

Estimating infall times of galaxies around clusters: Working to achieve accurate results based on observational data

Haoran Dou¹ and Heng Yu^{1*}

School of Physics and Astronomy, Beijing Normal University, Beijing 100875, PR China

Received 21 April 2025 / Accepted 22 May 2025

ABSTRACT

Context. The environment plays a crucial role in galaxy evolution, particularly for galaxies infalling into clusters. Accurately estimating the infall times of galaxies from observations can significantly enhance our understanding of the environmental effects on galaxy evolution.

Aims. This paper is aimed at evaluating existing methods for estimating infall times via the $R - V$ diagram. Here, we explore plausible strategies for improving the accuracy in estimating infall times and we discuss the fundamental limitations.

Methods. We utilised a TNG300-1 simulation and constructed the $R - V$ diagram that is directly comparable to the observations. Using the same dataset, we systematically compared four commonly used methods, including the projected radii, caustic profiles, and two discrete methods. A simple linear partition was also considered as a reference.

Results. Each method exhibits distinct characteristics. While the linear partition slightly outperforms other methods, all methods suffer from limited accuracy (≥ 2.6 Gyr), constrained by the intrinsic dispersion (2.53 Gyr) of infall times in the $R - V$ diagram. Given this limit, we explored two potential approaches that could improve the accuracy: (1) the infall time dispersion is smaller in more dynamically relaxed clusters and (2) employing two estimates of infall times instead of one reduces the dispersion to ≤ 1.5 Gyr. We further demonstrate that the intrinsic dispersion primarily arises from orbital overlap: galaxies in different orbital phases overlap with each other in the $R - V$ diagram and thus appear indistinguishable.

Conclusions. Orbital overlap fundamentally limits the accuracy of infall time estimation. The linear partition approach could be a simple and robust estimation.

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

1. Introduction

The evolution of galaxies is one of the fundamental questions in astrophysics. It is widely known that the environment plays an important role in galactic evolution (e.g. Dressler 1980; Dressler et al. 1997; Goto et al. 2003; Kauffmann et al. 2004; Alpaslan et al. 2015; Chen et al. 2017; Taylor et al. 2023; Shi et al. 2024; Zheng et al. 2024). Galaxies residing in denser environments, such as clusters and groups, tend to be red, gas-depleted, metal-rich, exhibiting an early-type morphology, and with lower star formation rates (SFRs), in contrast to field galaxies, which typically exhibit the opposite properties.

Tracing the infall process of galaxies into clusters offers a promising way to understand how the environment drives galaxy evolution. In simulations, fully six-dimensional (6D) kinematic information and complete motion histories allow for an accurate reconstruction of the orbits of galaxies (e.g. Mahajan et al. 2011; Oman et al. 2013; Oman & Hudson 2016; Rhee et al. 2017, 2020; Pasquali et al. 2019). However, observational data provide only three dimensions of kinematic information: two projected positions on the sky and one line-of-sight (LOS) velocity. Therefore, the galaxy infall process is commonly analysed via the $R - V$ diagram, also referred to as the projected phase space (PPS) in simulation studies. This diagram comprises the projected cluster-centric radius (R_{2D}) and the LOS velocity relative to the cluster centre (V_{10s}).

Several studies have explored the potential of the $R - V$ diagram for tracing the infall process and have proposed various methods for estimating galaxy infall times. Noble et al. (2013, 2016) employed caustic profiles to delineate regions corresponding to different infall stages. Oman et al. (2013) and Oman & Hudson (2016) used the Multidark Run 1 dark matter-only simulation to construct orbital libraries and calculated probability density functions of infall times at each location in the diagram. Rhee et al. (2017) divided the $R - V$ diagram into five zones designed to maximise the fraction of galaxies belonging to a specific infall population in each zone. Pasquali et al. (2019) introduced a set of quadratic curves to partition the $R - V$ diagram into eight zones, achieving an improved alignment with the mean infall time distribution compared with caustic profiles. These methods have been widely adopted in observational studies (e.g. Kim et al. 2023; Sampaio et al. 2021, 2024; Brambila et al. 2023; Oxland et al. 2024).

However, these methods have not been systematically evaluated under the same conditions. It remains unclear which method provides the most reasonable estimates of infall times or to what extent the infall times can be accurately estimated from observational $R - V$ diagrams. Although some works have reported accuracy levels, such as 2.58 Gyr in Oman et al. (2013) and 1.49–2.60 Gyr in Pasquali et al. (2019), there is no easy way to combine the intrinsic uncertainties with the observational data. This study aims to address that question through a consistent and comparative evaluation of different methods, while also exploring possible improvements.

* Corresponding author: yuheng@bnu.edu.cn

The simulation data used for this study are presented in Section 2, followed by a detailed description of the infall time distribution in the R – V diagram in Section 3. Section 4 compares four commonly used methods for infall time estimation alongside a simple linear partition method for reference. In Section 5, we explore possible approaches to improve accuracy and, in particular, we discuss the orbital overlap issue that contributes to the main dispersion in Section 5.3. Finally, we summarise the results in Section 6.

2. Data

2.1. Simulations

IllustrisTNG (hereafter, TNG) is a suite of large-volume, cosmological, gravo-magnetohydrodynamic simulations that includes a comprehensive model for galaxy formation physics (Springel et al. 2018; Pillepich et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Marinacci et al. 2018). TNG 300 has the largest volume ($205 h^{-1} \text{Mpc}^3$) of TNG and TNG 300-1 has the highest resolution level of TNG 300, including initial 2×2500^3 particles, with a dark matter cell mass of $4.0 \times 10^7 h^{-1} M_\odot$ and a baryonic matter cell mass of $7.6 \times 10^6 h^{-1} M_\odot$.

The public release of TNG 300 provides 100 snapshots between $z = 20$ and $z = 0$ and a group catalogue is provided for each snapshot. These catalogues contain halos (groups or clusters) identified with the standard friends-of-friends algorithm (Davis et al. 1985), as well as sub-halos (galaxies) identified with the Subfind algorithm (Springel et al. 2001; Dolag et al. 2009). The merger histories of sub-halos are constructed via SubLink (Rodríguez-Gomez et al. 2015) and LHaloTree (Springel et al. 2005). Hereafter, we use ‘clusters’ and ‘galaxies’ to refer to halos and sub-halos, respectively.

2.2. Cluster sample

We selected clusters from the redshift $z = 0$ snapshot, applying a lower mass threshold of $M_{200} > 10^{14} M_\odot$. As Rhee et al. (2017) demonstrated that galaxy behaviour in the R – V diagram shows negligible dependence on cluster mass, we excluded systems with $M_{200} < 10^{14} M_\odot$, which are typically classified as ‘groups’. To minimise the impact of cluster mergers, which can complicate member galaxy motions and introduce additional dispersion into the infall time estimation, we required the clusters to have all member galaxies within $3R_{200}$.

While TNG provides merger trees and motion histories for galaxies, it does not explicitly track clusters. Therefore, we use the central galaxies as proxies for the positions and motions of clusters. In most cases, a central galaxy always remains at the centre of a cluster. However, temporary misclassifications may occur in some snapshots, especially at high redshifts. Although such misclassifications generally do not affect the overall properties of clusters, they can introduce discontinuities in the infall trajectories of member galaxies. A cluster is excluded from our sample if its central galaxy is misclassified in more than three consecutive snapshots, as this would cause a significant disruption in the reconstructed trajectories. If the mis-classified snapshots are non-consecutive or no greater than 3, we simply discard those specific snapshots to retain as many clusters as possible.

Our final sample comprises 136 clusters, with M_{200} ranging from $1.01 \times 10^{14} M_\odot$ to $1.03 \times 10^{15} M_\odot$. Their characteristic radius (R_{200}) and mass (M_{200}) are obtained from the TNG group catalogues. Here, R_{200} is the radius within which the average matter

density is 200 times the critical density of the universe at the corresponding redshift and M_{200} is the mass enclosed within R_{200} .

2.3. Galaxy sample

We select member satellite galaxies associated with the 136 clusters from the redshift $z = 0$ snapshot, excluding central galaxies from our analysis. Galaxies that have undergone a temporary infall, meaning that they passed through a cluster once and subsequently exited, are also excluded. To ensure that the galaxies are well resolved, only those with stellar masses $M_* > 10^9 M_\odot$ are preserved. After applying these criteria, the sample comprises a total of 10 087 galaxies.

2.4. R – V diagram

To construct the observable R – V diagram from simulation data, we aligned the LOS direction with the x , y , and z axes in turn, thereby tripling the sample size to 408 clusters and 30 261 galaxies. The LOS velocity is the component of velocity along the line of sight. The projected radius is defined as the distance from a galaxy to the central galaxy on the projection plane. For example, when taking x axis as LOS, the LOS velocity is v_x , and the projected radius is:

$$R_{2D} = \sqrt{(y - y_0)^2 + (z - z_0)^2}, \quad (1)$$

where the sub-script 0 denotes the coordinates of the central galaxy. To stack all clusters together, the projected radii are normalised by R_{200} of host clusters, and LOS velocities are normalised and the LOS velocity dispersions (σ_{los}), which are the standard deviations of the LOS velocities of member galaxies. In the mean while, the Hubble flow correction is implemented as follow:

$$V_{\text{los}} = |v_i + H_0 \times r_i|, \quad (2)$$

where i represents the specific LOS direction (x , y , or z), while v_i and r_i are the corresponding components of cluster-centric velocity and radial vectors, respectively.

Moreover, the 3D phase space (hereafter referred to as ‘phase space’) serves as a reference for understanding the behaviour of galaxies. It is defined by the normalised 3D radius (r_{3D}/R_{200}) and the radial velocity (v_{3D}/σ_{3D}).

For brevity, in the rest of this paper, we use lowercase r and v to denote r_{3D}/R_{200} and v_{3D}/σ_{3D} , and uppercase R and V to denote R_{2D}/R_{200} and $V_{\text{los}}/\sigma_{\text{los}}$, respectively.

3. Infall time distribution in the R – V diagram

3.1. Infall process

Tracing the historical positions and velocities of the member galaxies along their main merger trees, we illustrate their overall infall process in Fig. 1. Here, the galaxy number distributions are shown in both the phase space and the R – V diagram at five look-back times.

A clear infall trajectory is visible in the phase space (upper panels of Fig. 1). Galaxies initially fall from the outskirts (right) to the centre (left), with radial velocities directed inwards (negative). Upon reaching the pericentre, the radial velocities become positive as the galaxies move outwards. Then, the galaxies splash back to the outer region, and their velocities approach zero near the apocentre. This cycle repeats, with orbital radii gradually shrinking until the galaxies eventually become virialised in

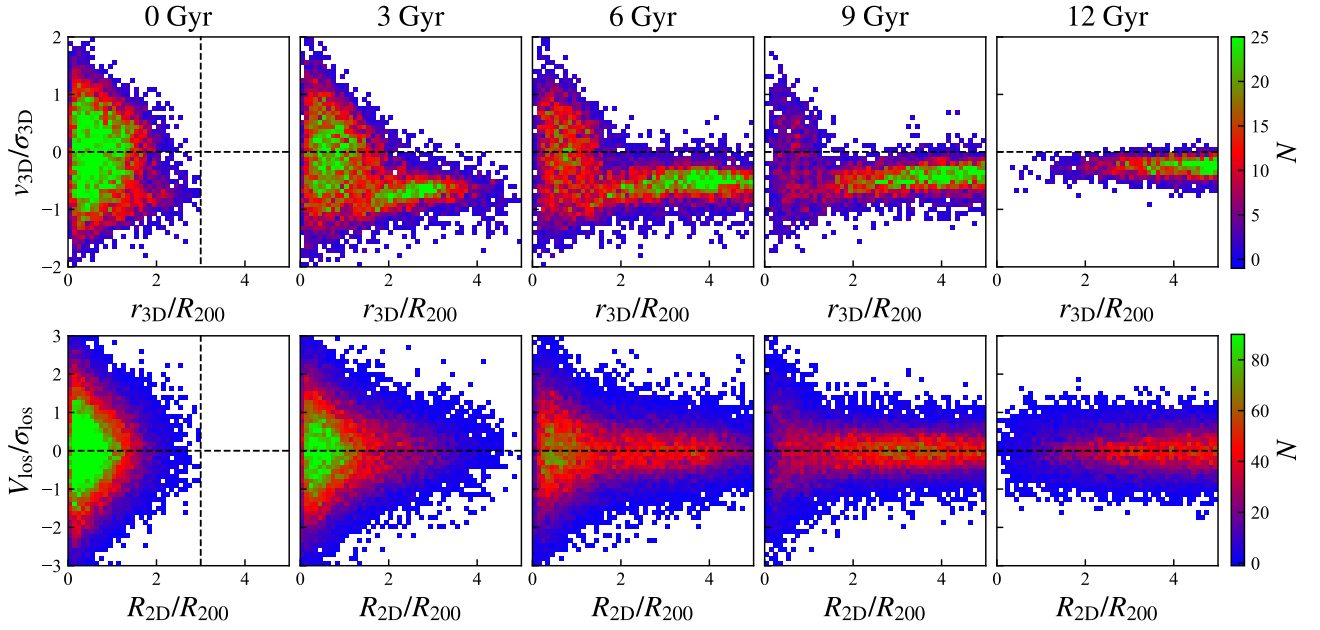


Fig. 1. Number distributions of the galaxies in the phase space and the $R - V$ diagram at five different look-back times. The vertical lines in the leftmost panels represent the $3R_{200}$ boundary for member galaxies.

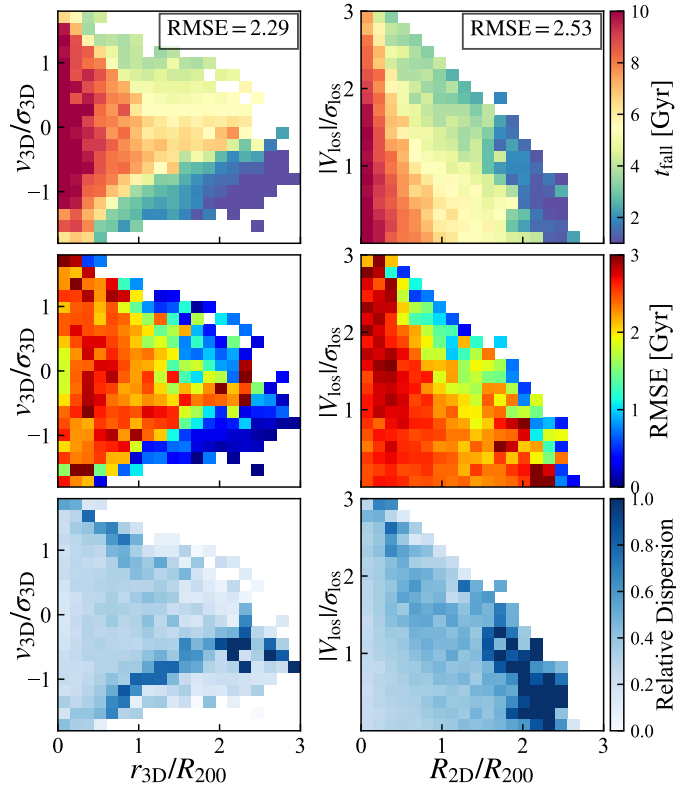


Fig. 2. Top: Median t_{fall} distributions in the phase space (left) and the $R - V$ diagram (right). Middle: Dispersion distributions, defined as RMSE relative to the median in each pixel. Bottom: Relative dispersion distribution, defined as the ratio of RMSE_{pix} to the median t_{fall} .

the core region. In the $R - V$ diagram (lower panels), a similar pattern is observed, although the distinction between the infalling and virialised regions is less pronounced due to pro-

jection effects. This infall process is consistent with previous studies (e.g. Mahajan et al. 2011; Oman et al. 2013; Rhee et al. 2017).

Moreover, the galaxy distribution in the $R - V$ diagram is symmetric about the x -axis. Assuming that galaxies are isotropically distributed around the cluster, each velocity component should follow a Gaussian profile centred on the average velocity. To simplify the analysis, we only apply the absolute value $|V|$ in the $R - V$ diagram in the remainder of this paper.

3.2. The infall time

We define the infall time (t_{fall}) of a galaxy as the look-back time when it first crosses $3R_{200}$ of the cluster. This boundary is sufficiently large to assign an infall time to each member galaxy and to ensure that no galaxy orbit exits. The specific choice of boundary primarily affects the absolute value of t_{fall} (e.g. when adopting R_{200} as the boundary, the t_{fall} values will be ~ 2 Gyr smaller) but has little impact on the relative relationship between t_{fall} and the position in the $R - V$ diagram. Therefore, slight variations in the definition of t_{fall} do not alter our overall results and conclusions.

The median t_{fall} distributions in both the phase space and the $R - V$ diagram are shown in the upper panels of Fig. 2. It is clear that galaxies with velocities and radii closer to zero tend to have larger t_{fall} values. In the phase space, galaxies with $t_{\text{fall}} < 2$ Gyr are concentrated in a region characterised by large radii and negative radial velocities. This finding is consistent with the infall trajectories shown in Fig. 1. We refer to these galaxies as ‘recent-infall galaxies’. In the $R - V$ diagram, recent-infall galaxies are also concentrated in the outer region with large radii.

To evaluate the accuracy of using the median values in pixels to estimate true infall times, we calculate the overall root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} (t_i - t_{\text{median,pix}})^2}, \quad (3)$$

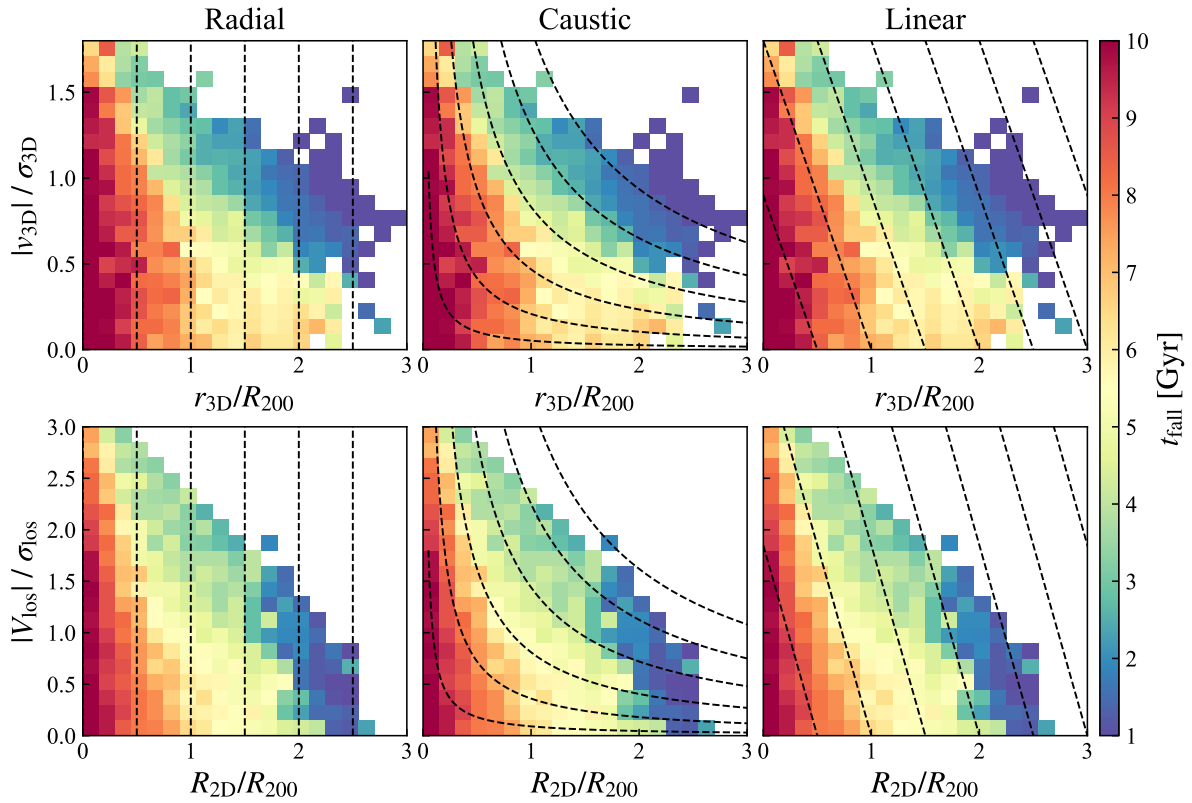


Fig. 3. Median t_{fall} distributions in the folded phase space (top) and the $R - V$ diagram (bottom). The three columns and corresponding dashed curves represent the projected radii, caustic profiles, and linear partitions, respectively.

where N_{tot} is the total number of galaxies, t_i represents the true infall times of individual galaxies, and $t_{\text{median,pix}}$ is the median infall time of galaxies in the corresponding pixel. The overall RMSE values reflect the fundamental dispersion in estimating infall times via kinematic data. As shown in the upper panels of Fig. 2, the dispersion in the $R - V$ diagram (2.53 Gyr) is greater than that in the phase space (2.29 Gyr), primarily because of the projection effect.

Then, we analyse the detailed dispersion distribution by calculating RMSE_{pix} , which is similar to RMSE but averages the error of galaxies in individual pixels:

$$\text{RMSE}_{\text{pix}} = \sqrt{\frac{1}{N_{\text{pix}}} \sum_{i=1}^{N_{\text{pix}}} (t_i - t_{\text{median,pix}})^2}, \quad (4)$$

where N_{pix} is the number of galaxies in the corresponding pixel. The resulting RMSE_{pix} distributions are displayed in the middle panels of Fig. 2. Consistent with the overall RMSEs, the dispersion is greater than 2 Gyr in most regions and even exceeds 3 Gyr in the inner region of the $R - V$ diagram. Only the region of recent-infall galaxies in the phase space exhibits low dispersion of $\text{RMSE} < 1$ Gyr.

We also examine the relative dispersion, defined as the ratio of RMSE_{pix} to the median t_{fall} , as shown in the bottom panels of Fig. 2. In the phase space, recent-infall galaxies exhibit low relative dispersions, as they have low absolute dispersions. The largest relative dispersions exist in a transition region between recent-infall galaxies and inner galaxies, corresponding to the region with a median t_{fall} of approximately 2–4 Gyr. In the $R - V$ diagram, recent-infall galaxies suffer from the largest relative dispersions because of their high absolute dispersions and small infall times.

4. Evaluating methods for estimating infall times

Here, we compare several methods for estimating infall times in the $R - V$ diagram. They can be divided into two categories: continuous tracers and discrete zones.

4.1. Continuous methods

There are two commonly used continuous methods. The projected radius (R) is the most convenient tracer, as it can be easily measured without spectroscopic observations. Although R is not typically used as a direct tracer of infall time due to the severe projection effect, it remains an effective and widely adopted tool for tracing environmental variations (e.g. Mahajan et al. 2011; Taranu et al. 2014; Haines et al. 2015; Maier et al. 2019; Lopes et al. 2023). Another one is the caustic line. Theoretically, the infall pattern in the phase space associated with a spherically symmetric density perturbation can be described by the caustic profile, which is nearly hyperbolic (Regos & Geller 1989; Diaferio & Geller 1997). Thus, many works have used the hyperbolic caustic profiles in $R - V$ diagrams to quantify the infall process of galaxies (e.g. Haines et al. 2012; Noble et al. 2013, 2016; Kim et al. 2023). Here, we introduce a continuous tracer, the caustic distance, which is defined as:

$$d_{\text{caustic}} = \sqrt{R \times V}. \quad (5)$$

The reference curves of the projected radii and caustic profiles are plotted in the left and middle panels of Fig. 3, respectively.

In addition to these two methods, another natural choice is the linear partition, which is similar to R but uses oblique lines, as illustrated in the right panels of Fig. 3. To quantify the linear

partitions, such as R and d_{caustic} , we introduced a new continuous tracer, the linear distance (d_{linear}), which is defined as the perpendicular distance from the origin to the oblique lines:

$$d_{\text{linear}} = \frac{|V| - kR}{\sqrt{1 + k^2}}, \quad (6)$$

where k is the slope of these lines. To determine the optimal slope, we test a range of slopes and use the Spearman rank correlation coefficient, ρ (Spearman 1904), to examine the monotonic correlation between t_{fall} and d_{linear} . A ρ closer to ± 1 indicates a stronger monotonic correlation. The Spearman coefficients as functions of slopes are illustrated in Fig. 4, with minima indicated by vertical lines. The optimal slopes are $k_{3\text{D}} = -1.8$ for the phase space ($\rho = -0.6$) and $k_{2\text{D}} = -3.7$ for the $R - V$ diagram ($\rho = -0.53$).

We also explored the dependence of the optimal slopes on galaxy stellar mass. Using the median stellar mass of $10^{9.8} M_{\odot}$ as the threshold, we divided all galaxies into two sub-samples: low-mass and high-mass. The corresponding Spearman correlation coefficients are plotted against slopes as blue and orange curves in Fig. 4. Generally, the low-mass sub-sample exhibits a relatively weaker monotonic correlation and favours a shallower slope. This trend is similar in both the phase space and the $R - V$ diagram. However, the differences in optimal slopes between the two sub-samples are small and have very little effect on the resulting correlation coefficients. Therefore, we adopted the slopes derived from the full galaxy sample, $k_{3\text{D}} = -1.8$ and $k_{2\text{D}} = -3.7$, for calculating d_{linear} in the remainder of this study.

We carried out a preliminary comparison of the three methods by overlaying their reference curves on the median t_{fall} distributions. This is shown in Fig. 3, where the phase space is folded to facilitate the application of these methods. The vertical lines of 3D radius, r , and projected radius, R , fail to trace the variation in t_{fall} due to the projection effect. The caustic profiles align well with the t_{fall} pattern in the phase space, especially for recent-infall galaxies. However, this alignment weakens in the $R - V$ diagram, where the pattern is slightly different from that in the phase space. Moreover, the inner caustic profiles are contaminated by some outer galaxies because they extend infinitely along the axes. In comparison, the linear partition better captures the variation in t_{fall} than does R , while also avoiding the inclusion of outer galaxies in the inner regions, such as caustic profiles.

We then quantitatively evaluate these methods in Fig. 5, where t_{fall} is plotted against the three tracers (R , d_{caustic} , and d_{linear}). The Spearman coefficients between t_{fall} and the tracers are given in the legends. We divided the tracers into 20 bins. Here, the median t_{fall} values of these bins are plotted as red curves, with error bars representing the 16th and 84th percentiles. The overall RMSEs relative to these medians are also provided in the legends.

In the phase space (upper panels of Fig. 5), the radius r exhibits a stronger correlation with t_{fall} ($\rho = -0.55$) but also greater dispersion (2.52 Gyr) than those of $d_{\text{caustic},3\text{D}}$ ($\rho = -0.53$, RMSE = 2.45 Gyr). The small error bars of the caustic method in the outer region with $d_{\text{caustic},3\text{D}} > 1$ highlight its effectiveness in estimating the t_{fall} values of recent-infall galaxies. In addition, the median t_{fall} trend of r exhibits some nonlinearity between $r = 1$ and 2. This arises because galaxies typically experience multiple orbits before virialisation, during which their orbital radii fluctuate repeatedly rather than decrease monotonically. In comparison, the median t_{fall} trend of $d_{\text{caustic},3\text{D}}$ is more linear, because caustic profiles closely follow the typical infall trajectories of galaxies. The linear partition has the strongest correlation ($\rho = -0.6$) and smallest dispersion (2.38 Gyr). Nonetheless,

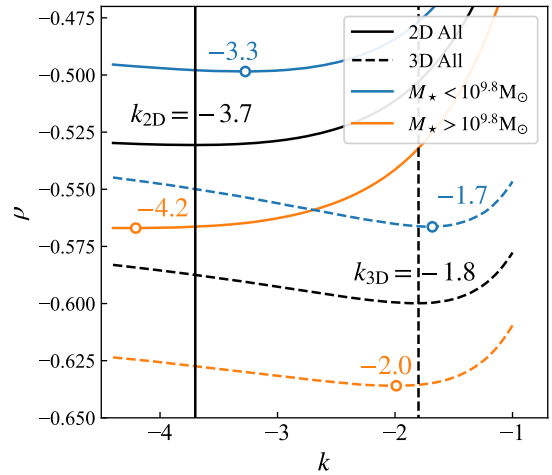


Fig. 4. Spearman coefficients, ρ , between the infall time, t_{fall} , and the linear distance, d_{linear} , as functions of the slope, k . Solid and dashed curves show results in the $R - V$ diagram and phase space, respectively. Black curves represent the full galaxy sample, with vertical lines marking the optimal k where ρ reaches minima. Blue and orange curves represent the low-mass and high-mass galaxies, respectively, with corresponding minima indicated by open circles.

its median t_{fall} trend still shows some nonlinearity, which is less pronounced than that of r but not fully eliminated as $d_{\text{caustic},3\text{D}}$.

In the $R - V$ diagram, all three methods show weaker correlations with t_{fall} and larger dispersions due to the projection effect. Among them, the caustic method is impacted the most severely. Its correlation remains lower ($\rho = -0.44$) than that of R ($\rho = -0.48$), but its dispersion rises significantly to 2.69 Gyr, even exceeding that of R (2.64 Gyr). The outer region with $d_{\text{caustic},2\text{D}} > 1$ is especially affected according to the error bars. This is expected because the pattern of recent-infall galaxies in the $R - V$ diagram no longer closely follows the caustic profiles as it does in the phase space (see Fig. 3). In contrast, the oscillatory patterns and non-linear median trends of the other two methods remain similar to those in the phase space. The linear partition, in particular, still achieves the strongest correlation ($\rho = -0.53$) and the smallest dispersion (2.56 Gyr). Therefore, despite its simple and empirical nature, the linear partition offers a better balance between the projected radii and caustic profiles.

4.2. Discrete methods

Moving beyond continuous tracers, two studies have attempted to manually divide the $R - V$ diagram into discrete zones, each associated with characteristic infall times. Rhee et al. (2017, hereafter R17) used zoom-in hydrodynamic simulations to explore the typical path of galaxies in the $R - V$ diagram, as they infall into the cluster potential. They divided the $R - V$ diagram into five distinct zones by maximising the fraction of galaxies belonging to a specific infall population in each zone. This method has been validated by several studies as effective in tracing galactic evolution in cluster environments (e.g. Brambila et al. 2023; Ding et al. 2023). Pasquali et al. (2019, hereafter P19) proposed an eight-zone discretisation method based on the mean infall time distribution in the $R - V$ diagram. Their quadratic curves outperformed both the caustic profiles and the projected radius, which is why they have been widely adopted (e.g. Sampaio et al. 2021, 2024; Oxland et al. 2024).

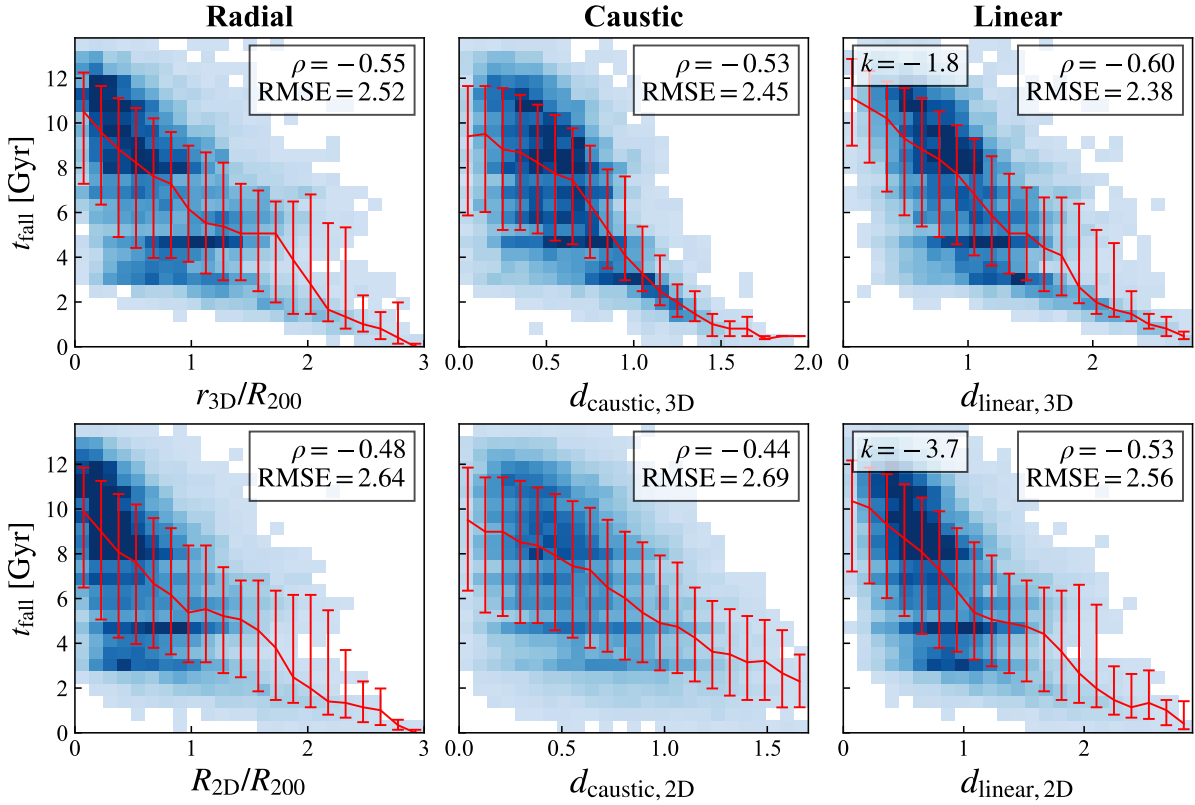


Fig. 5. Relationships between t_{fall} and the three tracers. The red curves indicate the median t_{fall} in 20 bins, with error bars indicating the 16th and 84th percentiles. The Spearman coefficient, ρ , between t_{fall} and each tracer is written in the corresponding upper right legend, together with the overall RMSE relative to the bin median. The slopes for calculating d_{linear} are given in the upper left corners.

Here, we compare the two discrete methods together with three continuous methods in Fig. 6. To facilitate the comparison, we discretised the continuous tracers into eight equal-width bins between the ranges of $0 < R < 2.6$, $0 < d_{\text{caustic}} < 1.5$, and $0 < d_{\text{linear}} < 2.6$, as illustrated by the reference curves in the upper panels. The P19 and R17 discretisations were adjusted from their original formulations, as their projected radii are normalised by R_{vir} , whereas we used R_{200} ($R_{200} = 0.73R_{\text{vir}}$, assuming an NFW profile with concentration of $c = 4$, Reiprich et al. 2013). The t_{fall} distributions of these zones are shown as violin plots in the lower panels, where red dots represent the medians and error bars represent the 16th and 84th percentiles. The overall RMSEs relative to these medians are listed in the upper right corners.

Considering the median t_{fall} trends, all methods exhibit monotonically decreasing trends from the inner to outer zones, except for R17, which shows a slightly inverted trend in the B – D zones. This may be because R17 aims to maximise the proportion of a specific population in each zone, rather than consider the overall trend of the median infall time. Additionally, while most median t_{fall} trends decrease to < 2 Gyr in the outermost zone, the trend of P19 flattens in the outer zones (No. 6–8) at ≈ 4 Gyr. This is because P19 excluded the region outside R_{vir} to avoid interlopers. The fractions of galaxies covered by each method are shown in the upper left corners of the bottom panels. While other methods cover nearly all member galaxies, P19 excluded $\sim 10\%$ of them, which are mainly recent-infall galaxies with $t_{\text{fall}} < 2$ Gyr. Therefore, this exclusion results in a flattening median t_{fall} trend.

For the dispersions, the approach from P19 and the linear partition as well exhibit the smallest dispersion of ≈ 2.6 Gyr, which is approximately 0.1 Gyr smaller than those of the other

three methods. This difference is fairly small compared with the dispersions themselves. As discussed in Section 5.3, the accuracy of estimating the infall time in the R – V diagram is limited by the intrinsic orbital overlap issue; this is the case regardless of whether complex discretisation, theory-based caustic lines, or a simple linear partition is used.

Additionally, the fractions of galaxies covered by each method are shown in the upper left corners of the bottom panels. While other methods cover nearly all member galaxies, P19 ignored $\sim 10\%$ of them because only the region within R_{vir} was covered to avoid interlopers. The ignored galaxies are mainly the recent-infall galaxies with $t_{\text{fall}} < 2$ Gyr. Therefore, the median t_{fall} trend of P19 flattens in the outer zones (No. 6–8) at ≈ 4 Gyr, while the median trends of other methods decrease to < 2 Gyr in the outer most zone.

5. Discussion

Since estimating the infall times of galaxies with their positions in the R – V diagram is limited by large intrinsic dispersion, we chose to explore two potential approaches to improve the accuracy. We also discuss the primary factors contributing to the dispersion.

5.1. The dynamic states of clusters

The spatial distribution and motion of member galaxies are closely associated with the dynamic state of the host clusters (e.g. Dressler & Shectman 1988; Wen & Han 2013). Non-relaxed or merging clusters typically exhibit a greater number

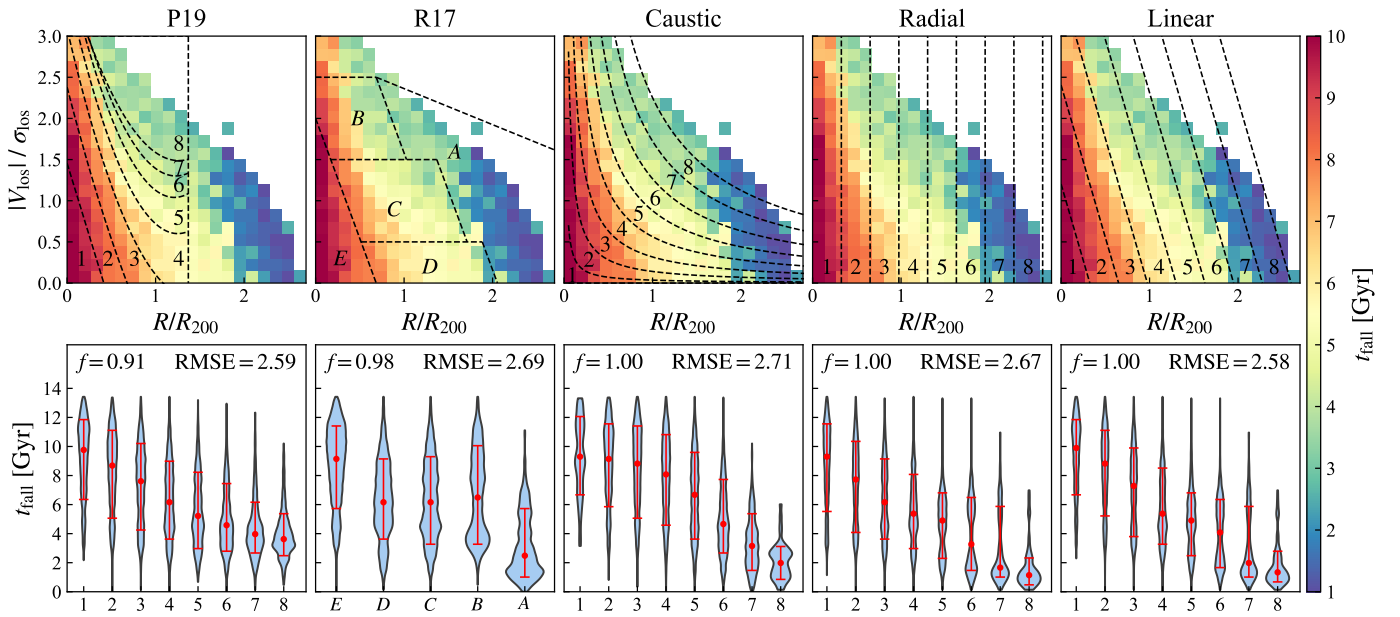


Fig. 6. Top: Reference curves for each method plotted over the median t_{fall} distribution in the $R - V$ diagram. Bottom: t_{fall} distributions of individual zones. The red dots represent the medians, and the error bars represent the 16th and 84th percentiles. The overall RMSE values relative to the zone medians are written in the upper right corners. The fraction of galaxies covered by each method is written in the upper left corner of the corresponding panel.

of infalling galaxies, wherein the motions are significantly disturbed. In contrast, relaxed and virialised clusters have more regular structures. Their member galaxies are concentrated towards the centres and have stable motions. Therefore, we want to explore how the dynamic state of clusters affects the estimation of infall time in the $R - V$ diagram.

We employ the ratio of the virial mass M_{vir} to M_{200} as an indicator of a cluster's dynamic state. The virial mass is calculated according to the virial theorem:

$$M_{\text{vir}} = \frac{3R_{200} \sigma_{\text{los}}^2}{G}. \quad (7)$$

The parameters M_{200} , R_{200} , and σ_{los} are all available in the observations. A virialised cluster has a theoretical mass comparable to its real mass, that is, $M_{\text{vir}}/M_{200} \lesssim 1$, whereas a less relaxed cluster has a larger M_{vir}/M_{200} . The M_{vir}/M_{200} distribution of our clusters is shown in Fig. 7. We divide the cluster sample into three equally sized sub-sets, with the thresholds indicated by the vertical dashed lines.

The median t_{fall} and RMSE_{pix} distributions in the $R - V$ diagram of the three cluster samples are shown in Fig. 8, with the overall RMSE values indicated in the legends. As expected, the dispersion clearly tends to increase with increasing M_{vir}/M_{200} . In the most relaxed sample with $M_{\text{vir}}/M_{200} < 1.01$, the intrinsic dispersion is 2.44 Gyr, which is 0.13 Gyr smaller than that of the least relaxed sample ($M_{\text{vir}}/M_{200} > 1.16$) and 0.09 Gyr smaller than that of all clusters together (Fig. 2). This trend is also evident in the detailed RMSE_{pix} distributions. The area with dispersion exceeding 3 Gyr increases in sequence.

We then apply the five methods for estimating infall times to the most relaxed sample, as shown in Fig. 9. The features of these methods remain consistent with their application to the entire galaxy sample (Fig. 6), but the accuracies are improved by ≈ 0.1 Gyr, in agreement with the 0.09 Gyr improvement in overall RMSE. Therefore, in observational studies, focusing

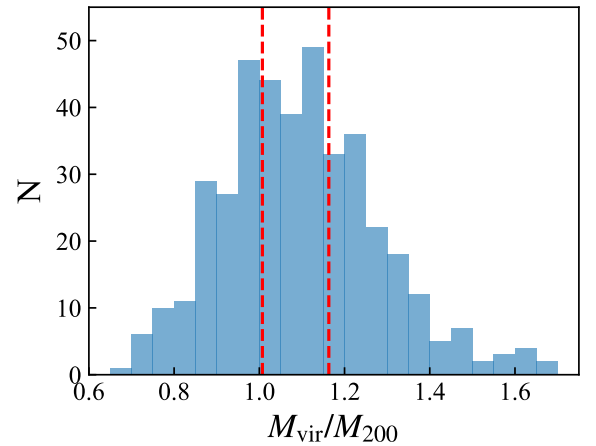


Fig. 7. M_{vir}/M_{200} distribution of 408 clusters. The vertical dashed lines represent the criteria for dividing clusters into three equally sized sub-samples.

on relaxed and virialised clusters may help reduce the dispersion in estimating infall times. Other dynamical indicators (e.g. Vallés-Pérez et al. 2025) may better probe the dynamical states of clusters. However, considering the RMSE difference of only 0.13 Gyr between our least and most relaxed samples shown in Fig. 8, more precise classification of clusters might not significantly improve the accuracy of infall time estimation.

5.2. Two estimates

When inspecting the t_{fall} distributions in Fig. 6, we observe a universal multimodality across them. A similar phenomenon has also been discussed in previous studies (e.g. Oman et al. 2013). Therefore, employing the median value as a single estimate for infall time would be inherently inaccurate, whether in pixels or

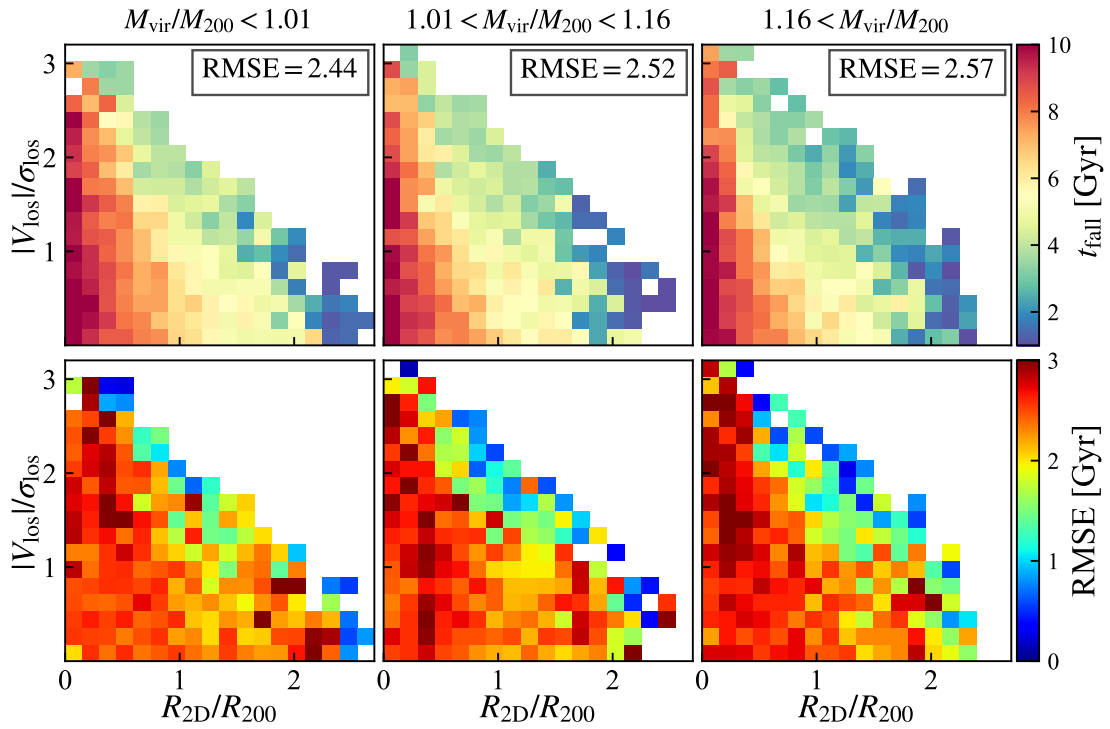


Fig. 8. Median t_{fall} (top) and RMSE_{pix} (bottom) distributions of three cluster samples in the $R - V$ diagram. The M_{vir}/M_{200} range of each sample is labelled in the titles. The overall RMSEs are written in the legends of the top panels.

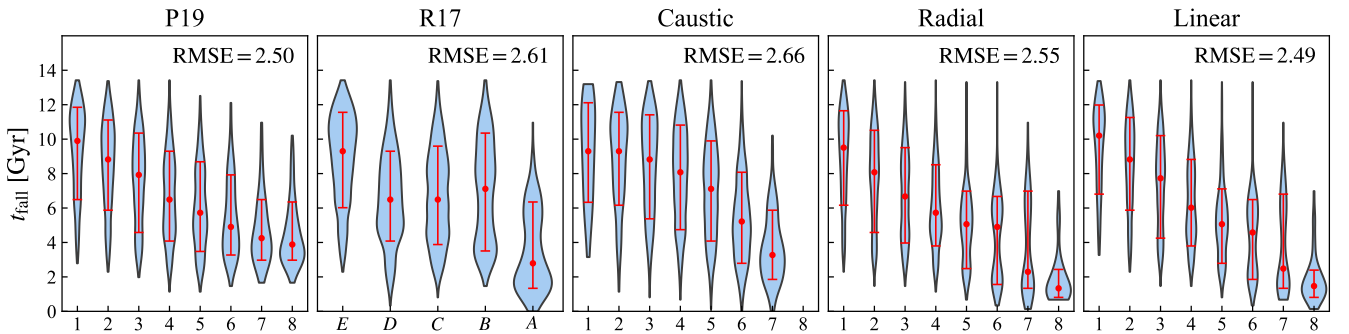


Fig. 9. Five methods applied to the most relaxed cluster sample, which have $M_{\text{vir}}/M_{200} < 1.01$. The elements are the same as those in Fig. 6. Note that no galaxy is located in the 8th zone of the caustic profiles; thus, it is empty.

in zones. Taking the linear partition as an example, the t_{fall} distributions in eight zones are plotted in Fig. 10, where the black dashed lines indicate the median infall times. It is clear that in most zones, the median value deviates from any local density peak. The specific zone ranges and median values are provided in Table 1.

Given this multimodality, it is reasonable to consider using multiple estimates rather than a single median value. For simplicity, we employ two estimates per zone. We employed the Gaussian mixture model (GMM) to model the t_{fall} distribution in each zone, with the component number fixed as two, implemented with the Scikit-learn (Pedregosa et al. 2011) package in Python. The fitted Gaussian components are presented as the red and blue curves in Fig. 10, with peaks indicated by dashed lines. In most zones, the two peaks provide a better match to the local density peaks in the t_{fall} distributions than the median value does. The two peaks serve as the two estimates of t_{fall} in each zone, which are listed as t_1 and t_2 in Table 1.

We note that two Gaussian components do not perfectly capture the multimodal nature of the t_{fall} distributions. For example, at least three peaks are visible in Zone 5. However, it is more convenient and practical in observations to employ two rather than more estimates. By considering various physical properties of a galaxy (e.g. colour, the star formation rate, and the gas fraction), it is possible to roughly determine whether it should have a younger or older infall time. However, dividing galaxies into more sub-samples would be tricky and unreliable.

Assuming that we can reliably divide galaxies into earlier-infall and later-infall populations, we calculate the dispersions when using two estimates. For each galaxy, we compared its two errors relative to both estimates and we used the smaller one to calculate RMSE, which is listed as RMSE_{two} in Table 1. Compared with the RMSE when the median value is used as the single estimate ($\text{RMSE}_{\text{median}}$), the use of two estimates significantly improves the accuracy to $\lesssim 1.5$ Gyr, even reaching ~ 1 Gyr in the outer sixth to eighth zones. This improvement clearly demon-

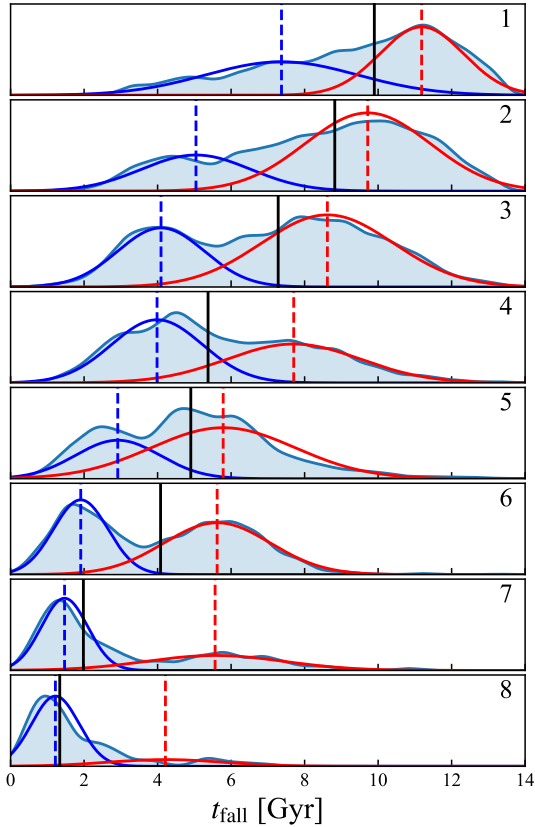


Fig. 10. Infall time distributions in the eight zones of the linear partition method. The zone numbers are indicated in the upper right corners. The blue shadings represent the t_{fall} distributions constructed via Gaussian kernel density estimation with a kernel size of 0.3 Gyr. The black vertical lines indicate the median t_{fall} in each zone. The blue and red curves are the two components obtained through the GMM, with the blue and red dashed lines indicating their peak locations.

states the advantage of using two infall time estimates in each zone over using one.

5.3. Orbital overlap

Throughout this paper, we highlight the large intrinsic dispersion of t_{fall} in the $R - V$ diagram. However, we need to address the cause behind this dispersion.

During a single-infall orbit, there is a nearly one-to-one correspondence between the t_{fall} of a galaxy and its position in the $R - V$ diagram. However, galaxies experience multiple orbits before being virialised, which is indicated by the oscillation of r with t_{fall} , as shown in Fig. 5. Consequently, galaxies in different orbits, of course with different t_{fall} , overlap with each other in the $R - V$ diagram. In other words, each position in the $R - V$ diagram corresponds to a complex mixture of galaxies in multiple orbital phases, naturally leading to the multimodality of the t_{fall} distribution discussed in Section 5.2. Here, we further illustrate the orbital overlap issue.

First, the overall t_{fall} distribution for all the galaxies is also multimodal. We employed the GMM to model the overall t_{fall} distribution, and this time, we freed the component number and determined the optimal model through the Bayesian information criterion (BIC). The best model has six components and the results are shown in the top panels of Fig. 11. The grey histogram

shows the overall t_{fall} distribution and the six dashed curves represent six Gaussian components.

Moreover, we divide galaxies into six populations on the basis of their orbital phases at $z = 0$:

- (a) Before the first pericentre;
- (b) From the first pericentre to the 1st apocentre;
- (c) From the first apocentre to the 2nd pericentre;
- (d) From the second pericentre to the 3rd pericentre;
- (e) From the third pericentre to 5th pericentre;
- (f) After the fifth pericentre.

These populations are arbitrarily determined to ensure that their t_{fall} distributions (shown as the red histograms in the top panels of Fig. 11) roughly align with the six Gaussian components. While this alignment is not exact, it still reflects the mixture of different orbital populations in the overall galaxy sample. The number distributions, median t_{fall} distributions, and RMSE_{pix} distributions of these six populations in the $R - V$ diagram are shown in the second, third, and fourth rows of Fig. 11, respectively.

Considering the number distributions, population (a) galaxies, which are in their first infalling orbits, are spread over the entire $R - V$ diagram. Some of them even sink into the innermost region rather than remaining in the outer regions. As the galaxies undergo subsequent orbits, the orbital radii shrink due to dissipation, and galaxies with longer orbital histories occupy the region with a smaller R value (see populations *d*, *e*, *f*). Therefore, the inner region of the $R - V$ diagram mixes more orbital populations than the outer region does, leading to greater dispersion in the inner region, as shown in Fig. 2. This also explains the reduced dispersion in more relaxed clusters (Section 5.1): relaxed clusters contain fewer population (a) galaxies, which occupy a larger area of the $R - V$ diagram.

For the median t_{fall} distributions in the third row of Fig. 11, the trend of population (a) remains consistent with that of all galaxies in Fig. 2: a larger t_{fall} in the inner region. However, the population (b) exhibits the opposite trend, with a larger t_{fall} at a larger R . These galaxies are moving outwards and are usually referred to as ‘splashback galaxies’. The opposite trends from these outwards-bound orbital phases further complicate the relationship between t_{fall} and the positions in the $R - V$ diagram. In the subsequent populations, the median t_{fall} varies little with position; this is likely due to the fact that these galaxies are close to virialisation and exhibit more isotropic motion.

Focussing on the dispersions, the overall RMSEs of these populations are reported in the legends of the third row of Fig. 11. Each population has a dispersion of <1.5 Gyr and, in particular, populations (a) and (b) have dispersions of ≈ 1 Gyr. Therefore, the overall dispersion of 2.53 Gyr results from the superposition of smaller dispersions of these orbital populations. The detailed RMSE_{pix} distributions in the bottom panels of Fig. 11 are fairly uniform across the $R - V$ diagram and their values are consistent with the overall RMSEs of the corresponding populations. Therefore, the complex patterns in the RMSE_{pix} distributions and relative dispersion distributions of all the galaxies (see Fig. 2) are also established by the superposition of these orbital populations.

These results highlight the dominant contribution of orbital overlap to the dispersion of the infall time estimation. The degeneracy in orbital phases remains challenging to resolve using observational data alone. As discussed in Section 5.2, incorporating the additional physical properties of galaxies may help partially mitigate this degeneracy, although further studies are needed to confirm its effectiveness.

Table 1. Infall time estimates in eight zones of the linear partition.

Zone	Range ^a	Median t_{fall} ^b	RMSE _{median} ^c	t_1 ^d	t_2 ^d	RMSE _{two} ^e	$f_1:f_2$ ^f	f_{intl} ^g
1	$0.000 < d_{\text{linear}} < 0.325$	9.90	2.57	7.37	11.18	1.40	0.47:0.53	0.00
2	$0.325 < d_{\text{linear}} < 0.650$	8.82	2.71	5.04	9.72	1.47	0.29:0.73	0.01
3	$0.650 < d_{\text{linear}} < 0.975$	7.28	2.70	4.09	8.62	1.44	0.35:0.65	0.03
4	$0.975 < d_{\text{linear}} < 1.300$	5.37	2.45	3.99	7.70	1.32	0.52:0.48	0.07
5	$1.300 < d_{\text{linear}} < 1.625$	4.90	2.20	2.92	5.78	1.43	0.31:0.69	0.18
6	$1.625 < d_{\text{linear}} < 1.950$	4.08	2.19	1.91	5.62	1.08	0.42:0.58	0.37
7	$1.950 < d_{\text{linear}} < 2.275$	1.98	2.58	1.47	5.56	1.09	0.62:0.38	0.60
8	$2.275 < d_{\text{linear}} < 2.600$	1.34	1.54	1.22	4.21	0.81	0.81:0.19	0.85

Notes. ^(a)The range of each zone. d_{linear} is calculated through Equation (6). ^(b)The median infall time in each zone, in Gyr. ^(c)The RMSE relative to the median infall time in each zone, in Gyr. ^(d)The two infall time estimates obtained by GMM, in Gyr. ^(e)The RMSE calculated with the smaller one of the two errors relative to the two estimates in each zone, in Gyr. ^(f)The relative proportion of two Gaussian components given by GMM. ^(g)The fraction of interlopers in each zone.

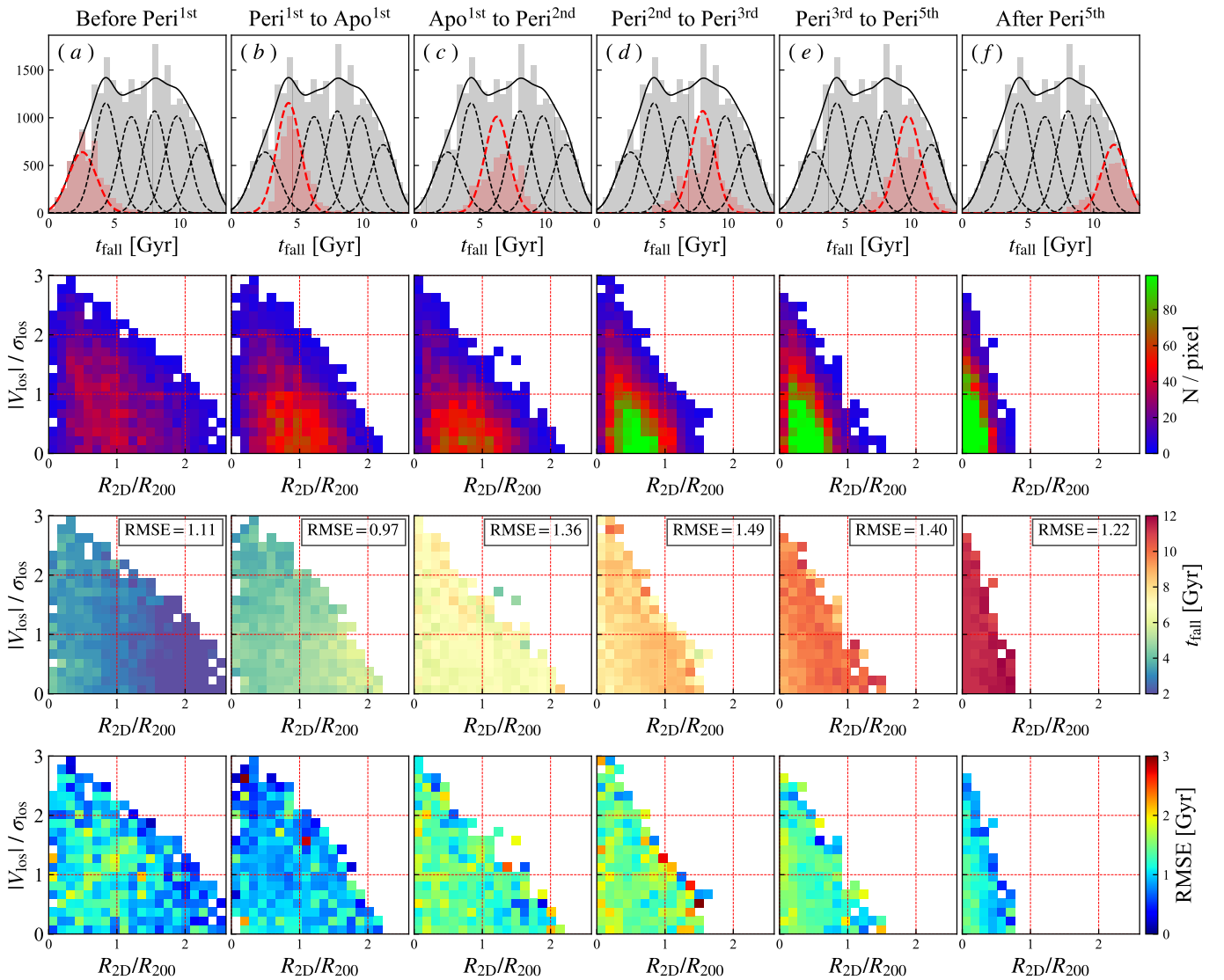


Fig. 11. Top panels: Infall time distributions. The grey histograms represent all the galaxies, and the dashed and solid curves represent the results of GMM fitting, which yields six Gaussian components. The t_{fall} distributions of the six orbital populations are shown as red histograms. These curves roughly align with the six Gaussian components, which are highlighted as red dashed curves one by one for visual clarity. Rows 2 to 4: Number distributions, median t_{fall} distributions, and RMSE_{pix} distributions in the $R - V$ diagram of the six orbital populations.

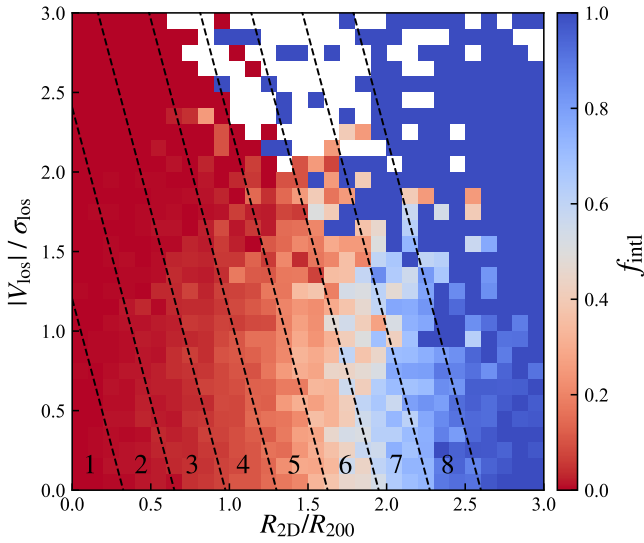


Fig. 12. Fraction of interlopers in the $R - V$ diagram. Reference lines of the linear partition method are plotted as dashed lines.

5.4. Other sources of dispersion

In addition to orbital overlap, several secondary sources contribute to dispersion. The projection effect plays a significant role. As shown in Fig. 1, the infall trajectory is clear in the phase space but is obscured in the $R - V$ diagram by the projection effect. All three continuous methods perform better when applied in the phase space than in the $R - V$ diagram, as illustrated in Fig. 5. In observational studies, measuring the distance and two projected velocity components is often extremely challenging, making the projection effect largely unsolvable.

Many galaxies infall into clusters as groups, which have their own internal dynamics (Diaferio 2009; Serra et al. 2011). These motions can introduce intrinsic and random noise in the $R - V$ diagram. However, with spectroscopic data and a sophisticated clustering algorithm, it is still possible to identify substructures within clusters and treat them separately (Yu et al. 2015; Qu et al. 2023).

5.5. Interlopers

In observations, the projection effect not only causes dispersions in infall time estimation, but also introduces interlopers that are not physically associated with the cluster, but lie within the $R - V$ diagram. Since our analysis in this paper is limited to true member galaxies, we aim to assess the potential impact of interlopers.

We define interlopers as galaxies that are not classified as members by TNG group catalogues but are located within $10R_{200}$ around our clusters. Galaxies beyond this distance have little chance of appearing in the location where the cluster members are seen in the $R - V$ diagram. Fig. 12 shows the fraction of interlopers in the $R - V$ diagram, along with the reference lines of the linear partition method. It is clear that interlopers mainly dominate the outer region with $R \gtrsim 2$, roughly around zone 7, zone 8, and farther outwards. Similar results have been reported in previous works (e.g. Oman & Hudson 2016). The interloper fractions in eight linear zones are listed as f_{intl} in Table 1. As expected from Fig. 12, interlopers constitute more than 50% of the galaxies in zones 7 and 8, whereas inner regions are dominated by true members. We therefore caution that our infall time

estimates in zone 7 and zone 8 may be unreliable in observation and should be interpreted with care.

6. Summary

Accurately estimating the infall time of galaxies infalling into clusters is crucial for understanding the environmental evolution of galaxies. However, achieving this estimation is challenging in observations. In this work, we utilised the TNG300-1 simulation to systematically evaluate five methods for estimating the infall time in the $R - V$ diagram, including the projected radii, caustic profiles, two discrete zoning methods, and a simple linear partition method.

These five methods exhibit distinct characteristics. Caustic profiles perform well for recent-infall galaxies in the phase space, but are less effective in the $R - V$ diagram because of the projection effect. The projected radii, R , which are also affected by the projection effect, perform better than the caustic profiles in the $R - V$ diagram. Of the two discretisations, P19 has the smallest dispersion (≈ 2.6 Gyr), although it excludes 10% of the infalling galaxies, whereas R17 shows slight non-monotonicity in the median t_{fall} trend. Compared with the other methods, the simple linear partition performs slightly better, showing the smallest dispersion, as that in P19, and a stronger correlation with infall time compared with the projected radii and caustic profiles.

However, the difference between the accuracies of the various methods is only ≈ 0.1 Gyr, which is negligible compared with the overall dispersion ($\gtrsim 2.6$ Gyr). In other words, these methods are all limited by the intrinsic dispersion of the infall time (2.53 Gyr), irrespective of their specific strategies for partitioning the $R - V$ diagram. Given this limit, we explored two potential approaches for improving the accuracy. First, more relaxed clusters exhibit less dispersion overall, allowing for more accurate estimations to be made. Second, considering the multimodality of the infall time distributions, using two estimates of the infall time in each zone (rather than a single median) improves the accuracy to $\lesssim 1.5$ Gyr. In observations, the appropriate choice between smaller or larger infall times could be guided by various physical properties of the galaxy. Both the single-estimate and two-estimate values and corresponding dispersions in eight zones of the linear partition are listed in Table 1, which could serve as a quantitative reference for future studies.

Finally, we demonstrate the orbital overlap issue that primarily contributes to the large intrinsic dispersion. The multimodal infall time distributions of all the galaxies are decomposed into six Gaussian components via the GMM. These components roughly align with six populations of galaxies in different orbital phases. The six orbital populations overlap with each other in the $R - V$ diagram, leading to orbital degeneracy among galaxies and consequently large dispersion in infall times. Additionally, secondary factors such as the projection effect and internal dynamics of infalling groups contribute to the overall dispersion.

Our results underscore the challenges and limitations of estimating infall times from observational data. Nevertheless, linear partitioning provides a simple and robust empirical framework for statistically linking the evolution of galaxies with their infall process. With upcoming surveys, such as DESI (Levi et al. 2013; DESI Collaboration 2016) and Euclid (Laureijs et al. 2011; Euclid Collaboration: Mellier et al. 2025), these methods will provide valuable insights into how the cluster environment influences galaxy evolution during the infall process.

Acknowledgements. We thank Lizhi Xie for providing constructive comments and valuable advice. We thank the IllustrisTNG collaboration for providing free and convenient access to the data used in this work, which can be accessed at www.tng-project.org. This work was made possible via several open-source software packages: AstroPy (Astropy Collaboration 2018), Matplotlib (Barrett et al. 2005), NumPy (van der Walt et al. 2011), Pandas (McKinney 2010), SciPy (Gommers et al. 2024) and Scikit-learn (Pedregosa et al. 2011).

References

- Alpaslan, M., Driver, S., Robotham, A. S. G., et al. 2015, *MNRAS*, 451, 3249
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123
- Barrett, P., Hunter, J., Miller, J. T., Hsu, J. C., & Greenfield, P. 2005, *ASP Conf. Ser.*, 347, 91
- Brambila, D., Lopes, P. A. A., Ribeiro, A. L. B., & Cortesi, A. 2023, *MNRAS*, 523, 785
- Chen, Y.-C., Ho, S., Mandelbaum, R., et al. 2017, *MNRAS*, 466, 1880
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371
- DESI Collaboration (Aghamousa, A., et al.) 2016, ArXiv e-prints [arXiv:1611.00036]
- Diaferio, A. 2009, ArXiv e-prints [arXiv:0901.0868]
- Diaferio, A., & Geller, M. J. 1997, *ApJ*, 481, 633
- Ding, Y., Zhu, L., van de Ven, G., et al. 2023, *A&A*, 672, A84
- Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, *MNRAS*, 399, 497
- Dressler, A. 1980, *ApJ*, 236, 351
- Dressler, A., & Shectman, S. A. 1988, *AJ*, 95, 985
- Dressler, A., Oemler, A., Couch, W. J., et al. 1997, *ApJ*, 490, 577
- Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, 697, A1
- Gommers, R., Virtanen, P., Haberland, M., et al. 2024, <https://doi.org/10.5281/zenodo.10909890>
- Goto, T., Yamauchi, C., Fujita, Y., et al. 2003, *MNRAS*, 346, 601
- Haines, C. P., Pereira, M. J., Sanderson, A. J. R., et al. 2012, *ApJ*, 754, 97
- Haines, C. P., Pereira, M. J., Smith, G. P., et al. 2015, *ApJ*, 806, 101
- Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, *MNRAS*, 353, 713
- Kim, K. J., Bayliss, M. B., Noble, A. G., et al. 2023, *ApJ*, 955, 32
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
- Levi, M., Bebek, C., Beers, T., et al. 2013, ArXiv e-prints [arXiv:1308.0847]
- Lopes, P. A. A., Ribeiro, A. L. B., & Brambila, D. 2023, *MNRAS*, 527, L19
- Mahajan, S., Mamon, G. A., & Raychaudhury, S. 2011, *MNRAS*, 416, 2882
- Maier, C., Ziegler, B. L., Haines, C. P., & Smith, G. P. 2019, *A&A*, 621, A131
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, *MNRAS*, 480, 5113
- McKinney, W. 2010, in *Proceedings of the 9th Python in Science Conference*, eds. S. van der Walt, & J. Millman, 56
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 477, 1206
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 475, 624
- Noble, A. G., Webb, T. M. A., Muzzin, A., et al. 2013, *ApJ*, 768, 118
- Noble, A. G., Webb, T. M. A., Yee, H. K. C., et al. 2016, *ApJ*, 816, 48
- Oman, K. A., & Hudson, M. J. 2016, *MNRAS*, 463, 3083
- Oman, K. A., Hudson, M. J., & Behroozi, P. S. 2013, *MNRAS*, 431, 2307
- Oxland, M., Parker, L. C., de Carvalho, R. R., & Sampaio, V. M. 2024, *MNRAS*, 529, 3651
- Pasquali, A., Smith, R., Gallazzi, A., et al. 2019, *MNRAS*, 484, 1702
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, 12, 2825
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, 475, 648
- Qu, C.-G., Yu, H., Diaferio, A., Sohn, J., & Liu, D. 2023, *RAA*, 23, 085007
- Regos, E., & Geller, M. J. 1989, *AJ*, 98, 755
- Reiprich, T. H., Basu, K., Ettori, S., et al. 2013, *Space Sci. Rev.*, 177, 195
- Rhee, J., Smith, R., Choi, H., et al. 2017, *ApJ*, 843, 128
- Rhee, J., Smith, R., Choi, H., et al. 2020, *ApJS*, 247, 45
- Rodríguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, *MNRAS*, 449, 49
- Sampaio, V. M., de Carvalho, R. R., Ferreras, I., et al. 2021, *MNRAS*, 503, 3065
- Sampaio, V. M., de Carvalho, R. R., Aragón-Salamanca, A., et al. 2024, *MNRAS*, 532, 982
- Serra, A. L., Diaferio, A., Murante, G., & Borgani, S. 2011, *MNRAS*, 412, 800
- Shi, K., Malavasi, N., Toshikawa, J., & Zheng, X. 2024, *ApJ*, 961, 39
- Spearman, C. 1904, *Am. J. Psychol.*, 15, 72
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, 328, 726
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, 475, 676
- Taranu, D. S., Hudson, M. J., Balogh, M. L., et al. 2014, *MNRAS*, 440, 1934
- Taylor, E., Almaini, O., Merrifield, M., et al. 2023, *MNRAS*, 522, 2297
- Vallés-Pérez, D., Planelles, S., & Quilis, V. 2025, *A&A*, 699, A1
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Comput. Sci. Eng.*, 13, 22
- Wen, Z. L., & Han, J. L. 2013, *MNRAS*, 436, 275
- Yu, H., Serra, A. L., Diaferio, A., & Baldi, M. 2015, *ApJ*, 810, 37
- Zheng, Y., Xu, K., Jing, Y. P., et al. 2024, *ApJ*, 969, 129