are satellites in $\geq 10^{13} M_\odot$ ($10^{14} M_\odot$) haloes (estimated using the galaxy stellar mass function of red galaxies from Baldry et al. 2012 and the satellite fraction and host halo mass distribution of van den Bosch et al. 2008) – they offer a useful proving ground for analysis techniques aimed at constraining the time-scale(s) of the quenching process(es) before attempting to tackle the more difficult galaxy group scale. In addition, it may be that a single environmental quenching mechanism is dominant in host haloes of all masses (see for instance van den Bosch et al. 2008, though Fillingham et al. 2015 argue the opposite). In this case the study of quenching in clusters can directly inform more difficult studies of lower mass hosts.

Some early work involved comparing semi-analytic models to observations. In models where the quenching occurs quickly after crossing the host halo’s virial radius, too many red satellites were produced. The disagreement suggests that quenching process had to be slow (Weinmann et al. 2006; Font et al. 2008; Balogh et al. 2009; Kimm et al. 2009; Weinmann et al. 2010). Hudson et al. (2010) showed that, while bulge colours do not depend on cluster-centric radius, the colours of discs are redder closer to the cluster centre. These results were modelled by Taranu et al. (2014), who found that star formation in discs declines with an exponential time-scale of ~ 3 Gyr, starting at cluster infall.

Wetzel et al. (2013, hereafter W13) concluded that satellites are quenched on time-scale of 2–6 Gyr after passing the virial radius of a larger host halo for the first time. W13 obtained this result by measuring quenched fractions of satellites and centrals in low-redshift Sloan Digital Sky Survey (SDSS) groups and clusters, and comparing these data with a satellite quenching model based on a halo infall time distribution from N -body simulations combined with an empirically calibrated model of the quenched fractions at higher redshifts. Similar results to those of W13 were obtained by Hirschmann et al. (2014). At higher redshifts ($z \sim 1$), Mok et al. (2014) found shorter time-scales of the order of 1 Gyr.

An alternative approach is to take advantage of galaxies’ positions in the observational projected phase space (PPS) of separation in the plane of the sky and line-of-sight velocity. Gill, Knebe & Gibson (2005) showed that, at the same projected radius, galaxies in different phases of their orbits have different kinematics. This was extended by Oman, Hudson & Behroozi (2013), who constructed a subhalo orbit library that allowed them to construct a detailed probabilistic mapping between position in PPS and subhalo infall time.

Mahajan, Mamon & Raychaudhury (2011) were the first to de-project the PPS to obtain constraints on star formation histories of galaxies falling into larger systems. They studied galaxies with recent (within 1–3 Gyr) or ongoing star formation, and concluded that star formation is efficiently quenched in a single passage through the cluster. Other authors have used the PPS to understand quenching at high redshift (Muzzin et al. 2014) or the effects of ram pressure stripping (Hernández-Fernández et al. 2014; Jaffé et al. 2015).

The aim of this paper is to study the star formation rates (SFRs) of galaxies based on their location in PPS, and model these using the orbit libraries of Oman et al. (2013). Whereas Mahajan et al. (2011) used a coarse binning of the populations (‘virial’, ‘back-splash’, ‘infalling’), in this paper we use detailed orbit libraries drawn from N -body simulations. Our model consists of two components: (1) an infalling population of galaxies (which are observed predominantly outside the virial radius and are assumed to have some ‘pre-processed’ quenched fraction) and (2) a simple model for quenching in which some fraction of the active infalling galaxies are quenched following a delay Δt after passing $2.5 r_{\text{vir}}$. The model then predicts the quenched fraction at any point in PPS. The infalling quenched fraction is fit from the PPS data simultaneously with the free parameters of the model (the efficiency of quenching and the time-scale of quenching). This allows us to account for ‘pre-processing’ in a natural way, and hence our results isolate the physical effects of infall of active satellites into their *current* cluster-mass ($\sim 10^{14.5} M_\odot$) host haloes. This differs from the approach of W13, in which the quenching time-scale refers to the time since infall into *any* halo, and so includes processing in the current host halo plus ‘pre-processing’ in host haloes of lower mass.

This paper is structured as follows: in Section 2, we describe our numerical and observed data samples. In Section 3, we describe our models and fitting method. In Section 4, we present the results of fitting our models to the observed data. We discuss our results and compare to other work in Section 5 and summarize in Section 6.

We assume the same cosmology used in the Bolshoi and Multi-dark Run 1 simulations with $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $\Omega_b = 0.0469$, $n_s = 0.95$, $h_0 = 0.70$, $\sigma_8 = 0.82$ (Prada et al. 2012).

2 DATA

2.1 Numerical simulations

We use the output of the Multidark Run 1 (MDR1) dark matter-only cosmological simulation to obtain a large sample of satellite orbits. The simulation has a box side length of $1 h^{-1}$ Gpc, 2048^3 particles, $8.63 \times 10^9 h^{-1} M_\odot$ mass resolution, $7 h^{-1}$ kpc force resolution, and uses the *WMAP* 7 cosmology. The simulation runs from redshift $z = 65$ to 0 and has outputs linearly spaced¹ in scalefactor a . The time resolution at $z = 0$ is of about 0.21 Gyr. For further details regarding MDR1, we refer to Prada et al. (2012). The simulation output was processed with the *ROCKSTAR* halo finder (Behroozi, Wechsler & Wu 2013a) and the merger tree code of Behroozi et al. (2013b). In order to use host-satellite linking in the merger tree as a proxy for cluster membership out to the largest apocentric radii of about $2.5 r_{\text{vir}}$ (Balogh et al. 2000; Mamon et al. 2004; Gill et al. 2005; Ludlow et al. 2009), we modified the merger tree code to create these links at distances of up to $2.5 r_{\text{vir}}$ (rather than the default $1.0 r_{\text{vir}}$).

2.2 Coordinates

We distinguish between two sets of cluster-centric coordinates: (r, v) the full ‘6D’ phase space coordinates, and (R, V) the projected coordinates consisting of the line-of-sight component of the velocity and the distance to the centre perpendicular to the line of sight. We arbitrarily adopt the third (z)-axis of the simulation as our

¹ The resolution in scalefactor doubles after $a \sim 0.7$. There are also a handful of irregularly spaced steps at small a .

projection axis. The radial projection between two points labelled 1 and 2 can be expressed as

$$R_{12} = \sqrt{(r_{2,x} - r_{1,x})^2 + (r_{2,y} - r_{1,y})^2}. \quad (1)$$

The projected velocity includes a correction for the Hubble flow, allowing projected coordinates from the simulation to be directly compared to observed line-of-sight velocity offsets.

$$V_{12} = |(v_{2,z} - v_{1,z}) + H(r_{2,z} - r_{1,z})|. \quad (2)$$

The absolute value encodes our assumption that observationally the distances of clusters and their satellites are not measured with sufficient accuracy to determine the sign of their relative velocity.

To facilitate comparison between clusters, both observed and simulated, we normalize all radial coordinates by the virial radius of the cluster r_{vir} , defined using the formula of Bryan & Norman (1998): the region enclosing an overdensity 360 times the background density at $z = 0$. For those accustomed to a definition in terms of the critical density, an approximate conversion valid at $z = 0$ is $r_{200c}/r_{\text{vir}} \sim 0.73$. We normalize velocity coordinates by the 3D velocity dispersion of the cluster, σ_{3D} . We assume that clusters are approximately spherically symmetric so that the observable 1D velocity dispersion σ_{1D} is $\sigma_{3D} \sim \sqrt{3}\sigma_{1D}$.

2.3 Observational sample of clusters

To obtain a large sample of clusters and their satellites, we use the cluster catalogue of von der Linden et al. (2007). This provides the right ascension, declination, redshift (z) and velocity dispersion (σ_{1D}) of 625 clusters. In Fig. 1 we show the cluster halo mass distribution of our sample, estimated from a relationship between σ_{1D} and halo mass calibrated with our simulation sample (Section 2.1):

$$\sigma_{1D}/9.9 \times 10^{-3} \text{ km s}^{-1} = (M_{\text{vir}}/M_{\odot})^{0.33}. \quad (3)$$

Our sample of satellites is drawn from the SDSS DR7 (Abazajian et al. 2009), supplemented with SFRs (Brinchmann et al. 2004; Salim et al. 2007) and stellar masses (Mendel et al. 2014). We

select from the galaxies in this catalogue those with spectroscopic measurements – and thus more reliable redshifts – of which there are 562 076.

We construct our sample of observed satellites as similarly as possible to our sample of simulated satellites. The virial radius of each cluster is estimated from its virial mass:

$$M_{\text{vir}} = \frac{4}{3} \pi r_{\text{vir}}^3 (360 \Omega_m \rho_{\text{crit}}). \quad (4)$$

A galaxy is flagged as a satellite candidate of a cluster if it is within $2.5 r_{\text{vir}}$ of the cluster centre and its line of sight (LoS) velocity offset $|\Delta v_{\text{LoS}}|$ from the cluster is less than $2.0 \sigma_{3D}$. The brightest cluster galaxies (BCGs) are included in the satellite population in our analysis. We note that by construction, the BCGs have coordinates $(R, V) = (0, 0)$ in their respective clusters since von der Linden et al. (2007) define the cluster centres in their catalogue as the location of the BCG. This assumes that the BCG is hosted by the cluster halo rather than a satellite halo, and no corresponding orbits would appear in our simulated orbit catalogues. The BCGs account for less than 2 per cent of our satellite sample and so we do not expect their presence in the observed sample to impact our conclusions.

The projected radius of the satellites is determined from their angular separation from their host cluster centre $\Delta\theta$:

$$\frac{R}{r_{\text{vir}}} = \frac{d_A \Delta\theta}{r_{\text{vir}}}, \quad (5)$$

where d_A is the angular diameter distance of the cluster. The velocity offset is calculated from the redshift offset:

$$\frac{V}{\sigma_{3D}} = \frac{c|z_g - z_c|}{(1 + z_c)\sqrt{3}\sigma_{1D}}, \quad (6)$$

where z_g is the redshift of the galaxy and z_c is the redshift of the cluster.

This process of associating galaxies to clusters yields a sample of 44 436 satellite candidates, of which we expect about half to be interlopers (see Section 3.2). The distribution of satellite stellar masses is shown in Fig. 1. We estimate the halo mass of the satellites, based on the method of Velander et al. (2014), and the conversion

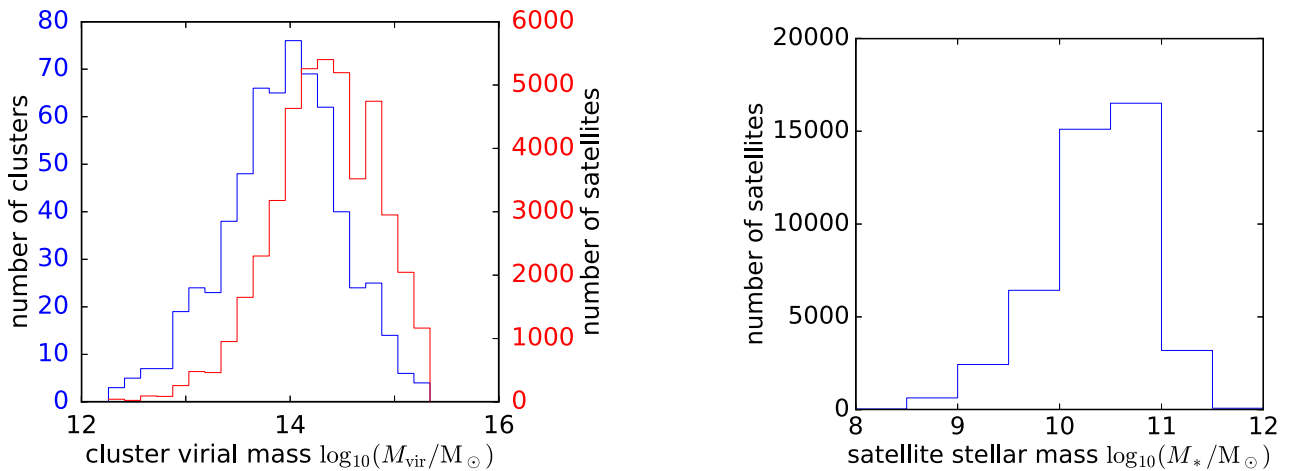


Figure 1. Left-hand panel: distribution of host cluster virial masses (dotted blue line) of the observed sample, inferred from the cluster velocity dispersion. Our numerical orbit sample is for hosts of $10^{13} < M_{\text{vir}}/M_{\odot} < 10^{15}$. The two data sets are well matched, given that the orbit sample is only very weakly sensitive to host halo mass (see Oman et al. 2013, and Figs 8, 9 and related discussion below). Also shown is the distribution of satellite galaxies as a function of the cluster mass they occupy (solid red line). Right-hand panel: distribution of satellite candidate stellar masses in the observed sample. Our numerical orbit sample cuts satellite haloes of less than $10^{11.9} M_{\odot}$. An estimate of the halo masses of the satellite candidates in the observed sample using the method of Velander et al. (2014) indicates a good match with this mass cut.

from M_{200c} to M_{vir} of Bullock et al. (2001), as

$$\log_{10} (M_{\text{vir}}/M_{\odot}) = 0.84 \log_{10} (M_{*}/M_{\odot}) + 3.09. \quad (7)$$

We estimate a systematic error in this conversion of up to 15 per cent, and a scatter of up to 0.4 dex.

3 METHOD

3.1 Orbit libraries

We use the same method to construct orbit libraries as Oman et al. (2013), which we summarize here. We define clusters as haloes of $>10^{13} M_{\odot}$ at $z=0$, of which MDR1 has $\sim 550\,000$. Satellites are within $2.5 r_{\text{vir}}$ of a cluster at $z=0$, and have a mass $>10^{11.9} M_{\odot}$ at infall such that they are well resolved and minimally sensitive to artificial disruption (Klypin et al. 1999; Kitzbichler & White 2008). We track the satellites back in time, noting the time of infall which we define as entry of the satellite within $2.5 r_{\text{vir}}$. This somewhat unconventional definition of infall has the advantage that satellites nearly never orbit back out past this radius. A satellite on a typical orbit takes 2.5–3.0 Gyr after infall to reach the virial radius, and pericentre occurs 3.5–4.0 Gyr after infall. Our final sample numbers $\sim 1\,200\,000$ satellite orbits.

3.2 Interlopers

An observed sample of cluster satellites is typically selected within some projected radius and velocity offset from the cluster centre. This defines a cylinder² in PPS. For reasonable selection cuts, this cylinder encloses a sphere centred around the cluster which contains the satellite galaxies of interest, but also a volume outside this sphere containing galaxies ‘projected into’ the cluster, which we term ‘interlopers’. Some of these interlopers will likely someday fall into the cluster, while others may eventually move off into a neighbouring structure. We supplement our simulated orbit catalogue with a sample of interlopers, allowing a fair comparison with our sample of observed cluster satellite candidates. We select all haloes with projected coordinates (projection along the z -axis of the simulation box) $\frac{R}{r_{\text{vir}}} < 2.5$, $\frac{V}{\sigma_{3D}} < 2.0$ that also have non-projected radius $\frac{r}{r_{\text{vir}}} > 2.5$. We apply the same mass cut as for our simulated satellite sample, yielding $\sim 1\,500\,000$ interlopers. The fraction of interlopers (compared to actual satellites) is a function of position in PPS, with interloper fraction increasing with increasing R and V (see Fig. 2).

From our catalogue of satellite orbits and interlopers, we construct a probability density of infall times for each position in PPS. Example probability density functions (PDFs) for a selection of points in the (R, V) plane (see boxes in Fig. 2) are shown in Fig. 3.

3.3 Comparison of observed and simulated samples

The observed and simulated catalogues of satellite candidates (i.e. including interlopers in both cases) are generally well matched. The distribution of observed cluster virial masses (Fig. 1) is similar to our cutoff of $>10^{13} M_{\odot}$ in the simulations, and the distribution of infall times as a function of (R, V) coordinates is only weakly sensitive to host mass (Oman et al. 2013, and Figs 8, 9 and related discussion below). The satellite halo mass distributions are

² More accurately the shape is that of a cone with its peak sliced off, but a cylinder is a good approximation for distant clusters.

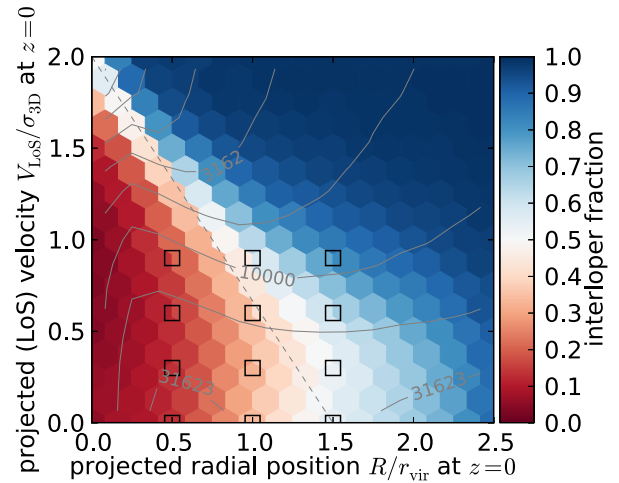


Figure 2. Fraction of ‘interlopers’ – satellite haloes which appear in a cluster in projection with $R < 2.5 r_{\text{vir}}$ and $V < 2.0 \sigma_{3D}$, but fall outside in 3D, that is $r > 2.5 r_{\text{vir}}$ – as a function of position in PPS. The halo population at small projected radii and velocities is dominated by bona-fide cluster members, while at large projected radii and velocities the interloper fraction tends to 1.0. The dashed line marks $\frac{V}{\sigma_{3D}} = -\frac{4}{3} \frac{R}{r_{\text{vir}}} + 2$ and approximately divides the two regions. Contours indicate the number of haloes in each bin. The black squares correspond to the locations of the subsamples shown in the panels in Fig. 3.

also well matched, with the offset between the observed and simulated samples being comparable to the estimated systematic error in the conversion from stellar mass to halo mass. Finally, the PPS distributions of the simulated and observed haloes are in excellent agreement, as shown in Fig. 4.

3.4 Quenching model

In this section, we briefly derive the likelihood function of our simple quenching model. We assume that when quenching occurs, individual galaxies transition rapidly from an active, star-forming state to a passive, quenched state. This is motivated by the bimodal distribution of specific star formation rate (SSFR, see for instance equation 13), also apparent as the ‘green valley’ in the colour distribution of galaxies (Kauffmann et al. 2003; Baldry et al. 2004). We assume that two galaxy populations exist, one inside and the other outside (but nearby) massive clusters, each with distinct passive fractions. We fit two time-scales: the first is a delay between cluster infall and the onset of quenching, the second is the time-scale for the transition between the quenched fraction of the population just outside the cluster to that inside.

The likelihood \mathcal{L} is defined in terms of a sum over probabilities P_i where i is an index running over all galaxies in the observed sample. Each galaxy has three properties of interest in the context of calculating the likelihood: a projected radius from its host cluster R_i , a line-of-sight velocity offset from the host cluster velocity V_i and a specific star formation rate SSFR_i ,

$$\ln \mathcal{L} = \sum_i \ln P_i. \quad (8)$$

If a galaxy is observed to be passive, with $\text{SSFR}_i < \text{SSFR}_{\text{cut}}(M_*)$ (we adopt a cut in SSFR between active and passive galaxies that depends on stellar mass, see Section 3.5 and Fig. 5), then P_i is the probability according to the model that the galaxy is passive p_{passive} . Conversely, if the observed galaxy is active, P_i is the probability

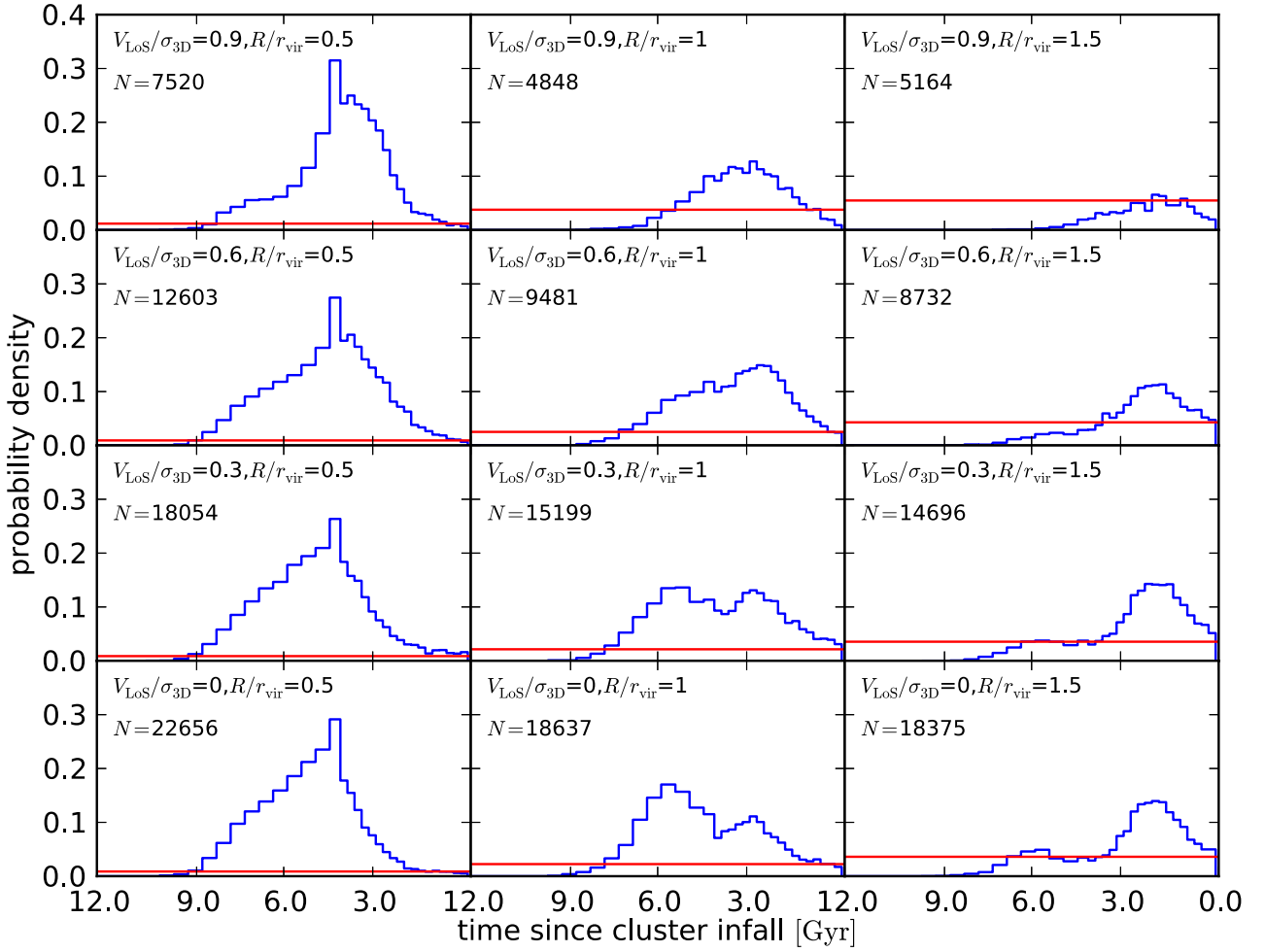


Figure 3. PDFs of time since cluster infall for a selection of points in PPS. The panels correspond to the locations marked by black squares in Fig. 2. The red line corresponds to the interloper probability: the ratio of the integrals of the red and blue curves is equal to the interloper fraction. The number of orbits used to construct each histogram is labelled N , giving a measure of the statistics of each PDF.

according to the model that the galaxy is active ($1 - p_{\text{passive}}$),

$$P_i = \begin{cases} p_{\text{passive},i}, & \text{SSFR}_i \leq \text{SSFR}_{\text{cut}}(M_*) \\ 1 - p_{\text{passive},i}, & \text{SSFR}_i > \text{SSFR}_{\text{cut}}(M_*), \end{cases} \quad (9)$$

$p_{\text{passive},i}$ is in turn defined in terms of the passive fraction outside the cluster $f_{\text{passive},\text{out}}$ and inside the cluster $f_{\text{passive},\text{in}}$, which are related by $\Delta f_{\text{passive}} = f_{\text{passive},\text{in}} - f_{\text{passive},\text{out}}$, and the probability $p_{q,i}$ that the cluster has quenched the galaxy,

$$p_{\text{passive},i} = f_{\text{passive},\text{out}} + p_{q,i} \Delta f_{\text{passive}}, \quad (10)$$

where $p_{q,i}$ is expressed as the integral of the product of two probabilities. The first is the probability $p_q(t)$ that at time t after infall the cluster has quenched the satellite. The functional form of $p_q(t)$ is an input of the model. The second is the probability $p_{\text{infall},i}(R_i, V_i, t)$ that the galaxy has a time since infall of t ,

$$p_{q,i} = \int_{t=0}^{t_f} p_q(t) p_{\text{infall},i}(R_i, V_i, t) dt, \quad (11)$$

where t_f is the age of the Universe $p_{\text{infall},i}$ is extracted from our orbit libraries, as illustrated in Fig. 3.

We choose $p_q(t)$ to represent a scenario where a galaxy is unaffected after infall until a time Δt has elapsed, then has an increasing probability of being quenched, parametrized by a time-scale τ . The

interpretation of τ is as a scatter in the quenching time of a *population* of satellites, rather than the time taken for an individual galaxy to transition from active to passive. We assume that this transition occurs rapidly, motivated by the bimodality in the SSFR distribution (Fig. 5; see also Wetzel, Tinker & Conroy 2012), and so make no attempt to model the full SSFR distribution at each point in PPS, instead reducing this distribution to a single parameter, the passive fraction

$$p_q(t) = \begin{cases} 0, & t \leq \Delta t \\ 1 - e^{-(t-\Delta t)/\tau}, & t > \Delta t \end{cases} \quad (12)$$

The model described above has four free parameters to be fit: Δt , τ , $f_{\text{passive},\text{out}}$ and $\Delta f_{\text{passive}}$. When discussing our results below, we often quote $f_{\text{passive},\text{in}}$ instead $\Delta f_{\text{passive}}$, which is equivalent, but, we feel, more intuitive. We also consider a single ‘combined’ time-scale $t_{1/2} = \Delta t + 0.69\tau$, the time when half of the galaxies that will be quenched by their host have become passive. In order to derive formal errors on this parameter, we perform additional fits where this parameter is fit directly (replacing Δt as a parameter to be fit).

We adopt flat priors on all parameters in the intervals $0 \leq f_{\text{passive},\text{in}} \leq 1$, $0 \leq f_{\text{passive},\text{out}} \leq 1$, $\Delta t \geq 0$ Gyr, $\tau \geq 0$ Gyr. In cases where we fit $t_{1/2}$ instead of Δt , we constrain $t_{1/2} > 0$ Gyr.

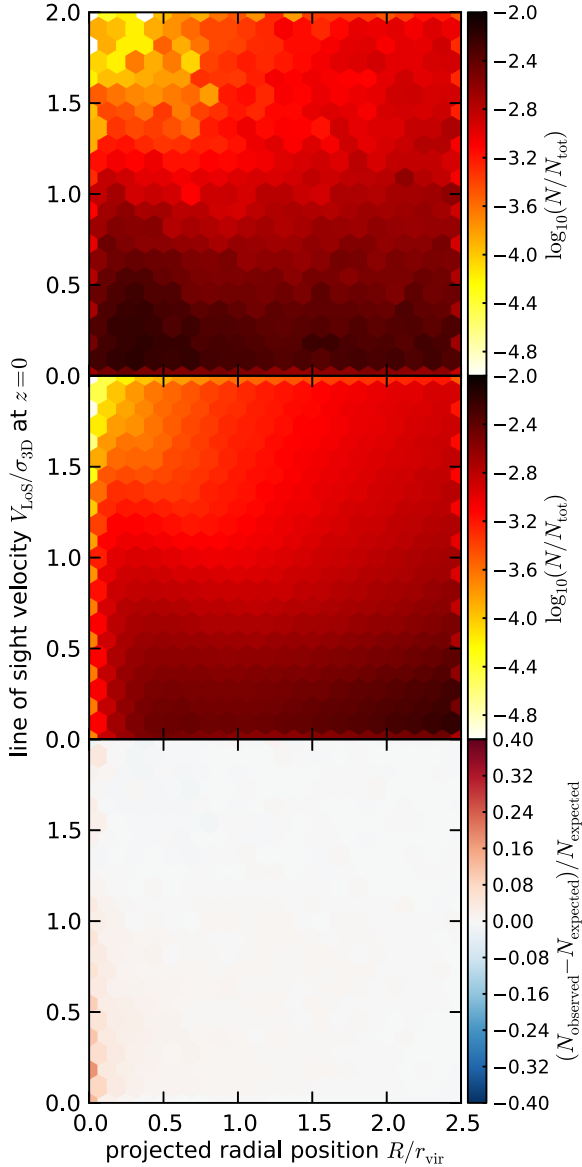


Figure 4. Upper panel: distribution of observed galaxy sample in PPS. Middle panel: distribution of simulated halo sample in PPS. The normalization of the two upper panels is such that the colour scales are directly comparable. Lower panel: fractional excess of observed galaxies relative to expectation from simulation halo counts as a function of PPS position. Most regions are limited to variations of a few per cent. The excess of observed galaxies at low R and V is consistent with the amount of artificial disruption of simulated haloes near the centres of clusters (Klypin et al. 1999; Kitzbichler & White 2008); a smaller proportion of observed galaxies are also absent due to fibre collisions in the SDSS.

3.5 Definition of ‘active’ and ‘passive’

We split our sample of cluster satellite candidates into an active and a passive population based on their SSFRs. The distribution of SSFRs shows a clear bimodality, and the relative size of the two populations is a function of PPS coordinates (see the upper-left panel of Fig. 6). The relative size of the passive population increases with decreasing R and V .

Whether a galaxy with a given SSFR should be classified as active or passive depends on its stellar mass M_* . Fig. 5 shows the distribution of SSFRs as a function of M_* for our sample of

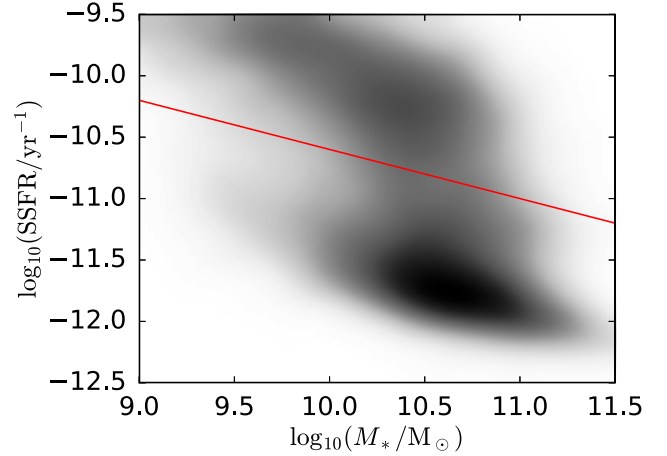


Figure 5. SSFR as a function of stellar mass for our sample of galaxies. The colour scale is logarithmic and represents the density of galaxies in this plane. The passive ‘red sequence’ and active ‘blue cloud’ populations are clearly visible. We separate the two ‘by eye’ with a line (see equation 13) to define our $\text{SSFR}_{\text{cut}}(M_*)$.

galaxies. The separation between the active and passive populations was determined ‘by eye’ and is illustrated by the red line at

$$\log_{10}(\text{SSFR}_{\text{cut}}/\text{yr}^{-1}) = -\frac{\log_{10}(M_*/M_{\odot})}{2.5} - 6.6. \quad (13)$$

4 RESULTS

In this section, we fit our simple quenching model to the observed sample of satellite galaxies. We split the observed data sample into two bins of host halo mass, 10^{13} – 10^{14} and 10^{14} – $10^{15} M_{\odot}$, and five bins of satellite stellar mass, evenly spaced in $\log_{10}(M_*/M_{\odot})$ between 9 and 11.5. For each of the 10 subsamples we produce an infall time PDF from our orbit libraries (including interlopers) using only orbits consistent with the host and satellite mass corresponding to each bin.

A visual depiction of our model (see Section 3.4), illustrated using the $14 < \log_{10}(M_{\text{host}}/M_{\odot}) < 15$ and $9.5 < \log_{10}(M_*/M_{\odot}) < 10$ subsample, is presented in Fig. 6. Given the four parameters $f_{\text{passive,out}}$, $f_{\text{passive,in}}$, Δt and τ , the model predicts the observable f_{passive} as a function of position in PPS. In the middle row of panels in Fig. 6, the leftmost panel shows the prediction of the model for our best-fitting parameter values. The next four panels show the effect on the prediction of changing individual parameters. The observed distribution of f_{passive} in the (R, V) plane is shown in the upper-left panel, and the density of galaxies w as a function of position in the (R, V) plane, relative to the maximum density, is shown in the upper-right panel. The likelihood of a given set of parameter values reflects the comparison between the model prediction and the observed data. The third row shows a visualization of this comparison: darker colour is used where there is a larger discrepancy between model and data, weighted by w .

The time resolution of the MDR1 simulation snapshots imposes a limit on the precision with which we can measure Δt , so we only evaluate \mathcal{L} for values of Δt spaced to match the simulation time resolution. The discrete nature of Δt makes a simultaneous multidimensional maximum likelihood search (e.g. Markov Chain Monte Carlo) somewhat unwieldy. Instead, we use an iterative grid search strategy, beginning with a coarse four-dimensional grid covering a wide area of our parameter space, and refining until we adequately

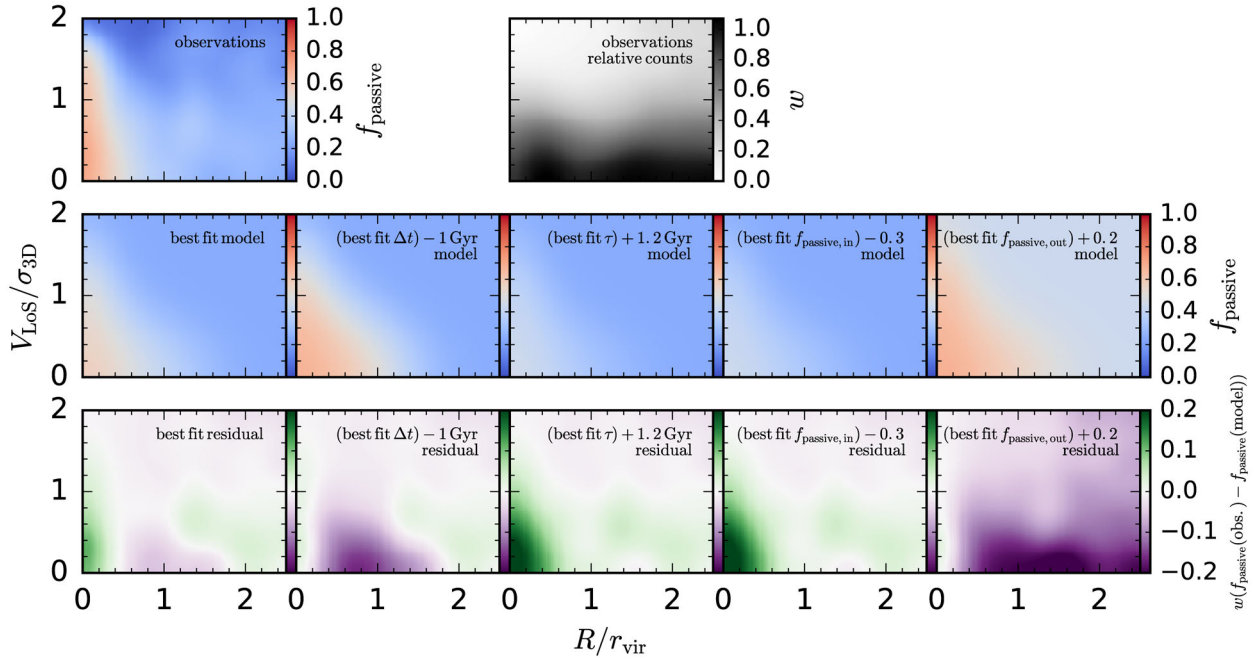


Figure 6. A visual representation of our quenching model. The top-left panel shows the observed passive fraction as a function of phase space coordinates for galaxies in the ranges $14 < \log_{10}(M_{\text{host}}/M_{\odot}) < 15$ and $9.5 < \log_{10}(M_{*}/M_{\odot}) < 10$, smoothed using a Gaussian kernel with $\sigma_V = \sigma_R = 0.2$ (all other panels use the same smoothing). The top-right panel shows the smoothed relative counts of observations in the phase space plane, which we denote w . The middle row shows the prediction of our model the parameter choices yielding the maximum likelihood (leftmost panel), and the effect of changing each of the parameters on the predictions; from left to right: Δt , τ , $f_{\text{passive,in}}$, $f_{\text{passive,out}}$. The bottom row shows the difference between the observations and the model predictions for each set of parameters from the second row, weighted by the relative counts: $w(f_{\text{passive}}(\text{obs.}) - f_{\text{passive}}(\text{model}))$.

sample the peak and 95 per cent confidence region of the likelihood distribution. The result of this fitting for the same subsample used illustratively in Fig. 6 is shown in Fig. 7. The panels along the diagonal show the posterior distributions for each of the four parameters, marginalized over all other parameters in each case. The dark (light) grey shaded regions show the 68 per cent (95 per cent) confidence intervals, and the solid vertical line indicates the parameter value at the global likelihood maximum (which may be distinct from the maximum of the marginalized distribution for a single parameter). The off-diagonal panels show the marginalized posterior distributions for pairs of model parameters. The two black contours show the 68 and 95 per cent confidence intervals, and the star symbol corresponds to the location of the global maximum likelihood (again, possibly distinct from the maximum of any given marginalized distribution). The other nine fits give qualitatively similar results; one key recurring feature of the posterior distributions is the degeneracy between Δt and τ , with shorter quenching times Δt being ‘compensated’ by longer transition time-scales τ , which will be discussed further in Section 5.

In Fig. 8 and Table 1, we summarize the results of fitting our model to each of the 10 data subsamples. All vertical error bars and quoted intervals represent 68 per cent confidence intervals (which in some cases are smaller than the symbols or run off the region shown); horizontal error bars simply indicate bin widths.

5 DISCUSSION

In the results of fitting our model, shown in Fig. 8, we note the expected trend in $f_{\text{passive,out}}$, with higher stellar mass galaxies outside the clusters that are more affected by ‘internal quenching’ having higher $f_{\text{passive,out}}$. Interestingly, in all cases we recover values of

$f_{\text{passive,in}}$ consistent with 1.0 within the 68 per cent confidence interval, and in most cases the best-fitting value is ≈ 1.0 , suggesting that quenching by clusters is 100 per cent efficient,³ in agreement with the conclusions of Mahajan et al. (2011). Of course, the *observed* fraction of quenched satellites will always be less than this, even in radial bins closest to the cluster centre, because these bins contain a fraction of satellites falling into the cluster for the first time, which have not yet had time to be quenched, and galaxies ‘projected into’ the cluster.

We find a flat or perhaps slightly decreasing trend in Δt with increasing stellar mass, and usually small values of τ (however, in a few cases up to several Gyr, though with 95 per cent confidence intervals still consistent with near-zero values). We examine the details of these trends in more detail below.

In the upper panel of Fig. 8, the lower best-fitting values of $f_{\text{passive,in}}$ in the lowest M_{*} bin stand out as peculiar (also, to a lesser extent, for the lower M_{host} bin in the $9.5 < \log_{10}(M_{*}/M_{\odot}) < 10$ bin). While this could be a sign that these lower mass galaxies are more resilient to quenching, considering other peculiarities in the fits in these bins we cautiously prefer an interpretation where $f_{\text{passive,in}} \approx 1.0$ in all cases. We first point out that the 68 per cent confidence intervals extend up to 1.0 in all cases. In all cases the marginalized posterior probability distribution for $f_{\text{passive,in}}$ (not shown) peaks at 1.0, but in these peculiar cases the global maximum likelihood is offset from the peak of the marginalized distribution. We believe that this is due to the proximity to the mass resolution limit in the simulations, which causes the satellite haloes contributing orbits for

³ Indicating that, once quenching by the cluster has had time to operate, 100 per cent of satellites have been quenched, but not that the cluster is responsible for quenching 100 per cent of the passive galaxies it contains.

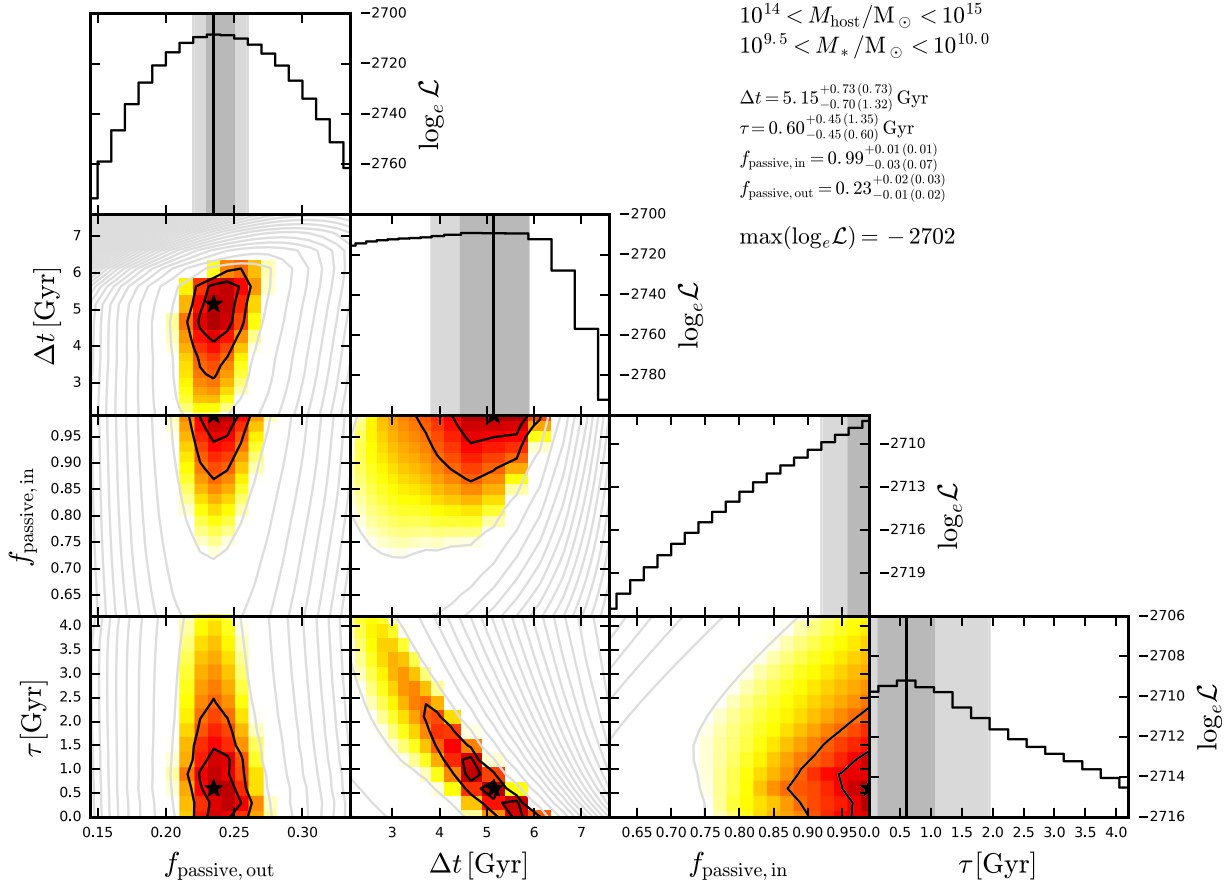


Figure 7. Marginalized likelihood distributions for individual model parameters (panels along the diagonal) and marginalized likelihood maps for pairs of model parameters (off-diagonal panels), for the data in the ranges $10^{14} < \log_{10}(M_{\text{host}}/M_{\odot}) < 10^{15}$ and $10^{9.5} < \log_{10}(M_{*}/M_{\odot}) < 10^{10.0}$. The location of the global maximum likelihood (i.e. without marginalization) is shown with a vertical black line (histograms) or a black star (maps). The 68 and 95 per cent confidence intervals are indicated by the dark and light shaded regions (histograms) and the inner and outer black contours (maps), respectively. In the maps, the pale grey contours show the overall shape of the likelihood distribution, with each contour representing an additional drop of 3 per cent relative to the maximum. The detailed distribution around the maximum is instead shown with a colour scale. The intervals given in the text labels are 68 (95) per cent confidence intervals.

use in this bin to be biased more towards the upper edge of the bin than they would otherwise be. Oman et al. (2013) showed that higher mass satellite haloes, which host higher mass galaxies, have orbits with preferentially smaller backsplash distances, which is easily understood as the effect of dynamical friction. If a PDF constructed from a collection of satellite orbits biased towards higher masses is used, when fitting the model, the lower mass galaxies have inferred times since infall that are biased low, driving down the fit quenching time-scales (Δt , τ , or both). This picture seems consistent with the time-scales plotted in the lower panel of Fig. 8, particularly for the higher M_{host} bin, which has a seemingly unrealistic⁴ best-fitting $\Delta t = 0$. This has a knock-on effect on $f_{\text{passive,in}}$, driving the best fit to lower values. This situation is exacerbated by the relatively low numbers of observed cluster satellite candidates in these mass bins (see Fig. 1). These difficulties are reflected in the statistical uncertainties derived from the posterior distribution; with the exception of those for $f_{\text{passive,out}}$, which is constrained primarily by the properties of interlopers, these are very large.

The trends seen in the τ parameter are also puzzling at first glance. This parameter turns out to be difficult to constrain using

our methodology, with 68 per cent confidence intervals up to several Gyr wide. Inspecting the marginalized posterior distributions (e.g. Fig. 7), we invariably find a strong degeneracy between τ and Δt . This is intuitive, as a rapid transition at a given time is numerically similar to a slightly slower transition that begins slightly earlier. We obtain tighter constraints by considering a representative single time-scale $t_{1/2}$. The trends and 68 per cent confidence intervals for this parameter combination for the higher (lower) M_{host} bin as a function of M_{*} are illustrated by the solid (dashed), red (pink) lines and corresponding shaded regions in Fig. 9. The intervals remain large for the subsamples with relatively low observed galaxy counts, but we verify via fits to a Monte Carlo sampling of the quenching time-scale distributions that the decreasing trend with increasing M_{*} is significant at 92 per cent (61 per cent) confidence for the lower (higher) M_{host} bin. Further efforts to understand these trends would likely benefit from a more sophisticated model which explicitly models the trends and fits data across the entire M_{*} and M_{host} range simultaneously.

We performed two tests to investigate the effect of changing the information contained in the infall time PDFs. In both cases, we used the same subsample used illustratively in Fig. 7. First, we reconstructed our PDFs binning only along the R direction in the (R, V) plane, effectively ignoring the velocity information and

⁴ Keeping in mind our definition of infall at $2.5 r_{\text{vir}}$.

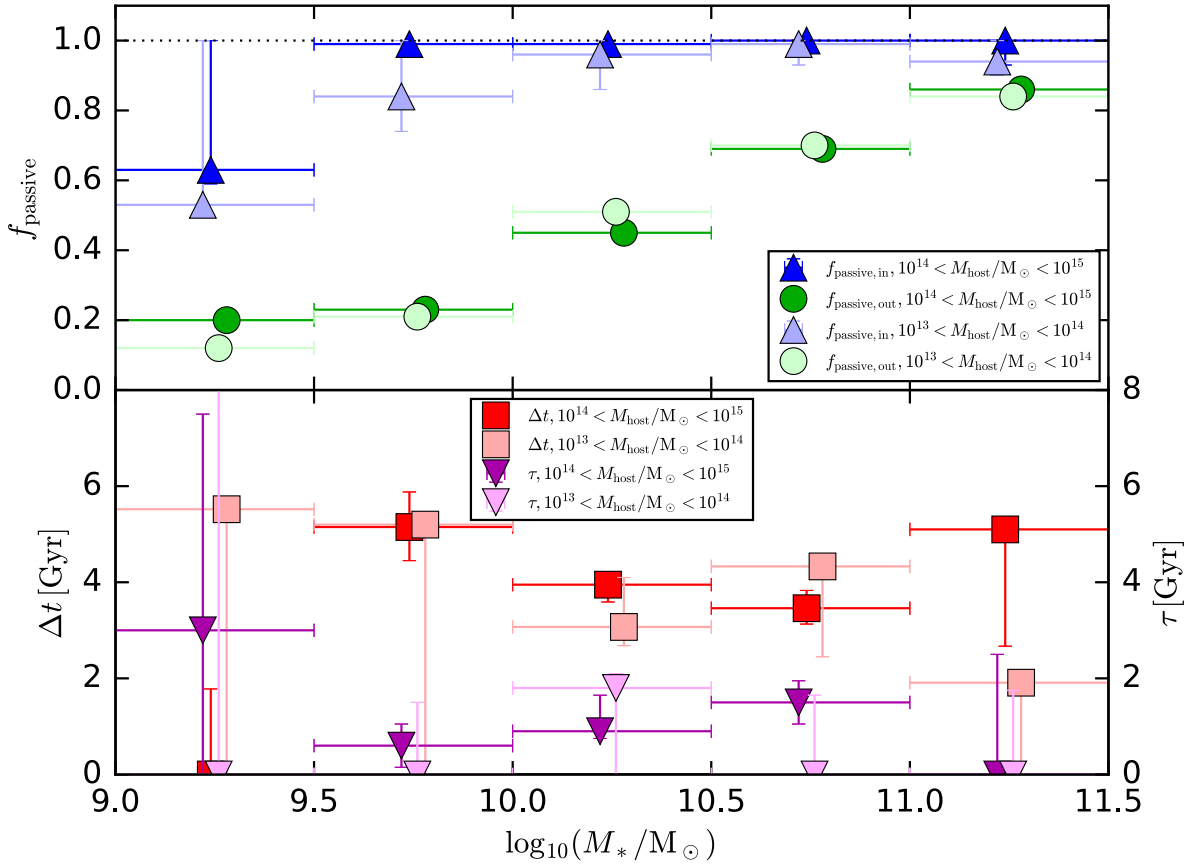


Figure 8. Maximum likelihood model parameters (upright triangles: $f_{\text{passive,in}}$; circles: $f_{\text{passive,out}}$; squares: Δt ; inverted triangles: τ) for the observational sample binned by stellar mass for host masses in the range $10^{13} - 10^{14} M_{\odot}$ (pale symbols) and $10^{14} - 10^{15} M_{\odot}$ (dark symbols). The horizontal error bars show the bin widths, the vertical error bars the 68 per cent confidence intervals. For clarity, the symbols have been slightly offset horizontally.

Table 1. Summary of our results. The first two columns indicate the host mass and stellar mass bins. The next four columns list the parameter choices yielding the maximum likelihood and the 68 per cent confidence intervals for each. The penultimate column lists $t_{1/2}$ (see text), including the 68 per cent confidence interval derived by marginalizing the posterior distribution over this parameter combination. Because we perform a new fit to derive errors on this parameter, and because of the approximate nature of a maximum likelihood grid search, the values are not always exactly equal to $\Delta t + 0.69\tau$, but are consistent within our quoted errors. The last column lists the maximum likelihood.

$M_{\text{host}} (M_{\odot})$	$M_{*} (M_{\odot})$	$f_{\text{passive,in}}$	$f_{\text{passive,out}}$	Δt (Gyr)	τ (Gyr)	$t_{1/2}$ (Gyr)	$\max(\log_e \mathcal{L})$
$10^{13}-10^{14}$	$10^9-10^{9.5}$	$0.53^{+0.47}_{-0.00}$	$0.12^{+0.02}_{-0.01}$	$5.52^{+0.00}_{-5.52}$	$0.00^{+11.00}_{-0.00}$	$5.32^{+2.70}_{-0.41}$	-237
	$10^{9.5}-10^{10}$	$0.84^{+0.16}_{-0.10}$	$0.21^{+0.02}_{-0.03}$	$5.20^{+0.00}_{-5.20}$	$0.00^{+1.50}_{-0.00}$	$5.20^{+0.75}_{-0.29}$	-927
	$10^{10}-10^{10.5}$	$0.96^{+0.04}_{-0.10}$	$0.51^{+0.02}_{-0.02}$	$3.07^{+1.03}_{-0.39}$	$1.80^{+0.30}_{-1.80}$	$4.17^{+0.22}_{-0.30}$	-2439
	$10^{10.5}-10^{11}$	$0.99^{+0.01}_{-0.06}$	$0.70^{+0.02}_{-0.02}$	$4.33^{+0.00}_{-1.88}$	$0.00^{+1.65}_{-0.00}$	$4.23^{+0.17}_{-0.35}$	-2032
	$10^{11}-10^{11.5}$	$0.94^{+0.06}_{-0.04}$	$0.84^{+0.01}_{-0.04}$	$1.91^{+0.09}_{-1.91}$	$0.00^{+1.75}_{-0.00}$	$1.85^{+0.98}_{-0.54}$	-260
$10^{14}-10^{15}$	$10^9-10^{9.5}$	$0.63^{+0.37}_{-0.04}$	$0.20^{+0.01}_{-0.02}$	$0.00^{+1.78}_{-0.00}$	$3.00^{+4.50}_{-3.00}$	$5.32^{+4.25}_{-5.32}$	-1008
	$10^{9.5}-10^{10}$	$0.99^{+0.01}_{-0.03}$	$0.23^{+0.02}_{-0.01}$	$5.15^{+0.73}_{-0.70}$	$0.60^{+0.45}_{-0.45}$	$5.66^{+0.29}_{-0.23}$	-2702
	$10^{10}-10^{10.5}$	$0.99^{+0.01}_{-0.01}$	$0.45^{+0.01}_{-0.01}$	$3.95^{+0.15}_{-0.36}$	$0.90^{+0.75}_{-0.15}$	$4.68^{+0.24}_{-0.28}$	-6730
	$10^{10.5}-10^{11}$	$1.00^{+0.00}_{-0.03}$	$0.69^{+0.02}_{-0.01}$	$3.46^{+0.37}_{-0.33}$	$1.50^{+0.45}_{-0.45}$	$4.19^{+0.73}_{-0.31}$	-6165
	$10^{11}-10^{11.5}$	$1.00^{+0.00}_{-0.07}$	$0.86^{+0.02}_{-0.02}$	$5.10^{+0.00}_{-2.43}$	$0.00^{+2.50}_{-0.00}$	$5.03^{+0.40}_{-0.63}$	-758

emulating the scenario where robust redshifts for cluster members are unavailable. The fit using this modified PDF is broadly similar to the one using the PDF including the velocity information. The preferred τ drops from $0.60^{+0.45}_{-0.45}$ to $0.00^{+0.15}_{-0.00}$ Gyr, and Δt increases from $5.15^{+0.73}_{-0.70}$ to $5.74^{+0.14}_{-0.35}$ Gyr. The values are consistent

within the quoted confidence intervals, and we note that, though Δt and τ vary individually, the combined time-scale $t_{1/2}$ increases by only 0.18 Gyr. The statistical uncertainties on all parameters are somewhat narrower when the velocity information is not used, which at first seems surprising. However, the maximum likelihood

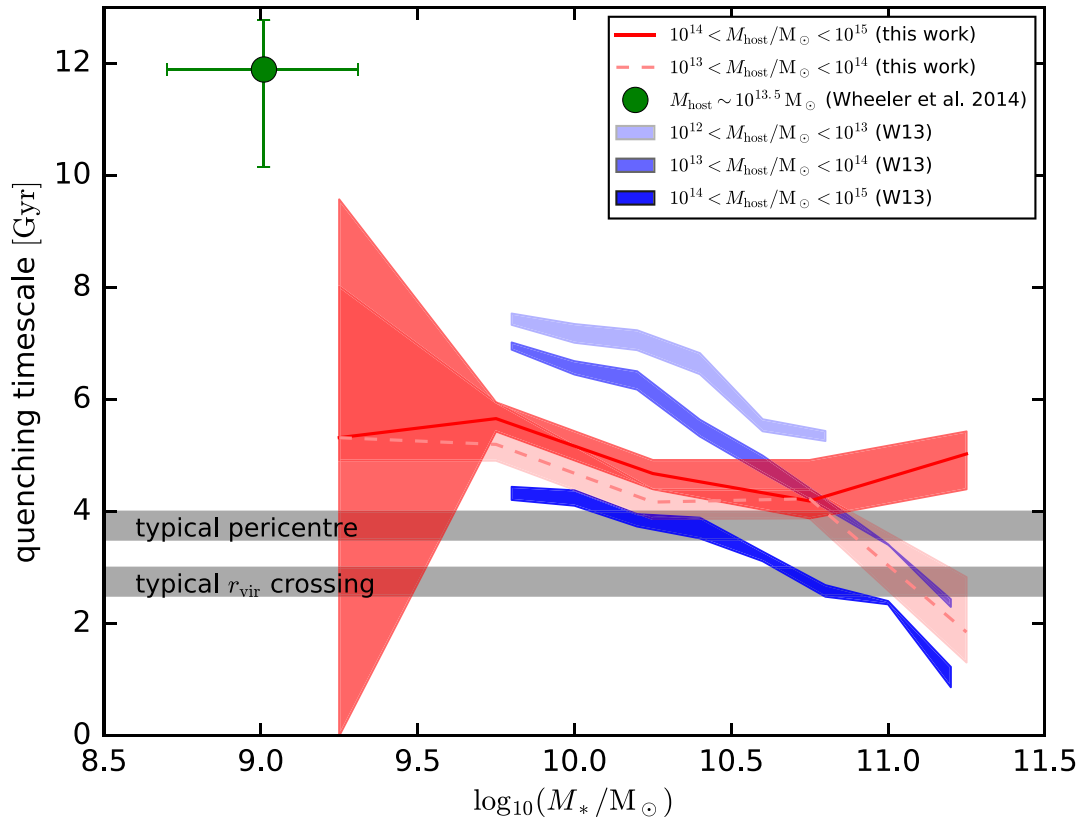


Figure 9. Comparison of our combined time-scale $t_{1/2}$ values (lines) and 68 per cent confidence intervals (shaded regions) as a function of M_* with those of W13 (shaded regions represent 68 per cent confidence intervals) and Wheeler et al. (2014, horizontal error bars representing the interquartile range of M_* for their sample, vertical error bars representing the uncertainty on satellite quenched fraction from 25 to 55 per cent). Host mass ranges or representative values are as shown in the legend. In both cases, their definitions of infall time differ from ours, but we attempt to correct for the differences. See Section 5 for the details of these corrections.

drops from -2702 to -2721 , a formally very significant ($\sim 5.3\sigma$) difference. The narrower confidence intervals are a natural consequence of the poorer fit: the χ^2 is larger, so the change in a parameter required to produce a given change in χ^2 shrinks, apparently leading to narrower confidence intervals, but this is an illusion due to a poorer model fit.

The second test we performed was to reconstruct the infall time PDFs by dividing the (R, V) plane into 50 bins in each direction (our fiducial PDFs use 100×100 bins). In this case we recover a formally somewhat better fit, with the likelihood increasing from -2702 to -2697 ($\sim 2.2\sigma$ significant). The best-fitting parameters are consistent within the confidence intervals; τ drops to $0.00^{+0.45}_{-0.00}$ Gyr and Δt increases to $5.76^{+0.12}_{-0.36}$ Gyr, again highlighting the degeneracy between the two parameters. For this reason we prefer to focus on the ‘combined’ time-scale $t_{1/2}$, but we note that our conclusion that τ prefers small values $\lesssim 2$ Gyr appears to be robust.

5.1 Comparison with other works

Our time-scales are not directly comparable to many previous studies of satellite quenching for two reasons. First, the radius at which ‘infall’ is defined ($2.5 r_{\text{vir}}$) is larger than most previous studies which typically adopt $1.0 r_{\text{vir}}$ (with varying definitions of ‘virial’). As noted above, we chose this large radius to avoid the ambiguity of tracking ‘backsplash’ subhaloes which would otherwise exit and re-enter the

virial radius. A correction for this difference is relatively straightforward, since the time for a typical subhalo to move from $2.5 r_{\text{vir}}$ to $1.0 r_{\text{vir}}$ is ~ 3 Gyr (in detail this depends on which virial definitions are assumed). Secondly, some previous studies define the time for quenching since the *first* time a subhalo falls into a larger halo of *any* mass. Thus, for example, 30 per cent of satellites falling into an $\sim 10^{14} M_{\odot}$ cluster halo had already become satellites of a lower mass group that then fell into the cluster-mass halo. Their quenching time therefore includes the time a satellite spent being ‘pre-processed’. In contrast, our methodology compares a quenched population ($f_{\text{passive,in}}$) with an infalling population that is already ‘pre-processed’ ($f_{\text{passive,out}}$) and so isolates the quenching that is due only to falling into the current $\sim 10^{14} M_{\odot}$ host halo. Consequently, due to the different definitions, if the infall radii were the same, our times since infall would always be shorter.

In Fig. 9, we show a comparison of our results with those of W13. We compare our combined time-scale $t_{1/2}$ with their t_Q parameter, which is similarly a combination of a delay and a transition time-scale (though in W13 the ‘transition’ time refers to the time for an individual galaxy to ‘fade’ from blue to red). In order to compare quantitatively – W13 uses a very different methodology to ours, but a sample with overlapping mass cuts – we have attempted to make a correction for the different definitions mentioned above. We correct for the offset between first infall (their preferred definition) and recent infall (which corresponds to crossing r_{200b}), which depends

on the host mass,⁵ using the data from their fig. 2. We also correct for the median time between crossing $2.5 r_{\text{vir}}$ (i.e. our infall time) and crossing r_{200b} , which is 2.5 Gyr. We make a further small correction for ‘ejected’ (i.e. ‘backsplash’) haloes using the time-scales in Wetzel et al. (2014). The total offsets we apply to the W13 results for their three host mass bins (low to high) are 2.2, 1.1 and -0.1 Gyr. The comparison is shown in Fig. 9. We find the same trend of a decreasing quenching time-scale with increasing M_* , though it appears the slope in our results is somewhat shallower. We also find a much weaker trend than W13, perhaps no trend, with M_{host} in the range probed by our sample. Some of the difference may be explained by the different treatment of ‘pre-processing’. At the high M_* , high M_{host} end, W13 find quenching times that, in our interpretation, correspond to quenching over a Gyr *before* first entering r_{vir} . This seems likely to be the signature of pre-processing in another group or cluster. In contrast, in our methodology which treats ‘pre-processed’ galaxies simply as part of the passive portion of the infalling galaxy population, and so isolates the effect of the final host, quenching times are restricted to around or after the time of the first pericentric passage (marked with a horizontal grey band in the Fig. 9).

We also plot for comparison in Fig. 9 the result of Wheeler et al. (2014).⁶ Again, the values are not directly comparable with our own, so we attempt to adjust them to match our definitions. We increase their reported time-scale by 1.7 Gyr to account for the difference between the infall time into any more massive host and the most recent infall into a more massive host, again guided by fig. 2 of W13 (the median host mass of the Wheeler et al. 2014 sample is $10^{13.5} M_{\odot}$), and a further offset of 2.5 Gyr to account for the travel time between $2.5 r_{\text{vir}}$ and r_{vir} .⁷ We omit the other results reported in Fillingham et al. (2015) from our comparison figure as they have no overlap in either M_{host} or M_* with our sample.

Recently, a number of authors (Wheeler et al. 2014; Fillingham et al. 2015; Wetzel, Tollerud & Weisz 2015; Mistani et al. 2016) have suggested that galaxies with $M \sim 10^9 M_{\odot}$ are significantly more resistant to quenching than satellites with both higher and lower masses. Our results for our lowest M_* bin are highly uncertain, but we do seem to find an increase towards $M_* \sim 10^9$, even though we cannot make any strong statements about the time-scale at this mass scale. However, the time-scales we find at higher masses are clearly lower than the 11.9 Gyr (value estimated assuming our definitions) found by Wheeler et al. (2014). Wetzel et al. (2015) infer a time-scale of ~ 8 Gyr for somewhat less massive $M \sim 10^{8.5} M_{\odot}$ satellites of the Milky Way and M 31, and Fillingham et al. (2015) find substantially lower time-scales, again around smaller hosts than those in our sample. This suggests that our results are plausibly consistent with the conclusion that $M_* \sim 10^9 M_{\odot}$ satellites are most resistant to quenching, but that the host halo mass dependence remains to be better understood.

5.2 Disentangling the physical mechanisms responsible for quenching

It is interesting to consider what the observed time-scale of quenching and its dependence on stellar mass reveals regarding the astrophysical mechanisms responsible for quenching star formation. A number of physical processes occur when a galaxy falls into a cluster halo. First, the accretion of dark matter and gas on to the halo is cut off while the satellite halo is still outside the virial radius. Secondly, ram pressure stripping may strip the hot gas halo as well as the cold gas from the disc. Finally, outflows due to galactic winds may deplete the gas that is available for star formation.

It has long been assumed in models of galaxy formation that a galaxy stops accreting gas on to its own halo when it becomes a satellite (e.g. Kauffmann, White & Guiderdoni 1993; Cole et al. 1994), and this is also observed in SPH simulations (Kereš et al. 2009). The cluster-centric radius at which this cutoff occurs is not well known. For example, Bahé et al. (2013, see their fig. 8) argue that satellite dark matter haloes stop growing within $\sim 2 r_{200c}$ of a cluster, but that their hot gas content is already reduced as far out as $5 r_{200c}$. Behroozi et al. (2014) have also shown that dark matter accretion ends roughly when the satellite is as far out as $\sim 2 r_{\text{vir}}$. So it is likely that the cutoff of accreting gas occurs further out than the fiducial virial radius, and closer to our ‘backsplash’ limit of $2.5 r_{\text{vir}}$.

Even if the supply of new gas is cut off, the existing reservoir of cold and hot gas is large enough to sustain star formation in excess of the Hubble time at typical SFRs: $t \sim (M_{\text{baryon}} - M_*)/\text{SFR}(M_*)$. Dividing both numerator and denominator by M_* gives $t \sim (f_{\text{baryon}}/f_* - 1)/\text{SSFR}$, where the fraction of mass in baryons $f_{\text{baryon}} \sim 0.15$ and the fraction of total mass in stars $f_* \sim 0.01$ for the lowest stellar mass galaxies (Hudson et al. 2015). Therefore, to explain the short quenching times, additional mechanisms are required to remove or heat the existing gas. As discussed in Section 1, ram pressure stripping of cold gas is clearly seen in galaxy clusters. The key results of this paper are that (i) the quenching occurs approximately at or shortly after pericentre passage, (ii) after the delay Δt , it is 100 per cent effective, (iii) that the time for low-mass galaxies to quench is slightly longer than the time for higher mass galaxies and (iv) the infalling population transitions to become the cluster population relatively quickly (once the delay Δt has elapsed), on a time-scale $\tau \lesssim 2$ Gyr. Because ram pressure stripping is strongest close to pericentre, the observed timing of the quenching is in broad agreement with the ram pressure stripping model. However, whether the remaining two observations are in accordance with this model is less clear. Ram pressure stripping is expected to be more effective for low mass satellites because the restoring force of the disc is lower, which would argue against the model. However, larger galaxies are more affected by dynamical friction and the ram pressure is very sensitive to speed; it is proportional to the square of the speed through the intra cluster medium. Smaller satellites are likely still proportionally more affected by ram pressure (Bahé & McCarthy 2015), so the trend remains puzzling unless either smaller satellites have some intrinsic property causing them to take longer to cease forming stars or ram pressure is not the dominant trigger of quenching for satellites of all masses in clusters, or some combination of both.

Recently, McGee et al. (2014) have advocated for the combination of a cutoff in gas accretion as a galaxy falls into a cluster, coupled with strong outflows driven by galactic winds (a model which they dub ‘overconsumption’), to explain the quenching time-scales. In their model, the key parameter is the mass-loading of the winds: $\eta = \dot{M}_{\text{out}}/\text{SFR}$, where \dot{M}_{out} represents the rate at which

⁵ The difference between first infall and most recent infall also depends on M_* , but this is a much weaker effect (Wetzel, private communication) that we neglect here.

⁶ We use the result as reported by Fillingham et al. (2015), which includes uncertainty estimates.

⁷ Wheeler et al. (2014) define infall based on friends-of-friends (FoF) group membership. This definition is unfortunately awkward for comparison; we simply assume that the edge of the FoF group corresponds to $\sim r_{\text{vir}}$.

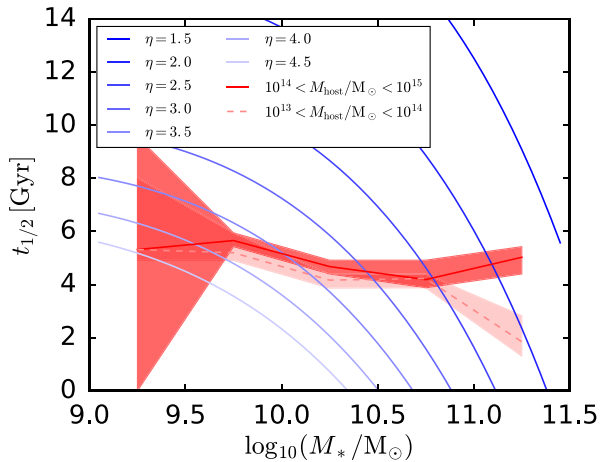


Figure 10. Comparison of our quenching time-scales as a function of M_* with the simple overconsumption models of McGee, Bower & Balogh (2014), parametrized by the mass-loading factor η .

gas is permanently ejected from the satellite’s halo. The quenching time is highly sensitive to η : if η is too low then the quenching time is longer than the Hubble time, too high and the quenching time rapidly approaches zero. McGee et al. (2014) and Balogh et al. (2016) argue that $\eta \sim 1.5$, independent of mass. We have used equation 7 of McGee et al. (2014) to fit their model to our quenching times, assuming no stripping. We have adopted the stellar-to-halo mass relation from weak lensing (Hudson et al. 2015), and the low-redshift SFR of Salim et al. (2007). Finally, we have also assumed that the ‘clock’ for (over)consumption starts ticking when the satellites crosses $2.5 r_{\text{vir}}$. The predictions for contours of constant η are shown in Fig. 10. The data are fitted with a slow varying η that ranges from 2.0 at high mass to 4.0 at low mass.

A slowly varying η model is in conflict, however, with other results on the mass loading of outflows. In particular, one would expect galaxies with shallower potential wells to have more efficient outflows. This is found in numerical simulations: Muratov et al. (2015) find that the mass-loading factor scales as $\eta \propto v_{\text{circ}}^{-1}$, and Keller, Wadsley & Couchman (2016) finds a constant $\eta \sim 8$ for low-mass systems, decreasing with a power-law slope of about -1.8 above $M_* = 10^{10} M_{\odot}$. From analytic arguments based on the scaling relations of low-mass galaxies and the baryonic TF relation, Dutton (2012) finds that $\eta \sim v_{\text{circ}}^{-2}$. These three models would predict higher values of η at low stellar mass, and hence short quenching times for low-mass galaxies for which winds efficiently remove the gas.

The above derived η is an upper limit on the true η , because ram pressure stripping of the hot gas reservoir will substantially reduce the amount of gas potentially available for star formation (Bahé & McCarthy 2015). Furthermore, the ‘effective’ η is likely to be higher for infalling satellites than similar counterparts in the field, because weak outflows that in field galaxies would return as a galactic fountain are instead stripped by ram pressure (Bahé & McCarthy 2015).

Overall, the timing of quenching near pericentre suggests that ram pressure stripping plays a role. However, the fact that low-mass galaxies have longer quenching time delays than high-mass galaxies is difficult to understand, because whatever the mechanism of gas removal, whether ram pressure stripping or galactic winds, it should be more effective in low-mass galaxies with shallower potential wells.

6 SUMMARY AND OUTLOOK

We have compared subhalo orbit libraries in PPS to SFRs of SDSS galaxies. This method isolates the environmental effects of the most recent host; in this paper, this is a cluster of mass $10^{13} - 10^{15} M_{\odot}$. The key results of this paper are as follows.

(i) Quenching occurs after a delay time Δt , measured from first crossing of $2.5 r_{\text{vir}}$. This delay time is typically 3.5–5 Gyr, with higher mass galaxies quenching slightly earlier. In most cases, this corresponds to times near or shortly after first pericentric approach. All galaxies are quenched on first infall, and before apocentre.

(ii) The delay time does not depend (or depends very weakly) on the host halo mass, over the relatively narrow range probed by our sample.

(iii) Once quenching begins, the time-scale τ for the galaxy population to transition from resembling the galaxies outside the cluster (described by $f_{\text{passive, out}}$) to those processed by the cluster (described by $f_{\text{passive, in}}$) is fairly short, $\lesssim 2$ Gyr, and usually consistent with 0 Gyr. Note that this time-scale is distinct from the time-scale for individual galaxies to transition from an active to a passive state, i.e. the time-scale for ‘crossing the green valley’.

(iv) After the delay has elapsed, the quenching is 100 per cent effective, i.e. all active galaxies that fell in longer ago than $\approx \Delta t + 2\tau$ are passive. The observed fraction of star-forming galaxies in rich clusters is therefore due to a combination of interlopers and galaxies that are falling in for the first time.

These results appear to be in reasonable agreement with some previous work (W13; Wheeler et al. 2014), after correction the fact that delay times in these works are measured at first accretion and at a different radius.

In this paper, we have shown how an orbit library can be used to deproject infalling, backslash and virialized populations. Here we have compared our projected models with SFR data in PPS, but only for a very simple parametrization of the SFR distribution; this could be extended to leverage the additional information contained in the full SFR distributions. Furthermore, there is no reason to limit the comparison to only SFR, particularly since it is well known that morphology is also correlated with environment. Tidal or harassment effects may also affect the *structures* of discs, possibly stripping them (reduction in stellar mass and radius) or puffing them up so that they are identified morphologically as bulges. In addition, here we have limited ourselves to the correlation between time since infall and PPS position, however much more information is contained in orbit libraries. As redshift surveys continue to improve and grow, we expect that increasingly subtle effects can be teased out of the data.

ACKNOWLEDGEMENTS

We wish to thank Michael Balogh, Coral Wheeler, Laura Sales and Andrew Wetzel for useful discussions. We thank Sara Ellison for providing tabulated stellar mass and SFR data and Andrew Wetzel for providing results in electronic form. We also thank the authors of the Bolshoi and MDR1 simulations for making their simulation outputs publicly available, and Peter Behroozi for assistance with the halo finder and merger tree codes. MJH acknowledges support from an NSERC Discovery grant. This research has made use of NASA’s Astrophysics Data System.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Bahé Y. M., McCarthy I. G., 2015, *MNRAS*, 447, 969
- Bahé Y. M., McCarthy I. G., Balogh M. L., Font A. S., 2013, *MNRAS*, 430, 3017
- Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681
- Baldry I. K. et al., 2012, *MNRAS*, 421, 621
- Balogh M. L., Navarro J. F., Morris S. L., 2000, *ApJ*, 540, 113
- Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, *ApJ*, 615, L101
- Balogh M. L. et al., 2009, *MNRAS*, 398, 754
- Balogh M. L. et al., 2016, *MNRAS*, 456, 4364
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, *ApJ*, 762, 109
- Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, *ApJ*, 763, 18
- Behroozi P. S., Wechsler R. H., Lu Y., Hahn O., Busha M. T., Klypin A., Primack J. R., 2014, *ApJ*, 787, 156
- Bluck A. F. L., Mendel J. T., Ellison S. L., Moreno J., Simard L., Patton D. R., Starkenburg E., 2014, *MNRAS*, 441, 599
- Boselli A., Gavazzi G., 2006, *PASP*, 118, 517
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, *MNRAS*, 321, 559
- Cheung E. et al., 2012, *ApJ*, 760, 131
- Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, 271, 781
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Dekel A., Sari R., Ceverino D., 2009, *ApJ*, 703, 785
- Dutton A. A., 2012, *MNRAS*, 424, 3123
- Fang J. J., Faber S. M., Koo D. C., Dekel A., 2013, *ApJ*, 776, 63
- Fillingham S. P., Cooper M. C., Wheeler C., Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., 2015, *MNRAS*, 454, 2039
- Font A. S. et al., 2008, *MNRAS*, 389, 1619
- Gill S. P. D., Knebe A., Gibson B. K., 2005, *MNRAS*, 356, 1327
- Giovanelli R., Haynes M. P., 1985, *ApJ*, 292, 404
- Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, *ApJ*, 600, 580
- Graves G. J., Faber S. M., Schiavon R. P., 2009, *ApJ*, 693, 486
- Gunn J. E., Gott J. R., III1972, *ApJ*, 176, 1
- Hernández-Fernández J. D., Haines C. P., Diaferio A., Iglesias-Páramo J., Mendes de Oliveira C., Vilchez J. M., 2014, *MNRAS*, 438, 2186
- Hirschmann M., De Lucia G., Wilman D., Weinmann S., Iovino A., Cucciati O., Zibetti S., Villalobos Á., 2014, *MNRAS*, 444, 2938
- Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, *ApJS*, 163, 1
- Hudson M. J., Stevenson J. B., Smith R. J., Wegner G. A., Lucey J. R., Simard L., 2010, *MNRAS*, 409, 405
- Hudson M. J. et al., 2015, *MNRAS*, 447, 298
- Jaffé Y. L., Smith R., Candlish G. N., Poggianti B. M., Sheen Y.-K., Verheijen M. A. W., 2015, *MNRAS*, 448, 1715
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kauffmann G. et al., 2003, *MNRAS*, 341, 33
- Keller B. W., Wadsley J., Couchman H. M. P., 2016, *MNRAS*, 00, 00
- Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, *MNRAS*, 395, 160
- Kimm T. et al., 2009, *MNRAS*, 394, 1131
- Kitzbichler M. G., White S. D. M., 2008, *MNRAS*, 391, 1489
- Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, *ApJ*, 522, 82
- Larson R. B., Tinsley B. M., Caldwell C. N., 1980, *ApJ*, 237, 692
- Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, *ApJ*, 692, 931
- McGee S. L., Bower R. G., Balogh M. L., 2014, *MNRAS*, 442, L105
- Mahajan S., Mamon G. A., Raychaudhury S., 2011, *MNRAS*, 416, 2882
- Mamon G. A., Sanchis T., Salvador-Solé E., Solanes J. M., 2004, *A&A*, 414, 445
- Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, *ApJS*, 210, 3
- Mistani P. A. et al., 2016, *MNRAS*, 455, 2323
- Mok A. et al., 2014, *MNRAS*, 438, 3070
- Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, *MNRAS*, 454, 2691
- Muzzin A. et al., 2014, *ApJ*, 796, 65
- Oman K. A., Hudson M. J., Behroozi P. S., 2013, *MNRAS*, 431, 2307
- Omand C. M. B., Balogh M. L., Poggianti B. M., 2014, *MNRAS*, 440, 843
- Peng Y.-j. et al., 2010, *ApJ*, 721, 193
- Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, *MNRAS*, 423, 3018
- Salim S. et al., 2007, *ApJS*, 173, 267
- Smith R. J., Lucey J. R., Hudson M. J., 2009, *MNRAS*, 400, 1690
- Smith R. J. et al., 2010, *MNRAS*, 408, 1417
- Taranu D. S., Hudson M. J., Balogh M. L., Smith R. J., Power C., Oman K. A., Krane B., 2014, *MNRAS*, 440, 1934
- van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, *MNRAS*, 387, 79
- Velander M. et al., 2014, *MNRAS*, 437, 2111
- von der Linden A., Best P. N., Kauffmann G., White S. D. M., 2007, *MNRAS*, 379, 867
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, *MNRAS*, 372, 1161
- Weinmann S. M., Kauffmann G., von der Linden A., De Lucia G., 2010, *MNRAS*, 406, 2249
- Wetzel A. R., Tinker J. L., Conroy C., 2012, *MNRAS*, 424, 232
- Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, *MNRAS*, 432, 336 (W13)
- Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2014, *MNRAS*, 439, 2687
- Wetzel A. R., Tollerud E. J., Weisz D. R., 2015, *ApJ*, 808, L27
- Wheeler C., Phillips J. I., Cooper M. C., Boylan-Kolchin M., Bullock J. S., 2014, *MNRAS*, 442, 1396

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