Quenched fractions in the IllustrisTNG simulations: comparison with observations and other theoretical models

Martina Donnari [1]★ Annalisa Pillepich [1]★ Dylan Nelson [2,3] Federico Marinacci [4]
Mark Vogelsberger [5] and Lars Hernquist[6]

[1]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[2]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany*
[3]*Zentrum für Astronomie, Institut für Theoretische Astrophysik, Universität Heidelberg, Albert-Ueberle-Str 2, D-69120 Heidelberg, Germany*
[4]*Department of Physics and Astronomy 'Augusto Righi', University of Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy*
[5]*Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[6]*Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*

ABSTRACT

We make an in-depth comparison of the IllustrisTNG cosmological simulations with observed quenched fractions of central and satellite galaxies, for $M_{stars} = 10^{9-12}$ M$_\odot$ at $0 \le z \le 3$. We show how measurement choices [aperture, quenched definition, and star formation rate (SFR) indicator time-scale], as well as sample selection issues (projection effects, satellite/central misclassification, and host mass distribution sampling), impact this comparison. The quenched definition produces differences of up to 70 (30) percentage points for centrals (satellites) above $\sim 10^{10.5}$ M$_\odot$. At $z \gtrsim 2$, a larger aperture within which SFR is measured suppresses the quenched fractions by up to $\sim$50 percentage points. Proper consideration of the stellar and host mass distributions is crucial: Naive comparisons to volume-limited samples from simulations lead to misinterpretation of the quenched fractions as a function of redshift by up to 20 percentage points. Including observational uncertainties to theoretical values of M_{stars} and SFR changes the quenched fraction values and their trend and/or slope with mass. Taking projected rather than three-dimensional distances for satellites decreases the quenched fractions by up to 10 per cent. TNG produces quenched fractions for both centrals and satellites broadly consistent with observations and predicts up to $\sim$80 (90) per cent of quenched centrals at $z = 0$ ($z = 2$), in line with recent observations, and higher than other theoretical models. The quantitative agreement of TNG and Sloan Digital Sky Survey for satellite quenched fractions in groups and clusters depends strongly on the galaxy and host mass range. Our mock comparison highlights the importance of properly accounting for observational effects and biases.

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

1 INTRODUCTION

At low redshift, galaxies in the Universe can be classified into two main populations according to their star formation activity: 'star-forming' galaxies with ongoing star formation, young stellar populations, and blue colours on the one hand and 'quenched' galaxies that have ceased to form young stars, have older stellar populations, and redder colours on the other.

The balance between these two populations – or the quenched fraction – is a strong function of galaxy mass, environment, and cosmic time. Overall, the fraction of quenched galaxies has been shown to increase with galaxy stellar mass at all redshifts ($z \lesssim 3$; Muzzin et al. 2013; Darvish et al. 2017; Fang et al. 2018). In dense environments such as groups and clusters, low-mass satellites with $M_{stars} \lesssim 10^{10}$ M$_\odot$ are frequently quenched, due to environmentally driven processes such as the inability to accrete new gas (Larson, Tinsley & Caldwell 1980) combined with the tidal (Merritt 1983) and

ram-pressure (Gunn & Gott 1972; Poggianti et al. 2017) stripping of pre-existing gas reservoirs. At the same time, more massive galaxies with $M_{stars} \gtrsim 3 \times 10^{10}$ M$_\odot$ can quench regardless of environment (Wetzel, Tinker & Conroy 2012; Wetzel et al. 2013; Peng, Maiolino & Cochrane 2015; Davies et al. 2019). In this case, the physical mechanism most often invoked is feedback from a supermassive black hole or active galactic nuclei (Di Matteo, Springel & Hernquist 2005; Hopkins et al. 2005; Croton et al. 2006).

Although there exists a broad qualitative consensus on the quenched fractions of galaxies versus mass, environment, and redshift, several quantitative discrepancies remain across observational compilations (Haines, Gargiulo & Merluzzi 2008; Geha et al. 2012; Lin et al. 2014; Fossati et al. 2017; Wagner et al. 2017; Jian et al. 2018). This is due in part to the different approaches that are taken by various studies, particularly to separate quenched and star-forming galaxies. Depending on redshift and wavelength coverage, different tracers are available to derive star formation rates (SFRs) of galaxies (Speagle et al. 2014), and these inherit systematic biases (Leja et al. 2019). Different group finding algorithms lead to different environmental measures and classifications of central versus

★ E-mail: donnari.m@gmail.com (MD); annalisa.pillepich@gmail.com (AP)

satellite status (Yang et al. 2007; Campbell et al. 2015; Gao et al. 2020). Different surveys contain various selection effects, integrate galaxy star formation out to different radii and apply different dust corrections (Salim et al. 2007), and sample a diversity of galaxy and host halo mass distributions versus redshift (Pintos-Castro et al. 2019).

Comparison of observational data with theoretical models can shed light on the impact of the systematic uncertainties, assumptions, and biases (Donnari et al. 2019a; Katsianis et al. 2020). Modern large-volume cosmological hydrodynamical simulations such as Illustris, EAGLE, Horizon-AGN, IllustrisTNG, SIMBA, Romulus, and FABLE fundamentally aim to reproduce the observed dichotomy of the galaxy population, of both star-forming and quenched galaxies, through a physical quenching mechanism. As a result, they broadly recover the trend of increasing quenched fraction versus mass (Furlong et al. 2015; Donnari et al. 2019a, 2021). At the same time, they also self-consistently capture gas dynamical processes such as stripping, enabling a view into the quenching of satellites (Bluck et al. 2016; Bahé et al. 2017; Tremmel et al. 2019; Appleby et al. 2020; Donnari et al. 2021).

Recent semi-analytical models continue to refine the processes that generate the quenched galaxy population (Brennan et al. 2015; Croton et al. 2016; Baugh et al. 2019; Fontanot et al. 2020; Henriques et al. 2020), although – despite explicit calibration against observationally inferred quenched fractions – discrepancies remain (Asquith et al. 2018). This highlights the difficulty of producing a full galaxy population with consistent stellar mass functions and SFR distributions across cosmic epochs. Wang et al. (2018) and Ayromlou et al. (2021) compare the quenched fractions between the Munich SAM L-Galaxies (Henriques et al. 2015) and the EAGLE and IllustrisTNG simulations, respectively, quantifying differences that arise due to details of the quenching processes.

We focus herein on the IllustrisTNG simulations and the associated TNG physical model for galaxy formation (Weinberger et al. 2017; Pillepich et al. 2018a), making comparisons to observations as well as other theoretical models. The outcomes of the TNG simulations have been validated against a broad set of observations: the colour distribution of red versus blue galaxies at $z = 0$ (Nelson et al. 2018a), the distributions of metals in the intracluster medium (Vogelsberger et al. 2018), the galaxy mass–metallicity relation (Torrey et al. 2018), the galaxy size–mass relation across redshifts and for star-forming versus quenched galaxies (Genel et al. 2018), the O VI content of the circumgalactic medium (Nelson et al. 2018b), the dark matter fractions within massive galaxies (Lovell et al. 2018), the environmental dependences of the cold gas contents of satellites (Stevens et al. 2019), and the star formation main sequence (MS) and quenched fractions of central galaxies (Donnari et al. 2019a).

In our recent companion paper (Donnari et al. 2021), we present the diverse pathways that TNG galaxies take towards quenching within the hierarchical growth of structure. Having postponed observational comparisons there, in this paper we now aim to properly and robustly compare the outcome of the TNG simulations to empirical inferences of central and satellite quenched fractions as a function of stellar masses, halocentric distance, host halo mass, and redshift. We focus mainly on galaxies with stellar mass between 10^{9-12} M$_\odot$ residing in groups and clusters, and we primarily compare to the Sloan Digital Sky Survey (SDSS) data by creating mock catalogues following the procedures adopted in Wetzel et al. (2012, 2013).

This paper is organized as follows: In Section 2, we introduce the simulations and analysis methods. In Section 3, we show how different observational issues impact the inferred quenched fractions. In particular, we contrast several definitions of star-forming versus quenched galaxies, physical apertures, and SFR tracers, also considering observational uncertainties, host mass estimates, satellite selections, and central/satellite misclassification. A comparison to other theoretical models and observations is presented in Section 4, with a particular focus on the SDSS data. We conclude and summarize in Section 5.

2 METHODS AND DEFINITIONS

2.1 The IllustrisTNG simulations

In this work, we chiefly use the TNG100 and TNG300 cosmological magneto-hydrodynamical simulations of the IllustrisTNG[1] project (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018a; Pillepich et al. 2018b; Springel et al. 2018). These two volumes have side lengths of ∼100 and ∼300 Mpc, and mass resolutions (average gas cell mass or star particle mass) of $m_{baryon} \sim 1.4 \times 10^6$ and $\sim 1.1 \times 10^7$ M$_\odot$, respectively. The third and highest resolution volume is TNG50 (Nelson et al. 2019b; Pillepich et al. 2019), which simulates a smaller volume of ∼50 cMpc across, with roughly 15 times higher mass resolution, $m_{baryon} \sim 8.5 \times 10^4$ M$_\odot$. In this paper, we use TNG50 in comparison to other models and observations.

The TNG simulations are performed with the AREPO code (Springel 2010), which solves for gravity and ideal magnetohydrodynamics (Pakmor, Bauer & Springel 2011). They include a well-explored and well-validated physical model for the key galaxy formation physics. All details of this underlying galaxy formation model are described in the two TNG method papers (Weinberger et al. 2017; Pillepich et al. 2018a). In short, the simulations evolve dark matter, gas, stars, and supermassive black holes, accounting for the physical processes of gas radiation (heating, cooling, and the background radiation field), star formation (above a density threshold following the two-phase interstellar medium model of Springel & Hernquist 2003), stellar population evolution and chemical enrichment (tracking supernovae Types Ia, II, and AGB stars), galactic-scale winds from stellar feedback (see Pillepich et al. 2018a), supermassive black holes (formation, mergers, and gas accretion), and black hole feedback (combining a high accretion rate thermal mode with a low accretion rate kinetic mode; see Weinberger et al. 2017).

2.2 Central and satellite populations

Substructures corresponding to haloes and subhaloes are identified via the Friends-of-Friends (FoF; Davis et al. 1985) and SUBFIND algorithms (Springel et al. 2001; Dolag et al. 2009). In this work, we differentiate between galaxies residing in groups and clusters, from the general or 'field' population. Unless otherwise stated, we define the virial radius as R_{200c}, the radius within which the mean enclosed mass density is 200 times the critical density of the Universe. The total mass enclosed within this radius is M_{200c}, and in general we consider all haloes with $M_{200c} \geq 10^{12}$ M$_\odot$ to be hosts.

We refer to hosts with mass in the range of 10^{12-14} M$_\odot$ as groups, and to those above this mass as clusters. All galaxies residing within one virial radius of the centre of their parent host FoF halo are satellites, except for the single galaxy centred at the minimum of the gravitational potential, which is typically the most massive and labelled as the central. We also call 'central' the galaxies that are associated with the field population, i.e. those that are not

---

[1]http://www.tng-project.org

associated with any groups or clusters. Throughout this analysis, the central sample may include backsplash galaxies, namely galaxies that might have been satellites in the past but are not at the time of inspection.

At $z = 0$, TNG300, TNG100, and TNG50 have 3733, 182, and 24 groups and clusters above $M_{200c} \geq 10^{13}\,M_\odot$, respectively, and more than 39,000, 2800, and 1600 satellites with stellar mass $M_{stars} \geq 10^9\,M_\odot$. The most massive cluster in TNG300, TNG100, and TNG50 has a halo mass of $10^{15.2}$, $10^{14.6}$, and $10^{14.3}\,M_\odot$, respectively – see also table 1 of Donnari et al. (2021). The minimum stellar mass we consider in this paper is $M_{stars} \geq 10^9\,M_\odot$, corresponding to at least about 100, 1000, 1500 star particles in TNG300, TNG100, and TNG50, respectively.

2.3 Stellar masses and SFRs

We adopt throughout a common definition of galaxy stellar mass as the sum of all gravitationally bound stellar particles within twice the stellar half-mass radius. We note that this is not comparable to any observational choice and differences can be large in general, especially at the high-mass end (see Pillepich et al. 2018b, for a discussion on this). However, for the purposes of the quenched fractions, the stellar mass choice implies a somewhat negligible shift of the functions along the x-axis; see Appendix A.

We characterize galaxies by their 'instantaneous' SFR by taking, directly from the simulation output, the sum of the SFRs of all gas cells that are gravitationally bound and within $2R_{star,h}$, twice the stellar half-mass radius.

We also consider the SFR measured over different time-scales, measured indirectly from the mass of recently formed stellar particles within the same aperture. We consider time intervals of both 200 and 1000 Myr (see Section 3.1.2). We refer to other upcoming works the task of extracting SFRs from the simulated galaxies by mocking what is typically done in observations.

Finite numerical resolution implies that there is a minimum resolvable SFR. To properly handle this issue, we randomly assign an SFR value in the range of $(10^{-5}–10^{-4})\,M_\odot\,yr^{-1}$ to those galaxies with log (SFR) below the resolution limit of the simulation, which otherwise would have SFR ~ 0 (see Donnari et al. 2019a, for more details). The existence of such galaxies is of the essence for those authors who simply average SFRs instead of distinguishing between star-forming and quiescent galaxies in some manner. Even if their actual distribution is unknown, those galaxies would distribute well below the MS, thus populating the quenched region. Additionally, we note that we do not identify starburst or green-valley galaxies.

3 RESULTS: THE SYSTEMATIC UNCERTAINTIES OF GALAXY QUENCHED FRACTIONS

Different measurement decisions and sample selection effects can have a critical impact on the comparison of observational data with theoretical models. In fact, they can have critical effects in whatever comparison, whether across observational samples, among simulated results, and for the simulation versus observations comparisons. Here we undertake an analysis of how key choices shape the inferred quenched fractions of TNG300 galaxies. Under the assumption that the TNG galaxies are sufficiently realistic, the systematic uncertainties quantified in the following apply to any observational and simulated sample.

3.1 The effects of different measurement choices

3.1.1 Star-forming versus quenched definitions

We explore several criteria to separate quenched versus star-forming galaxies. Throughout the paper, we do not make distinctions between *quiescent* and *quenched* galaxies, which will all be called quenched.

Fig. 1 shows how the star formation MS and quenched fractions, both as a function of galaxy stellar mass, depend on definition. The grey shaded area represents the Poissonian errors,[2] shown for the black curve only for clarity. All galaxies (left-hand panels) are contrasted against centrals (middle panels) and satellites only (right-hand panels), at $z = 0$ (top) and $z = 2$ (bottom). The grey point clouds show the full underlying galaxy population, while coloured curves indicate the boundary between the MS and the quenched galaxies or the quenched fraction medians, respectively, according to the definitions in the legends. The 'quenched' definitions we consider are all are based on SFR, and defined as follows (see Donnari et al. 2019a, for UVJ-based criteria).

Specific SFR $< 10^{-11}\,yr^{-1}$: According to a commonly adopted definition in the literature (Franx et al. 2008; Fontanot et al. 2009; Sherman et al. 2020), 'quenched' galaxies are defined as those having sSFR $\leq 10^{-11}\,yr^{-1}$ (purple lines). We note from the onset that, at $z > 0$, this fixed threshold fails to capture the evolution of the median MS, and a redshift-dependent threshold should be preferred to capture galaxy quenching across cosmic epochs.

1 dex below the extrapolated star-forming MS: Following Donnari et al. (2019a) and Pillepich et al. (2019), we iteratively locate the locus of the MS and label galaxies as quenched if they fall 1 dex or more below it. Specifically, we stack all galaxies in 0.2 dex bins of stellar mass, measure the median SFR in the bin, iteratively remove quenched systems, re-calculate a new median, and repeat until the median SFR in the mass bin converges. Beyond a certain stellar mass the locus of the MS in TNG, as well as in observations, bends downwards towards lower sSFR values. We therefore extrapolate the MS, by adopting a linear fit within two different stellar mass ranges – $M_\odot = 10^{8–10.5}M_{stars}$ (black lines) and $M_\odot = 10^{9–10.2}M_{stars}$ (orange lines). We also use the fit MS directly without extrapolation (green lines).

1 dex below the median SFR: We define the MS as the median value of the SFR values in a given stellar mass bin, using the same 0.2 dex bins as above, and label quenched galaxies as those 1 dex or more below the median (red lines).

Besides the methods listed above, other criteria have been used in the literature, such as the Gaussian Mixture Model (Bisigello et al. 2018; Hahn et al. 2019) and, more traditionally, colour–colour and colour–magnitude diagrams (Williams et al. 2009; Whitaker et al. 2010; Wuyts et al. 2011; Quadri et al. 2012, and reference therein, and see also Donnari et al. 2019a for the simulated UVJ diagram of TNG galaxies).

Fig. 1 shows that all the definitions are largely in agreement for galaxies below $\sim 10^{10.5}\,M_\odot$ at $z = 0$. As a result, they all give the same quenched fractions of ~ 10 per cent for centrals and ~ 80 per cent for satellites, on average. However, for galaxies above $\sim 10^{10.5}\,M_\odot$ the behaviour becomes more complex. At $z = 0$, the linear extrapolations of the MS (black and orange lines) as well as the fixed threshold log (sSFR/yr) < -11 (purple lines) return similar quenched fractions

---

[2]Strictly speaking, the error bars for the quenched fractions should be estimated via a binomial proportion confidence interval. We have compared the Poissonian errors that we adopt throughout to those from a Wilson score interval, finding them consistent in the majority of the regimes.
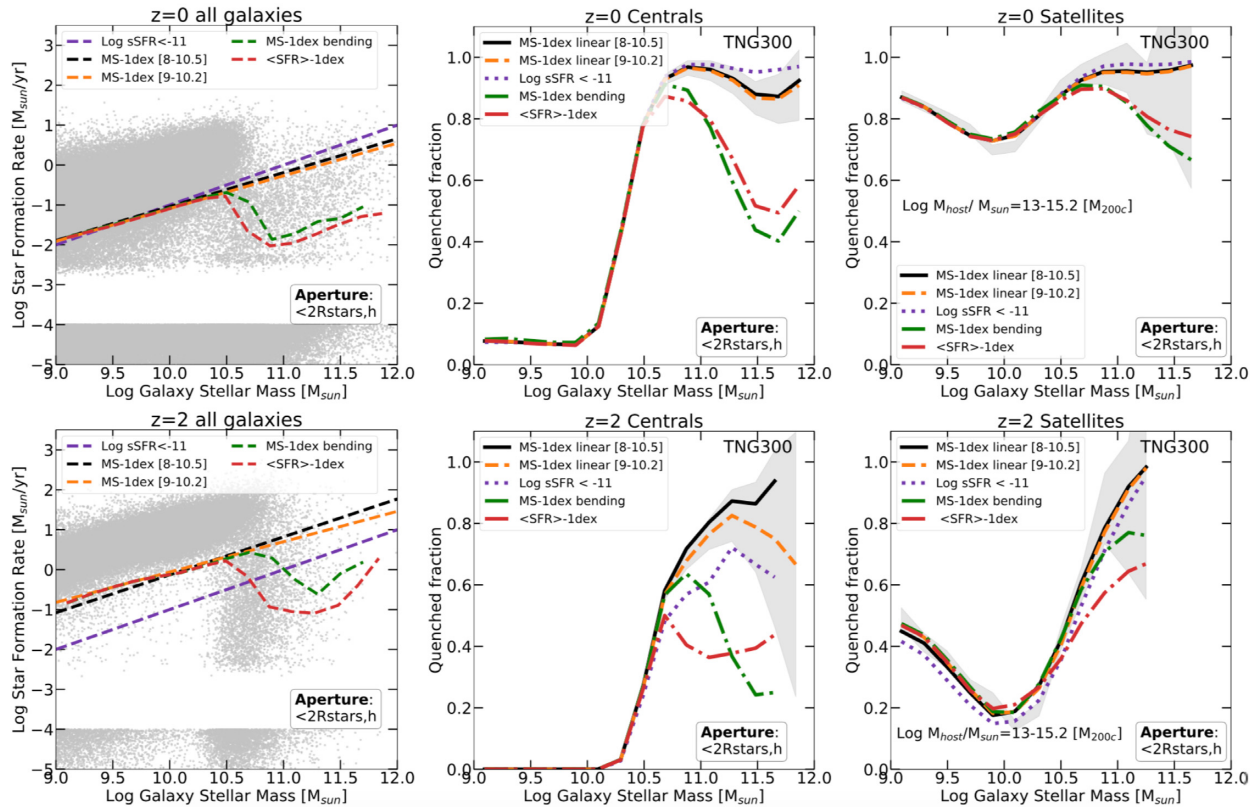
Figure 1. Quenched definitions. The SFR–$M_{\rm stars}$ plane for the TNG300 simulation (left-hand panels), as well as the quenched fraction versus stellar mass for centrals (middle panels) and satellites (right-hand panels), at $z = 0$ (top) and $z = 2$ (bottom). In the left-hand panels, coloured dashed lines define star-forming galaxies and thus trace the lower end of the MS, as labelled in the legend. These same colours show the corresponding quenched fractions in the middle and right-hand panels. The grey shaded area represents the Poissonian errors (shown for the black line only for clarity). Although the definition of 'quenched' has little impact for low-mass galaxies below $10^{10.5}$ M$_{\odot}$, above this mass the inferred quenched fractions are highly sensitive to this choice. This holds for both centrals and satellites, is true at both $z = 0$ and 2, and can qualitatively change the interpretation of high-mass quiescence. These considerations must hence be taken into account whenever results from different observational or simulated samples are compared.

of ∼90–100 per cent for both centrals and satellites.[3] On the other hand, by following the bending MS or the median SFR values (green and red lines), the inferred quenched fractions are lower by up to ∼ 40–50 percentage points for centrals, and up to ∼ 20–30 percentage points for satellites.

The results are similar at $z = 2$, but with even more dramatic possible differences. It should be noted that at higher redshift the SFR–$M_{\rm stars}$ plane is shifted towards higher SFR values: As a result, the fixed criterion log (sSFR/yr) < −11 fails to follow the evolution of the MS and is a poor separator of the quenching boundary. Whereas we discourage the usage of such cut at $z > 0$, here we show it for reference and to raise awareness among the readers by contrasting it against a more appropriate choice. Throughout the paper, we also we use the fixed criterion whenever the TNG outcome is compared to other works that adopt this method, in order to make a significant comparison.

As at redshift 0, also at higher redshifts the low-mass end quenched fractions are insensitive to definition. However, the separation be-

tween quenched versus star-forming galaxies becomes critical for galaxies more massive than $\gtrsim 10^{10.7}$ M$_{\odot}$. For these massive systems, the $z = 2$ quenched fraction of centrals (satellites) can vary from 30 to 100 per cent for centrals, and from 60 to 100 per cent for satellites taken collectively. The simulation predicts that $z = 2$ high-mass galaxies with $M_{\rm stars} \simeq 10^{11.5}$ M$_{\odot}$ have a low quenched fraction of only 30 per cent if one does not extrapolate from the low-mass MS (green lines), or a high quenched fraction of 80 per cent if doing so (orange and black lines, lower middle panel).

Overall, the shape and the actual values of the star formation MS, as well as the quenched fraction of galaxies, sensitively depend on definition at the high-mass end. This is particularly severe for central galaxies, and to a lesser degree, also for satellites. We hence conclude that the only way to make meaningful comparisons across samples is by matching the adopted star-forming versus quenched definitions.

3.1.2 Averaging time-scales for SFRs

As discussed in Section 2.3, the most direct measurement of SFRs from a hydrodynamical simulation is the instantaneous SFR of gas, which is not a direct observable. In observed data, the SFR of a galaxy can be derived from a variety of different techniques, typically based on available wavelength coverage of the SED. These different tracers are thought to be sensitive to recent SFRs over different time-scales (see also Donnari et al. 2019a).

---

[3]It is important to note that this agreement depends on sampling, especially at the highest mass end. In TNG100, for example, where there are fewer high-mass galaxies than in TNG300, fitting the MS over different stellar mass ranges implies quenched fractions at $M_{\rm stars} \simeq 10^{11.0-11.5}$ M$_{\odot}$ that differ by up to 20 percentage points.
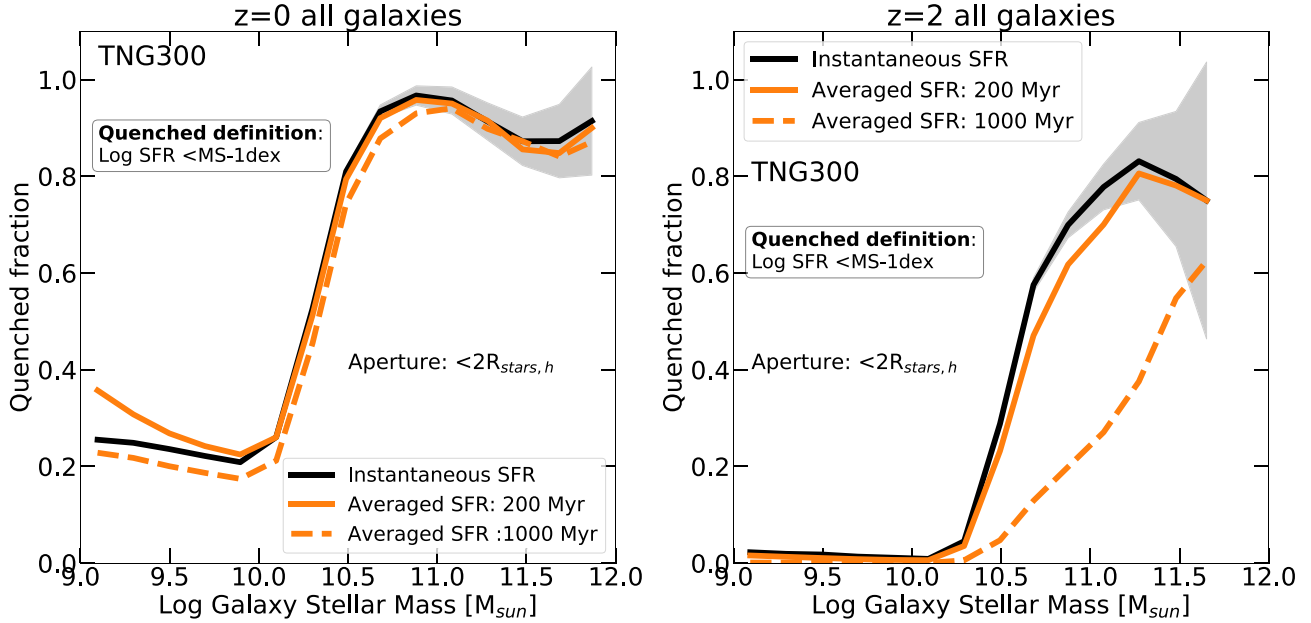
Figure 2. Averaging time-scales for SFR quenched fractions of TNG300 galaxies at $z = 0$ (left) and $z = 2$ (right), for the instantaneous SFR measured from the gas (black) and from the amount of stellar mass formed over the last 200 Myr (orange solid) and 1000 Myr (orange dashed). In all cases, quenched galaxies are selected via the linearly extrapolated MS definition (see the text). While at a redshift of 0 the impact of averaging time-scales is negligible, for $z > 2$ where SFRs are higher and have higher variability, it can be significant.

In Fig. 2, we compare the quenched fractions of galaxies inferred based on adopting different star formation indicator time-scales: instantaneous (black curves) versus time averaged over 200 Myr (solid orange) and 1000 Myr (dashed orange). To label quenched galaxies, we adopt the extrapolated MS definition, and again contrast results from $z = 0$ (left-hand panel) to $z = 2$ (right-hand panel), including all galaxies in TNG300.

At $z = 0$, the impact of different averaging time-scales compared to the instantaneous SFR is almost always negligible. The small differences at the low-mass end ($<10^{9.5}$ $M_\odot$) are likely due to resolution effects (Donnari et al. 2019a; see Appendix A).

At higher redshifts, however, the SFRs of galaxies are overall higher, and more time variable, making indicators more sensitive to different time-scales. At $z = 2$, while there is no appreciable difference between the instantaneous and 200 Myr SFRs (black versus solid orange curves), the longer Gyr time-scale produces substantially lower quenched fractions. For galaxies above $10^{10.5}$ $M_\odot$, this effect lowers the inferred quenched fraction by up to $\sim$40 per cent. This results from how galaxies shift on the SFR–M_{stars} plane under different averaging time-scales, particularly below the MS: While the locus of the MS is not strongly affected by the averaging time-scales (see fig. 2 of Donnari et al. 2019b), the SFRs of galaxies in the process of being quenched (e.g. those that are falling off or are below the MS) can vary dramatically depending on the SFR tracer.

3.1.3 Apertures

The measured value of the SFR depends on the physical aperture within which it is measured. In the following, we show this using the outcome of TNG300: However, given the demonstrated realism of many properties of the TNG galaxies (see Nelson et al. 2019a, for a partial compendium), we expect this issue to apply a priori also to observed galaxies.

To explore the impact of the aperture choice, Fig. 3 shows the quenched fraction trends for an instantaneous SFR measured within

twice the stellar half-mass radius $2R_{stars,h}$ (solid curves) versus for all gravitationally bound resolution elements (dashed curves). The latter is an effective aperture that is typically straightforward to use with simulation data, given the functioning of most galaxy finders, but it is in practice never attained when dealing with observational data. We show results for $z \in \{0, 2, 3\}$ (different panels), including in each case two different quenched definitions, MS 1 dex linearly extrapolated (black), and $\log(sSFR) < -11 \, \mathrm{yr}^{-1}$ (orange).

At $z = 0$, the variations due to different physical apertures are minimal. Conversely, at higher redshift $z > 2$ and for galaxies more massive than $10^{10.5}$ $M_\odot$ we find that larger apertures lead to lower quenched fractions. That is, measuring SFRs within the galaxy body only (solid lines) results in a quenched fraction that is up to $\sim$60 per cent higher, as compared to more global SFRs (dashed lines). It is clear that these differences substantially impact the interpretation of high-redshift galaxy quenching. For instance, if one adopted the $\log(sSFR/yr) < -11$ definition at $z = 2$ and extracted an obvious value from the simulation – the SFR of all gas cells – the conclusion would be that only a small minority, $\lesssim$10 per cent, of massive galaxies with $M_{stars} \simeq 10^{11.5}$ $M_\odot$ are quenched. Changing the aperture to twice the stellar half-mass radius leads to the opposite conclusion: that the majority of galaxies, $\gtrsim$60 per cent, are quenched at the same mass and redshift.

In the TNG model, star formation occurs in extended regions beyond the main bright body of galaxies (Donnari et al. 2019a). Quenching also occurs 'inside-out' (Nelson et al. 2019b, their fig. 16), with star formation first suppressed in the centres of galaxies, and later in the outskirts. As a result, SFRs measured within, for example, radii of 5 pkpc versus 30 pkpc can be non-negligibly different. This is particularly relevant for transitional galaxies, and hence can especially impact high redshift results. As observations can only access a star formation tracer within some aperture, defined by instrumental considerations (e.g. a fibre-fed spectrograph such as SDSS; Salim et al. 2007), or by sensitivity limits (e.g. resulting in loss of sufficient S/N in an emission line such as Hα; Förster
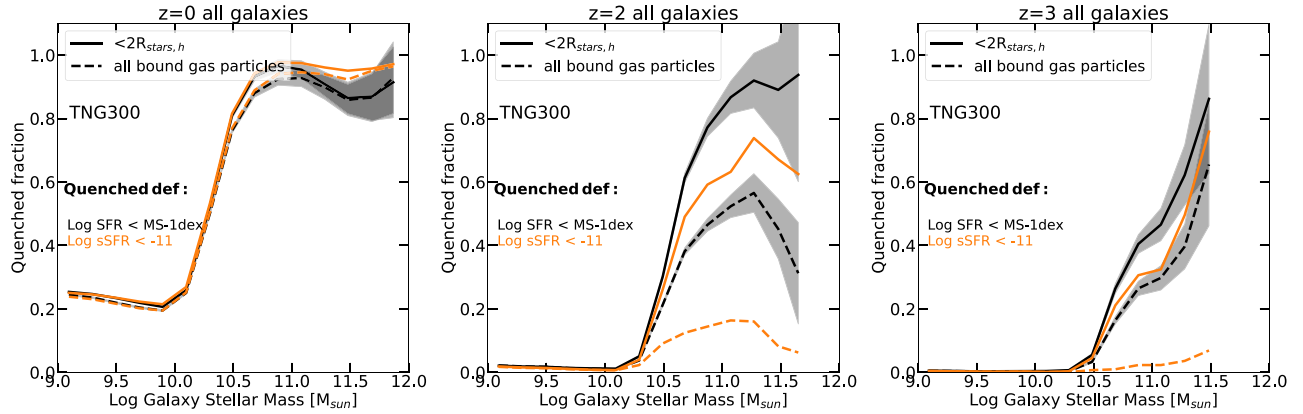
Figure 3. Different apertures and quenched definitions. We show the trend of quenched fraction versus stellar mass for all galaxies in TNG300 at $z = 0$ (top left), $z = 2$ (top middle), and $z = 3$ (top right). In each case, we contrast two different physical apertures within which star formation is measured, either restricting to $2R_{stars,h}$ (solid) or including all gravitationally bound gas (dashed). We also compare the impact for two different quenched definitions, MS 1 dex (black) and sSFR < 11 (orange). Measuring SFRs in larger apertures substantially lowers the inferred quenched fractions for high-mass $z \gtrsim 2$ galaxies.

Schreiber et al. 2018), it is important to appropriately match this choice in any comparison: for theory versus data, data versus data, or theory versus theory. Indeed, different choices in SFR measurements have been proposed as the underlying reason for the long-standing inconsistency between observed stellar mass functions from $0 < z < 2$ and observed SFRs over these same epochs (Leja et al. 2020).

3.1.4 Statistical errors for observationally inferred values

To explore the impact of statistical uncertainties on the two key observationally inferred quantities – stellar masses and SFRs – Fig. 4 again shows the quenched fractions of centrals (top panel) and satellites (bottom panel) at $z = 2$. The results at a redshift of 0 are similar and omitted for brevity. The direct output of the simulation, with M_{stars} and SFR values assumed to be known exactly, is shown with solid black lines. We then add a random Gaussian component to the simulated stellar masses (with $\sigma = 0.2$ dex) and SFRs (with $\sigma = 0.6$ dex) in order to roughly match observational uncertainties in these quantities, as inferred from SED fitting and SFR calibrators, respectively (see also Genel et al. 2018; Leja et al. 2019; Anand, Nelson & Kauffmann 2021). The resulting quenched fractions are shown with dashed curves.

Regardless of the redshift and of whether a galaxy is a central or a satellite, the most important effect of including reasonable statistical uncertainties is a 'shallower' trend of the quenched fraction with galaxy stellar mass. This is particularly clear for centrals, where the slope of the median line is smaller with uncertainties included. We think that, when measurement uncertainties are included, more low-stellar mass galaxies scatter into higher stellar mass bins than the other way around. The sharpness of the transition from predominately star forming to predominately quenched galaxies is also reduced, i.e. smoothed out, by the observational uncertainties: The latter is intricately linked with the physics of quenching in the simulations (also producing the blue to red colour transition; Nelson et al. 2018a), most notably through the high- to low-state black hole feedback transition (Weinberger et al. 2018). It has been suggested that this transition is somewhat too rapid in TNG (e.g. Terrazas et al. 2020), but our result highlights that not only the normalization, but also the slope, of galaxy population relations can change due to systematic uncertainties (see also Stevens et al. 2019, for a related conclusion with respect to the M_{HI}–$M_\star$ relation for TNG).

Beyond this steepening effect, the inclusion of systematic errors also changes the actual values of galaxy quenched fractions, depending on mass. With respect to the error-free case (solid lines), galaxies below (above) $\sim 10^{10.5}$ M$_\odot$ show a higher (lower) quenched fraction, by ~ 10 per cent. One consequence is the interpretation of whether or not low-mass quenched field centrals (Geha et al. 2012) exist in the simulations.

3.2 The effects of sample selection

3.2.1 Host halo mass distributions

The most relevant quantity that determines environmental processes is the total halo mass of the host potential wells where satellite galaxies reside. Both observations and simulations agree in that higher host mass implies larger quenched fractions (at least for low-mass galaxies; see Donnari et al. 2021, for an extended discussion). A multitude of observational techniques are adopted to infer host halo masses (via abundance matching or mass proxies like X-ray cluster properties or optical-based richness), each affected by its own systematic bias and/or random errors (Old et al. 2014, 2015, 2018). For example, Old et al. (2018) find a systematic bias for several methods used to infer halo masses, depending on the cluster mass. They are indeed significantly overestimated for lower mass clusters, by ~ 10 per cent at 10^{14} M$_\odot$ and larger than 20 per cent for $< 10^{13.5}$ M$_\odot$ (see also Old et al. 2014, 2015, for a related topic).

While a discussion of systematic bias and uncertainties of different halo mass measuring is beyond the scope of this paper, here we comment on the importance of matching the nominal host mass distributions when comparing the satellite quenched fractions across samples.

First, because the satellite quenched fractions are a strong function of host mass, it is essential to match at the very least the upper and lower boundaries of the sampled host mass distributions. Fig. 5 shows the quenched fraction of satellites at $z = 0$ for three host mass ranges with distributions inherited from the volume-limited nature of the TNG300 simulated box. The minimum host mass is 10^{13}, 10^{12}, and 10^{11} M$_\odot$ for the black, orange, and red curves, respectively. In all cases, the upper limit of the host mass distributions is $10^{15.2}$ M$_\odot$, the most massive host of the volume. Solid and dashed curves denote M_{200c} versus M_{200m} mass definitions, respectively, where the latter denotes the mass enclosed in a sphere whose mean density
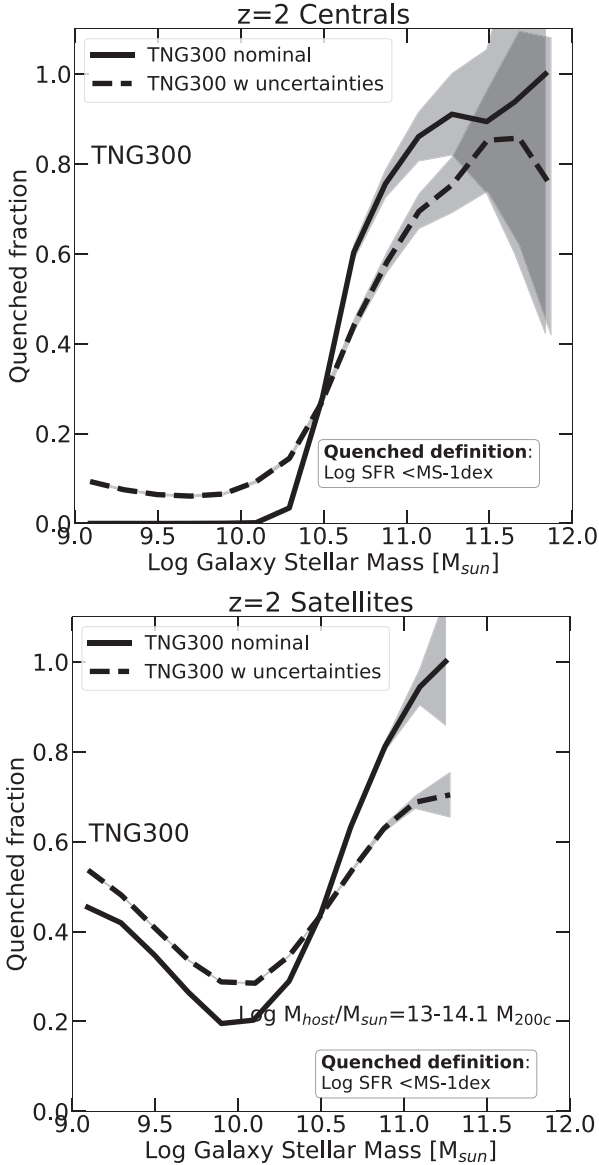
Figure 4. Observational uncertainties. Quenched fractions of centrals (top panel) and satellites (bottom panel) at $z = 2$. Solid curves show direct results from TNG300, assuming 0 uncertainty, while dashed curves include observationally motivated random errors on the galaxy stellar mass and SFR measurements. The most important effect of including these systematic uncertainties is to make the trend of quenched fraction with mass less steep, as galaxies are scattered between bins. The same findings apply almost unchanged at all redshifts studied here ($z \leq 3$).

is 200 times the *mean* density of the Universe. It is evident that not properly matching the lower host mass boundary produces large systematic differences, up to 20 percentage points if the minimum host mass is e.g. 10^{13} versus 10^{12} M$_\odot$. Secondly, whereas confusing the nominal definition of halo mass (e.g. M_{200c} versus M_{200m}) is a priori a problem, in comparison to what has been discussed so far it is a lesser issue, but it still impacts the results by up to 10 per cent. The situation is similar across redshifts, $z \leq 2$. In the next sections, we further comment on the importance of matching the shape of the host mass function within the considered host mass range.
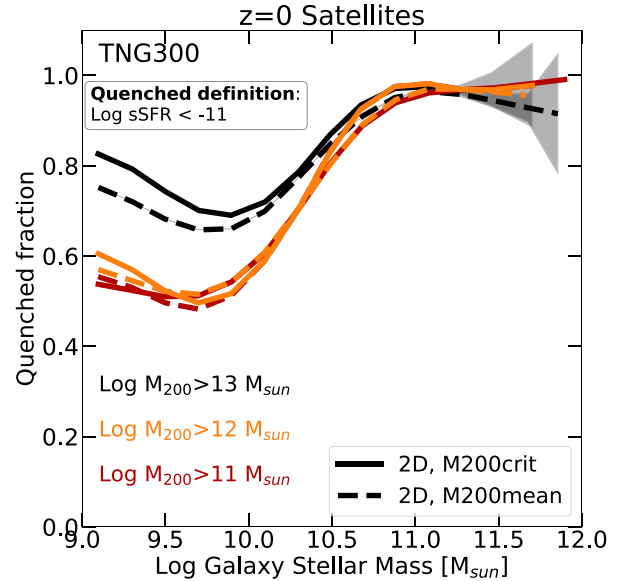

Figure 5. Host halo mass. TNG300 quenched fraction of satellites at $z = 0$ for three host mass ranges where the minimum host mass reads 10^{13}, 10^{12}, and 10^{11} M$_\odot$ for the black, orange, and red curves, respectively. In all cases, the upper limit of the host mass distributions is $10^{15.2}$ M$_\odot$, the most massive host of the TNG300 volume. Solid and dashed curves denote M_{200c} and M_{200m} mass definitions, respectively. Not properly matching the lower host mass boundary implies systematic differences of up to 20–25 percentage points at the low-mass end.

3.2.2 3D versus 2D projected distances for satellites

After host halo mass, another important quantity for an environmental study of galaxy evolution is the halocentric distance: the distance between a satellite galaxy and the centre of its host (or parent) dark matter halo. We can measure distances in two ways: in two-dimensional projection, on an imaginary sky plane perpendicular to the line of sight, and in three dimensions. While the latter is the obvious and more common choice in simulations, it is fundamentally inaccessible in observations, which must rely on either pure sky-projected distance or a combination of projected distance and line-of-sight velocity separation (i.e. an unknown mixture of line-of-sight peculiar velocity and line-of-sight physical distance). The same issues arise for group identification and membership assignment itself, i.e. finding host haloes and associating satellites to them (Yang et al. 2007; Lim et al. 2017).

In Fig. 6, we show radial profiles of satellite quenched fractions, comparing these two choices: two-dimensional cylindrical satellite selection with radius R_{200c} and encompassing the entire FoF along the z-direction and projected 2D distances (solid curves) versus three-dimensional spherical satellite selection within R_{200c} and 3D distances (dashed curves). We slice the galaxy population in two different ways. In one case, we include all satellites above a stellar mass threshold of $M_{stars} > 10^9$ M$_\odot$, splitting the sample into four different host mass bins (left-hand panel). Separately, we include all hosts above a halo mass threshold of $M_{halo} > 10^{13}$ M$_\odot$, splitting the sample into four different satellite stellar mass bins (right-hand panel).

In terms of global trends, for all satellite masses in all hosts, the fraction of quenched galaxies is higher closer to the host centre (see also Donnari et al. 2021; fig. 4). This halocentric distance trend is stronger for lower mass hosts (yellow lines), because distant satellites have not yet saturated towards very high quenched fractions. The
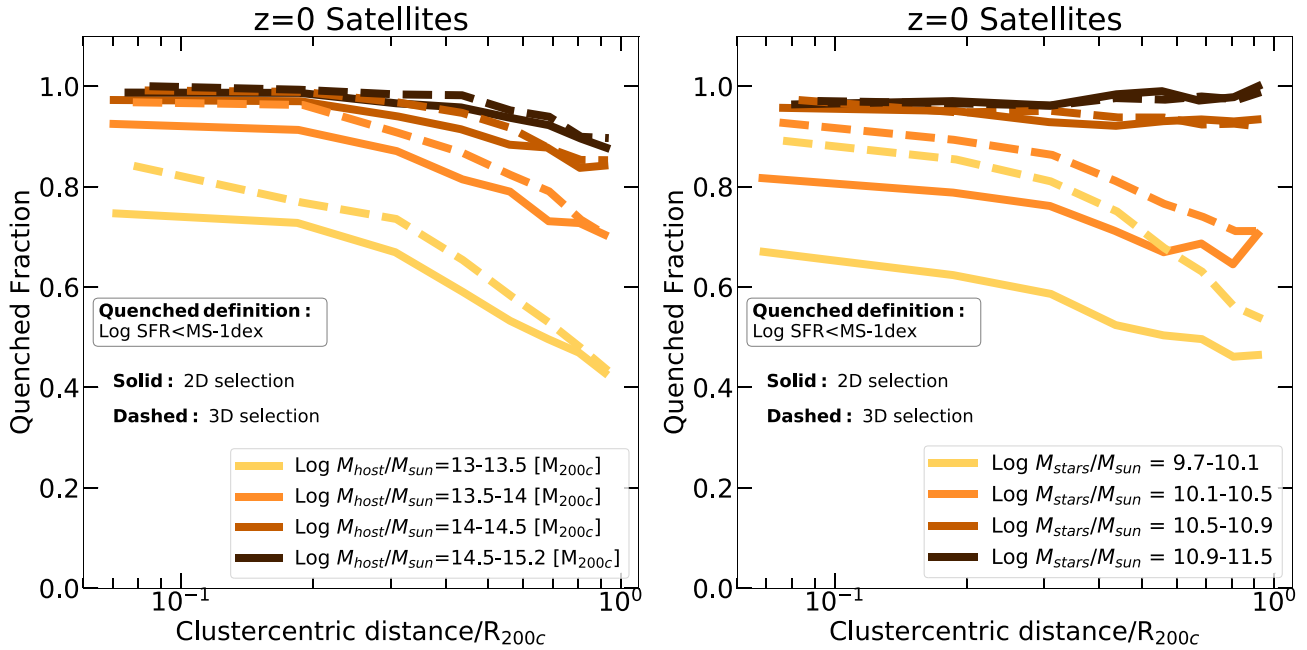
Figure 6. 3D versus 2D distances. Radial profiles of satellite quenched fractions as a function of 3D membership identification and distances (dashed), versus 2D membership identification and projected distances (solid). In the former case, galaxies within a radius of R_{200c} are classified as satellites, while in the latter case galaxies within a cylinder of this same radius and encompassing the entire FoF along the z-direction are labelled as satellites. We show two decompositions of the satellite + host population. First, including all satellites with $M_{stars} = 10^{9-12}$, we split the sample into four host halo mass bins (left-hand panel). Secondly, stacking all host haloes with $M_{host} = 10^{13-15.2}$ we split the satellites into four stellar mass bins (right-hand panel). Quenched fractions are always lower for the 2D case because of contamination effects, by up to 20 percentage points, e.g. in the case of low-mass satellites (right-hand panel, yellow curves).

distance trends are also stronger for lower stellar mass satellites, and become negligible above $M_{stars} \gtrsim 10^{10.5}$ M$_\odot$ for the same saturation reason, i.e. because the quenched fraction of such massive galaxies in the field is already high due to supermassive black hole feedback-based quenching.

The impact of two- versus three-dimensional sample selection and distance calculations is manifest in both panels. In fact, the differences in Fig. 6 probably underestimate the effects, as in the 2D measurements along the z-direction we here only account for galaxies within the FoF hosts, and not for any galaxy within a line-of-sight velocity range or photometric-redshift uncertainty that may correspond to projections along the line of sight of even many Mpc. Quenched fractions are always lower for the 2D case, which is due to two contamination effects. First, satellites at large physical distances (and thus lower quenched fractions) but small projected distances contaminate small radius bins and suppress the quenched fraction. Secondly, and similarly, field galaxies that are not physically associated with or affected by the host, due to large physical but small projected distances, likewise contaminate $\leq R_{200}$ bins with field systems, suppressing the inferred quenched fractions.

The quantitative change depends on host and satellite mass. For host haloes less massive than 10^{14} M$_\odot$ (left-hand panel), the 2D-based technique reduces the quenched fraction by ~ 10 per cent with respect to the 3D method, decreasing to 0 towards higher host mass. For satellites with stellar masses less than $10^{10.5}$ M$_\odot$, the quenched fractions with the 2D technique are lower by ~ 20–30 percentage points when compared to the 3D method. This difference increases towards zero distance – near the centre of the parent halo. We note (but do not show) that such differences are less clear if we consider quenched fractions versus satellite stellar mass, stacking all hosts together, likely because the statistics are dominated by large distance orbits where the effect is always smaller.

3.2.3 Misclassification of centrals and satellites

Because group finding in observational galaxy surveys is inexact, regardless of using spectroscopic or photometric redshifts, some level of misclassification is always present. Namely, some central galaxies are inevitably misidentified as satellites of higher mass haloes, while at the same time some fraction of satellites are misidentified as isolated centrals (see e.g. Campbell et al. 2015; Tinker 2020). This leads to a contamination effect for any measured quenched fraction, and it has only recently become appreciated that the resulting bias can be important for comparisons (in the GAMA survey: Bravo et al. 2020).

Whereas we do not attempt to replicate the typical observational procedures for galaxy group finding in the projected sky by adopting the simulation data, we quantify for this effect by randomly reassigning labels from 'central' to 'satellite', and vice versa, with given mixture fractions. That is, we take some per cent of centrals f_{cen}, at random, and relabel them as satellites. Simultaneously, we take some per cent f_{sat} of satellites, at random, and relabel them as centrals. We denote this transformation by [cen, sat] = $[f_{cen}, f_{sat}]$. For a given observational group finder, these fractions can be estimated by careful mocks on synthetic catalogues.

Our fiducial choice is based on our comparisons below to SDSS catalogue data as described in Wetzel et al. (2012, 2013). Namely, averaging across all halo masses, the purity and completeness of satellites are both ~ 80 per cent, while the purity and completeness of centrals are slightly $\gtrsim 80$ per cent (Tinker, Wetzel & Conroy 2011; Fig. C1). This implies that the average fraction of galaxies that are classified as centrals when they are in fact satellites, and vice versa, is about one in ten. For our purposes, this translates into [cen, sat] = [10, 20], such that we relabel 10 per cent of centrals to satellites, and 20 per cent of satellites to centrals. We neglect the

additional complexity of the host halo mass dependence of these contamination fractions. A similar assessment of the satellite/central misclassification is also shown in terms of averaged SFRs by Stevens et al. (2021).

Even if we do not show it here, we have quantified the impact of this reassignment. At the high-mass end ($M_{stars} > 10^{10.5}$ M$_\odot$) of both centrals and satellites, the misclassification does not imply any significant difference in terms of quenched fractions. Conversely, at the low-mass end ($M_{stars} < 10^{10}$ M$_\odot$), we find a higher (lower) quenched fraction for centrals (satellites) of about 10 percentage points. We show the impact of this reassignment in concert with other analysis choices below in Section 4.2 when we quantitatively compare to the SDSS data.

3.2.4 Matching the stellar mass distribution

The fraction of quenched galaxies is a strong function of mass. As a result, subtleties in the mass distributions of both host (halo mass) and galaxy (stellar mass) can influence comparisons. Here, instead of matching to a given observational mass distribution, we perform a more illustrative experiment. Specifically, we generate two very differently shaped mass distributions from TNG300: a volume-limited sample, as naturally emerges from any cosmological volume simulation with no additional cuts, and a 'flat' sample, with uniform statistics in bins of logarithmic mass. This can occur in practice in a survey where a magnitude limit is convolved with increasing volume towards higher redshift, such that progressively brighter galaxies are included more frequently and in turn make the sample mass distribution more top heavy than in reality.

Fig. 7 shows these two mass distributions, constructed for $z = 0$ satellites in TNG300: the nominal volume-limited case (orange), where the number of galaxies declines rapidly towards larger masses, and an artificially flat distribution (black). For the latter, we randomly subselect satellites residing in hosts with $M_{halo} > 10^{12}$ M$_\odot$ to obtain an equal number within fixed 0.2 dex bins.

The corresponding quenched fractions as a function of host mass M_{200c} are also shown in Fig. 7 (upper right panel), where we focus on $z = 0$. Satellites are stacked in three stellar mass bins: $M_{stars} = 10^{9-10}$ M$_\odot$ (circles), $M_{stars} = 10^{10-11}$ M$_\odot$ (squares), and $M_{stars} > 10^{11}$ M$_\odot$ (triangles). We define quenched systems using the log (sSFR) < -11 yr^{-1} definition. Although the two galaxy mass distributions are quite different, we see that the resulting quenched fractions differ by at most ~ 10 per cent at any host mass, where the flat distribution returns higher values.

A much stronger impact is found when we consider all galaxies, both centrals and satellites, as a function of redshift. In Fig. 7 (lower right panel), we show the median quenched fractions for three lower limits of galaxy stellar mass: $M_{stars} > 10^{\{9,10,11\}}$ M$_\odot$ (circles, squares, and triangles). For the two lower stellar mass thresholds, at all redshifts, the flat galaxy mass distribution results in quenched fractions that are ~ 20 percentage points higher. Such open-ended mass selections are often considered observationally, but these act very differently in practice between the two samples. Bins in the volume-limited selection are dominated by galaxies near the lower edge of the mass cut, while for the flat selection bins are dominated by much higher mass galaxies, depending on the volume.

Comparisons of quenched fractions between simulations and observational surveys (and in general among any galaxy sample) must account for the impact of the differing shape of the sampled mass distributions.

4 COMPARISON TO OBSERVATIONS AND OTHER THEORETICAL MODELS

4.1 TNG quenched fractions versus observations

In this section, we compare the TNG model to several observational data sets that measure the quenched fractions of central and satellite galaxies separately. These observations are based on a diversity of data with different characteristics, which we enumerate in Table 1 – redshift range, stellar and host mass range, SFR tracer, criteria to separate star-forming versus quenched galaxies, and adopted initial stellar mass function. Comparison requires adopting appropriate analysis choices particular to each data set.

We first show a number of 'at-face-value' comparisons, adopting analysis choices common to all the selected data sets. Fig. 8 presents the quenched fractions of central galaxies (left column) and satellites (right column), for three redshifts: $z = 0.1$ (top row), $z = 0.35$ (middle row), and $z = 0.65$ (bottom row). The observational surveys are denoted with different grey symbols as indicated in the legends: Wetzel et al. (2013) (SDSS, pentagons) and McGee et al. (2011) (SDSS/GEEC, crosses), Davies et al. (2019) (GAMA, plus), Schaefer et al. (2019) (SAMI, stars), Lin et al. (2014) (Pan-STARRs, squares), Jian et al. (2018) (Subaru HSC, triangles), Fossati et al. (2017) (3HST/CANDELS, circles), and Wagner et al. (2017) (CLASH, diamonds). For clarity, we exclude observational error bars.

We adjust the observational data points only to transform them from the Salpeter (s) to Chabrier (c) IMF where required, adopting $\log (M_\star^c) = \log (M_\star^s) - 0.24$ and $\log (SFR^c) = \log (SFR^s) - 0.15$. We do not attempt to mock or reproduce the operational definitions of SFRs and stellar mass estimates, nor the apertures within which they are measured: All these differ across the observational analyses.

For comparison, we select satellites from the TNG100 (blue curves), TNG300 (orange curves), and TNG50 (green curves) simulations residing within hosts more massive than 10^{13} M$_\odot$ at each redshift. Only for TNG100 satellites at $z = 0.1$ we additionally show the quenched fractions taking the minimum host mass to be 10^{11} M$_\odot$ (thin blue curve). For TNG, SFRs and stellar masses are taken within twice the stellar half-mass radius, and the SFRs are instantaneous (see Section 3.1.2). In all panels and for all the simulations, shaded areas represent the Poissonian error. Furthermore, to account for the limited simulation volumes, we give estimates of the cosmic-variance uncertainties by measuring the quenched fractions in eight subboxes of 50 Mpc on a side for TNG100 and by computing the *jackknife* error represented by the striped blue areas in all panels. We note that the smallest volume simulation, namely TNG50, must be affected by at least such an amount of uncertainties due to sample/cosmic variance.

For group membership, we require satellites to fall within a cylinder of radius R_{200} and depth encompassing the entire FoF along the z-direction (random) of the simulation. We account for systematic uncertainties by adding random Gaussian errors to the simulated data ($\sigma = 0.2$ dex for $M_\star$, and $\sigma = 0.6$ dex for SFR, as before). We define quenched galaxies as having log (SFR) 1 dex or greater below the MS, which is not identical to all the observational definitions employed. However, as demonstrated in Fig. 1, at low redshift the definition based on the distance from the recursive MS gives very similar results to that with a fixed SFR cut, which is the choice adopted by most observational works. We also note that the adopted host mass range does not exactly match all the observations (see Table 1).

Overall, the TNG model returns qualitatively and, in some cases, quantitatively good agreement with the observations as a whole, at both low and high stellar masses, for centrals and satellites, and at all
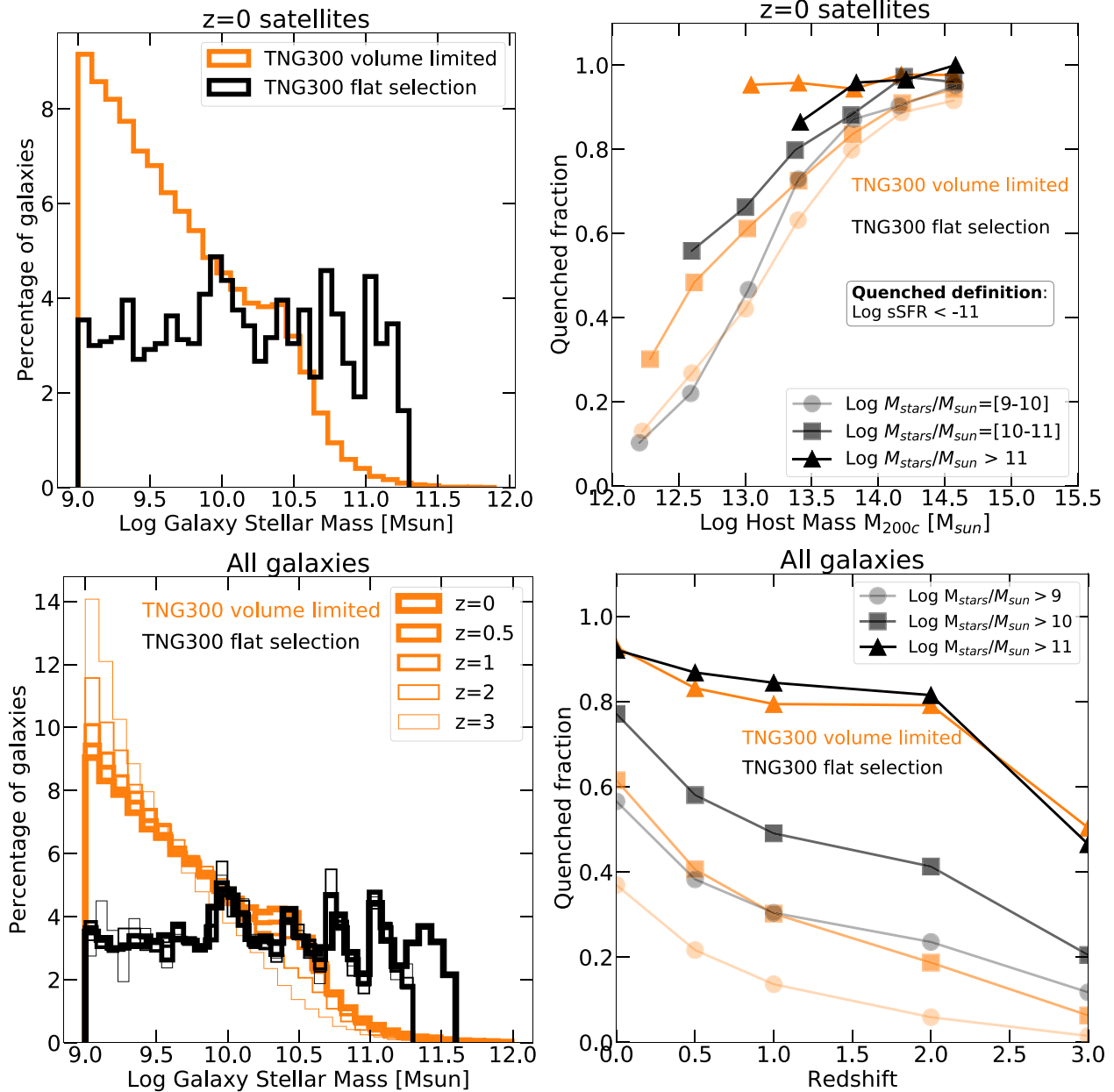
Figure 7. Different sample mass distributions. We construct two fundamentally different galaxy distributions, both out of the galaxies simulated within the TNG300 volume: a volume-limited sample (orange) and a flat distribution (black). First, we consider satellite galaxies residing in all host haloes above $10^{12}\,\mathrm{M}_\odot$ (top row), showing the satellite mass distributions themselves (top left) and the resulting quenched fractions as a function of host mass (top right), for three different satellite galaxy stellar mass bins. In this case, the impact on inferred quenched fractions is small: The flat distribution produces up to ~ 10 per cent higher quenched fractions, particularly for intermediate-mass hosts. Secondly, we consider all galaxies with $M_\star > 10^9\,\mathrm{M}_\odot$ together with their redshift evolution, showing the resulting mass distributions (lower left) and quenched fractions as a function of redshift (lower right) for three different minimum stellar mass thresholds. Here, the two synthesized galaxy samples have substantially different quenched fractions. The flat sample, as is more common observationally, produces quenched fractions ~ 20 per cent higher than the volume-limited sample, as would be intrinsically drawn from a cosmological simulation volume, for all most massive galaxy mass bins.

redshifts available. Particularly, for massive centrals at any redshift, TNG50 returns a lower quenched fraction with respect to TNG100 and TNG300 and to observations, likely due to a combination of a low statistics and mass resolution (Appendix C).

In more detail, the best agreement is found at low redshift $z = 0.1$ and for low-mass centrals ($\lesssim 10^{10}\,\mathrm{M}_\odot$) where the TNG model is remarkably consistent with the SAMI and GAMA data (Davies et al. 2019; Schaefer et al. 2019). Both simulations and observations show

quenched fractions of less than 5 per cent in this regime. Similar agreement is also present at the high-mass end ($M_{\mathrm{stars}} \gtrsim 10^{11}\,\mathrm{M}_\odot$), where the quenched fractions of TNG galaxies are consistent with both SDSS (Wetzel et al. 2013) and SAMI surveys. The simulations exhibit reasonable agreement with Davies et al. (2019) and McGee et al. (2011), which have a larger scatter than the other samples. For massive satellites ($\gtrsim 10^{10}\,\mathrm{M}_\odot$, right-hand panel), we see good quantitative agreement with Wetzel et al. (2013) and McGee et al.

Table 1. Summary of the observational data used for comparisons in Figs 8 and 10. Column 1: type of data, observations or models. Column 2: references. Column 3: survey. Column 4: redshift range. Column 5: stellar mass range. Column 6: host mass range. Column 7: SFR tracers. Column 8: threshold used to divide star-forming from quenched galaxies. In all results considered here, a Chabrier IMF (Chabrier 2003) is adopted but for Lin et al. (2014) that use a Salpeter IMF (Salpeter 1955).

Observations/ models	Reference	Survey/ simulation	Redshift	$\log M_{stars}$ range ($M_\odot$)	$\log M_{host}$ range ($M_\odot$)	SFR tracer	Quenched criteria
Observations ($z \simeq 0.1$)	McGee et al. (2011)	SDSS	0.08	9.4–11.2	–	SED (UV)	$\log$ sSFR $= -11$
	Schaefer et al. (2019)	SAMI	<0.11	8.2–11.5	>12.5	H α	$\log$ sSFR $= -12$
	Davies et al. (2019)	GAMA	0.1–0.2	9–11.4	11.4–15	H α	$\log$ sSFR $= -10.5$
	Wetzel et al. (2013)	SDSS	0.1	9.5–11.5	12–15	H α/D_n4000	$\log$ sSFR $= -11$
Observations ($z \simeq 0.35$)	Lin et al. (2014)	Pan-STARRS1	0.2–0.5	9.25–11.35	12.5–14	SED	$\log$ sSFR $= -11$
	Jian et al. (2018)	HSC	0.2–0.5	9.1–11.9	>13.6	SED (UV–opt)	$\log$ sSFR $= -10.1$
	Wagner et al. (2017)	CLASH	0.15–0.4	10.1–11.4	14.8–15.7	SED	$D_n(4000)$
Observations ($z \simeq 0.65$)	Fossati et al. (2017)	3D *HST*/CANDELS	0.5–0.8	9.6–11.4	>13	–	UVJ
	Jian et al. (2018)	HSC	0.5–0.8	9.1–11.9	>13.6	SED (UV–opt)	$\log$ sSFR $= -10.1$
	Lin et al. (2014)	Pan-STARRS1	0.5–0.8	9.25–11.35	12.5–14	SED	$\log$ sSFR $= -11$
	Wagner et al. (2017)	CLASH	0.4–0.8	10.1–11.4	14.8–15.7	SED	$D_n(4000)$
Models	Schaye et al. (2015)	EAGLE	0	9–12	14–14.5	–	$\log$ sSFR $= -11$
	Bahé et al. (2017)	Hydrangea	0–2	9.25–11.35	14–14.5	–	$\log$ sSFR $= -11$
	Tremmel et al. (2019)	RomulusC/25	0	8–11.2	>14	–	$\log \Delta$MS < 1 dex
	Ayromlou et al. (2021)	L-Galaxies (TNG)	0–2	9–12	14–14.5	–	$\log$ sSFR $= -11$

(2011). We note that the mismatch for low-mass satellites is, at least in part, apparent and due to different host mass selections (see findings of Fig. 5), as we suggest with the thinner and lighter TNG100 curve in the top right panel that includes hosts down to $10^{11}\,M_\odot$. However, in TNG we do not recover the very low quenched fractions of low-mass satellites as inferred from GAMA observations (Davies et al. 2019). We explore this possible issue in more detail below.

At intermediate redshift $z = 0.35$, TNG exhibits a fraction of quenched centrals systematically lower than what is found in observations, across stellar mass. This is particularly true in comparison to Jian et al. (2018) where the discrepancy is up to ~ 40 percentage points. This also appears to be the case for satellites, where TNG generally has quenched fractions lower than Jian et al. (2018) by ~ 20 percentage points, although we note the significant discrepancy between this data set and all other observations, by roughly the same amount. On the other hand, the quenched fractions of high-mass TNG satellites broadly agree with Lin et al. (2014) and Wagner et al. (2017), to better than ~ 10 per cent. We speculate that disagreement for low-mass centrals may also be driven by differing galaxy classifications. For example, in both Lin et al. (2014) and Jian et al. (2018), and differently from TNG and Wagner et al. (2017), this classification is based on the richness of groups (N_{rich}): 'Field' galaxies are those not associated with any groups with $N_{rich} > 2$, 'groups' span $10 < N_{rich} < 25$, and 'clusters' are defined as having $N_{rich} > 25$. Because of this and possibly other reasons, we note that some observational estimates appear mutually inconsistent, depending on mass regime.

At $z = 0.65$, TNG massive centrals ($\gtrsim 10^{10.5}\,M_\odot$) show quenched fractions in striking agreement with both Lin et al. (2014) and Jian et al. (2018). On the other hand, the comparison to 3DHST/CANDELS galaxies (Fossati et al. 2017) shows an offset, with the simulations having a quenched fraction lower by ~ 20 percentage points at the low-mass end, while also being higher by ~ 25 percentage points at the high-mass end. We reiterate that observed fractions of quenched massive satellites at this redshift are mutually inconsistent, and differ by up to 40 per cent. The degree to which the simulations are consistent with data therefore depends on the particular data set, as well as the stellar and host mass regimes.

More targeted comparisons between TNG and other observational data sets at these intermediate redshifts, but with no distinction between central and satellite galaxies, can be found in fig. 3 of Donnari et al. (2019a).

In conclusion, the salient features of the observed quenched fractions of both centrals and satellites are broadly recovered by the TNG model. However, due to a combination of different systematics, sample selections, host mass ranges, membership definitions, and the issue of central/satellite misclassification, it is challenging to draw a satisfactory comparison. This is true not only for model versus observation comparisons, but also among observations themselves (Speagle et al. 2014; Donnari et al. 2019a). Particularly at the low-mass end, models and observations rarely show good agreement (see also Bahé et al. 2017; Tremmel et al. 2019).

4.2 A more robust comparison: mocking the SDSS measurement choices and sample selection

We proceed with a more robust and well-posed comparison, by focusing on data from the SDSS as compiled and analysed by Wetzel et al. (2012, 2013), by using the same SDSS catalogues used in their papers. To do so, we adopt the most important choices and definitions of those works, their systematic impacts on TNG quenched fractions having been demonstrated in Section 3. In particular, we

(i) classify satellite galaxies based on a two-dimensional, projected, cylindrical selection geometry (see Fig. 6);

(ii) adopt the same virial radius R_{200m} and corresponding host mass definitions, based on a sphere whose mean density is 200 times the *mean* density of the Universe;

(iii) define satellites as quenched using the observationally employed criterion of $\log$ (sSFR) $< -11\,yr^{-1}$ (see Fig. 1);

(iv) account for the effects of misclassification between centrals and satellites with fractions [cen, sat] $= [10, 20]$ (see Section 3.2.3);

(v) include appropriate systematic uncertainties with 0.2 dex random errors for stellar masses, 0.3 dex for halo masses, and 0.6 dex for SFRs (see Fig. 4);

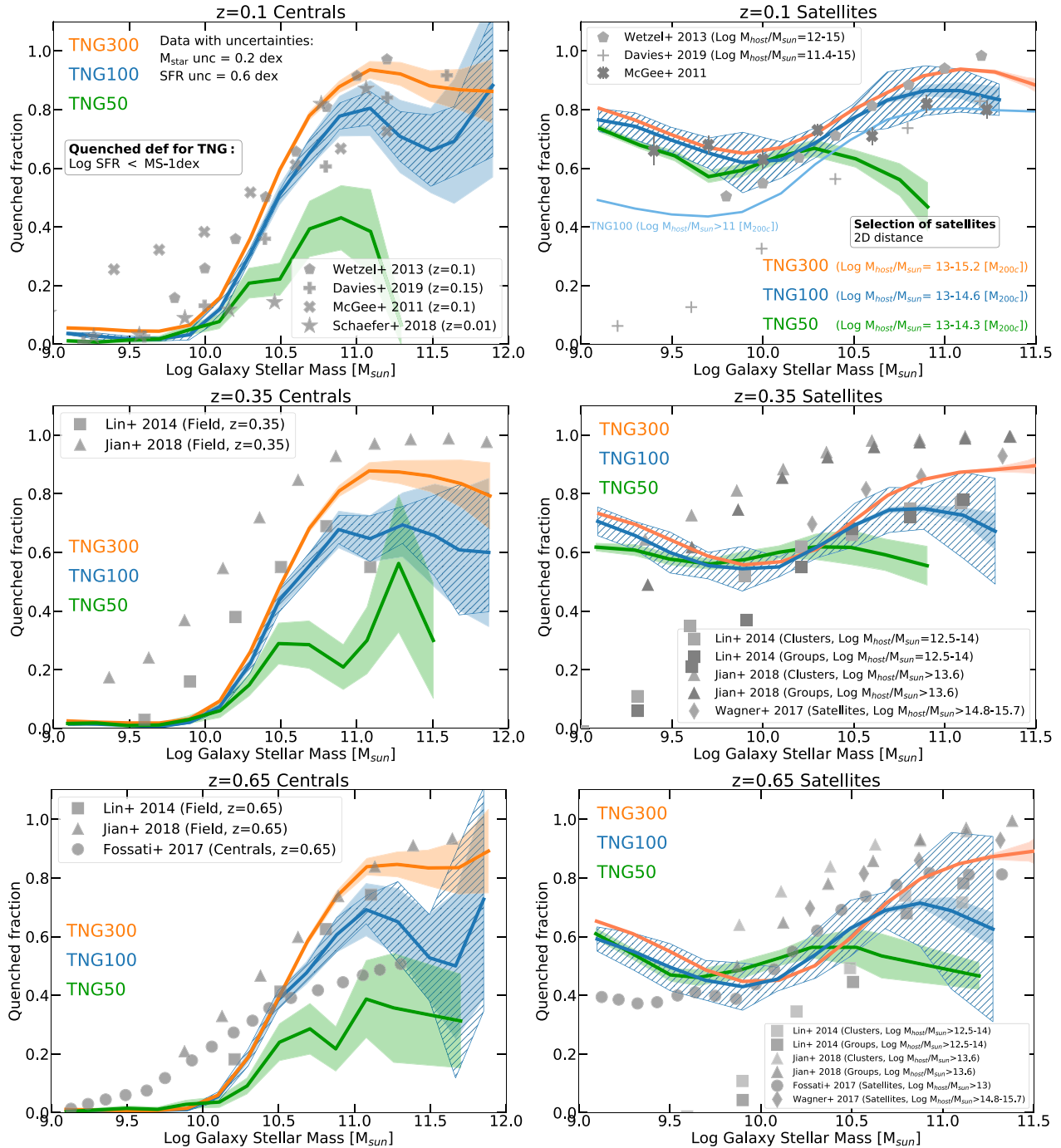(vi) match the galaxy stellar mass and host mass distributions (see Fig. 7 and further details in Appendix B).

Figure 8. Nominal comparison to observations at low redshift. Quenched fractions in TNG100 (blue), TNG300 (orange), and TNG50 (green) for central galaxies (left column) and satellites (right column) at $z = 0.1$ (top), $z = 0.35$ (middle), and $z = 0.65$ (bottom). Shaded areas represent the Poissonian errors, while the striped areas around TNG100 represent the cosmic variance uncertainties. We roughly mock the satellite selections of the diverse data sets (see the text) that are shown with different grey symbols and indicated in the legend: Wetzel et al. (2013) (SDSS, pentagons) and McGee et al. (2011) (SDSS/GEEC, crosses), Davies et al. (2019) (GAMA, plus), Schaefer et al. (2019) (SAMI, stars), Lin et al. (2014) (Pan-STARRs, squares), Jian et al. (2018) (Subaru HSC, triangles), Fossati et al. (2017) (3HST/CANDELS, circles), and Wagner et al. (2017) (CLASH, diamonds). Overall, the TNG quenched fractions are qualitatively (and often, quantitatively) in good agreement with observations as a whole, although discrepancies are visible in certain regimes.

Both SFR and stellar masses in TNG are based on the simulations output, with no mocking strategy, and are evaluated within twice the stellar half-mass radius, as per Section 3.1.2. However, even if we do not mock the tracers used by SDSS to infer galaxies SFRs, we expect Hα to return SFRs on very short time-scales – roughly 5–

10 Myr – and therefore closer to the instantaneous SFR values that we adopt.

The results are shown in Fig. 9: global quenched fraction as a function of galaxy stellar mass (top panels), halocentric distance (middle panels), and host mass (bottom panels). We select galaxies,
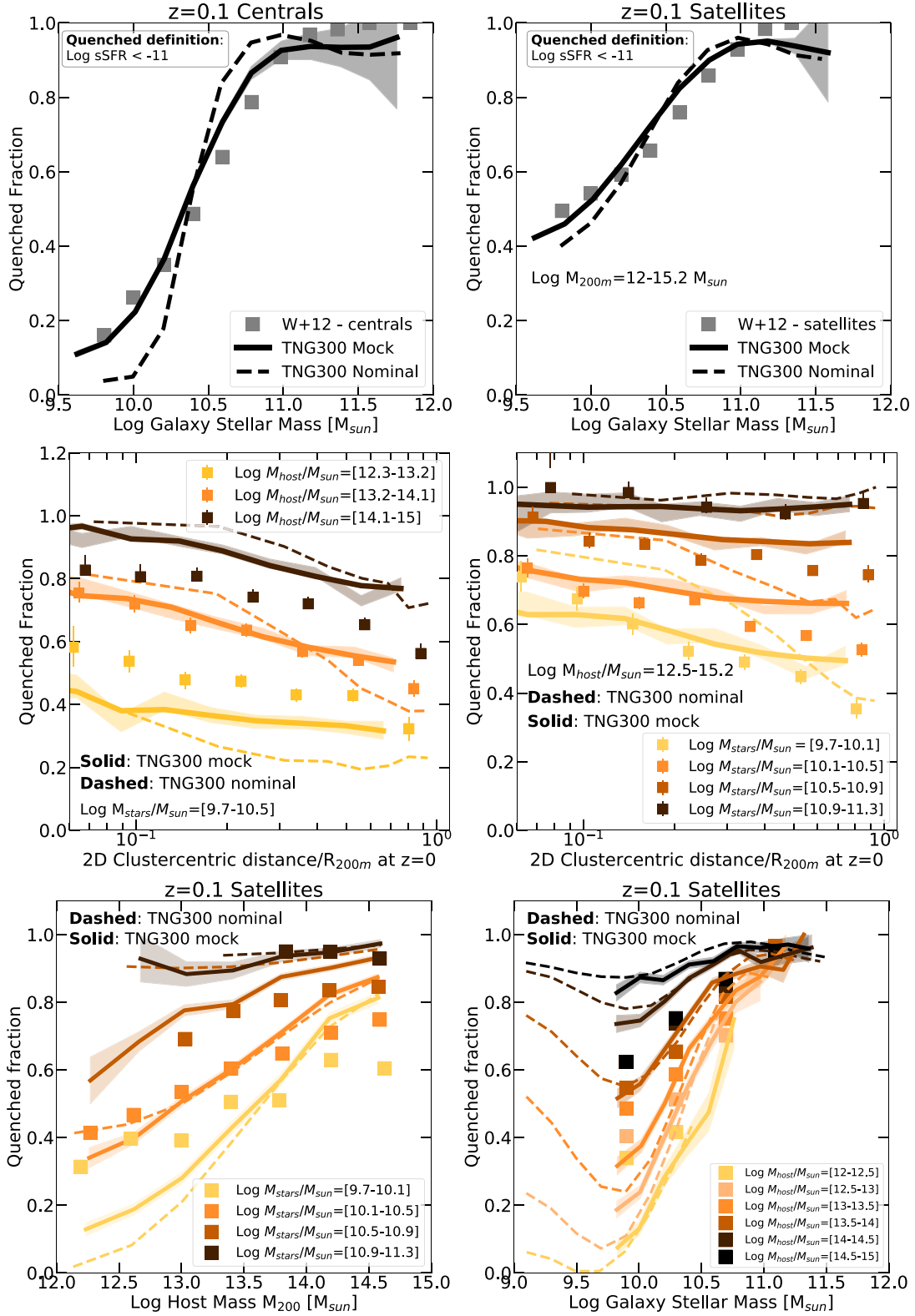
Figure 9. TNG300 nominal versus TNG300 mock versus SDSS results. Quenched fractions as a function of stellar mass (top), halocentric distance (middle), and host mass and galaxy stellar mass (bottom), at $z = 0.1$. In all panels, we show results of the nominal, i.e. direct TNG300 results (dashed) and the more-realistically mocked TNG300 results (solid; see the text for more details). Shaded areas represent the Poissonian errors. The global quantitative agreement between TNG and SDSS is excellent, as visible in the top row, while specific trends as a function of halocentric depend on the galaxy stellar and halo mass regime under consideration, generally being best for intermediate group-mass haloes.

centrals and/or satellites where appropriate, from the TNG300 snapshot at $z = 0.1$, the median observed redshift.

We denote with dashed curves the direct output of the simulation ('TNG300 nominal'), for which we adopt previous fiducial analysis choices: M_{200m} and R_{200m}, $\log(\text{sSFR}) < -11 \, \text{yr}^{-1}$, satellite selection based on spherical 3D geometry, unmatched mass distributions, and with no added systematic errors. In contrast, we show with solid curves the results obtained when applying the previously enumerated and more correct measurement choice sample selections ('TNG300 mock'). In all panels of Fig. 9, square symbols indicate observational data from SDSS (Wetzel et al. 2012, 2013), grey in the top panel and colour coded in the middle and bottom panels for different mass bins, which are matched to TNG.

The importance of the complete mock procedure is apparent in both top panels. As our ultimate objective is to make a statically robust and systematically correct comparison with data, the difference between the dashed (nominal) and solid (correctly mocked) lines is significant. Specifically, the full mock brings the quenched fractions of TNG galaxies, both centrals and satellites, ensemble averaged across the mass-matched galaxy population, into quantitative agreement with the SDSS results. Here, a key effect is the inclusion of the Gaussian errors, which make the trends with mass shallower and boost the low-mass normalization. These differences are small, but noticeable.

When comparing quenched fractions versus halocentric distance (middle panels), the un-mocked simulation results (dashed lines) only roughly reproduce the observed trends, and have various normalization offsets depending on distance and mass. As we have previously shown, the geometrical selection and distances based on either two-dimensional or three-dimensional methods are critical here (see Fig. 6). Once we consider the mock results (solid lines), which account for the projection effects and the possible effects of satellite/central misclassification, the agreement with the SDSS data at intermediate host ($10^{13-14} \, M_\odot$) and stellar masses ($10^{10-11} \, M_\odot$) is significantly better. While agreement at the group scale is excellent (see curve for host mass $10^{13.2-14.1} \, M_\odot$), the TNG quenched fractions for satellites in lower (higher) mass hosts are roughly lower (higher) by up to ~ 20 percentage points than observed (middle left panel).

The dependence on host halo mass is therefore stronger in the TNG simulations than in the data. Here we are fundamentally probing environmental processes, which are not related to SMBH-driven quenching and which are fully emerging features of the underlying physical model: Therefore, this excessively strong dependence on host mass points towards an issue in the interplay between stripping processes and the background halo medium, i.e. either in the restoring gravity of satellite galaxies and/or the background density of the intragroup medium. It is, however, unclear what aspect of the TNG model could be modified to impact this particular observable trend. On the one hand, the low-mass end satellite quenched fractions in TNG are consistent within 10–15 percentage points across variations in baryonic mass resolution of a factor of 1024 (see appendixes C and A of Donnari et al. 2021). On the other hand, as the quenching of low-mass satellites in high-mass hosts is due in similar proportions to pre-processing in smaller hosts at earlier times and environmental effects of the hosts in which they are found today (see Donnari et al. 2021), it is difficult to pinpoint in what mass and redshift regime the main culprit of the disagreement may lie. Finally, at the low-host mass end, the yellow curves and data points include satellite galaxies in hosts whose centrals have comparable stellar masses, i.e. small groups with possibly two dominant galaxies: These are systems for which larger systematic issues may arise also observationally in terms of, e.g. host mass labelling.

Considering the dependence on satellite stellar mass (middle right panel), we find generally good agreement to better than ~ 10 per cent. The only noticeable tension is that the data favour a quenched fraction that declines more rapidly with halocentric distance than in the simulations. This is particularly clear for the lowest mass satellites, below $M_{\text{stars}} < 10^{10} \, M_\odot$, where the data decline by nearly 40 per cent from the host centre to the virial radius, while the decline in TNG is less than 20 per cent. As above, this could be related to the radial profiles of background gas density in the hosts being too shallow, which would cause ram-pressure stripping to increase more weakly with decreasing radius. For instance, feedback from both supernovae and SMBHs are known to significantly distribute baryons in the CGM and IGM of galaxies in the TNG simulations (Nelson et al. 2019b; Pillepich et al. 2018a; Terrazas et al. 2020), and this baryonic redistribution could be indirectly probed by environmentally induced quiescence.

Finally, we stack satellites at all distances to provide an in-depth view of the results in the top right panel of Fig. 9. We show the dependence on host halo mass for different satellite mass bins (Fig. 9, lower left panel), as well as the dependence on satellite mass for different host mass bins (Fig. 9, lower right panel). As before, we see excellent agreement for intermediate-mass hosts, but the trend with host mass is somewhat too large in TNG, which results in quenched fractions below the SDSS data for the lowest mass hosts, and above the SDSS data for the most massive hosts. This is only true for intermediate- and low-mass satellites with $M_{\text{stars}} \lesssim 10^{10.5} \, M_\odot$ – more massive satellites are in good agreement, although here the quenched fractions are set by SMBH-driven quenching in the field (see Donnari et al. 2021). For the most massive clusters available in the sample, $M_{\text{host}} \simeq 10^{14.5} \, M_\odot$, the quenched fractions of intermediate (low)-mass TNG satellites are too high by ~ 15 (~ 20) per cent, as shown by the orange (yellow) lines.

Overall, we see that the impact of the mock procedure (solid versus dashed lines) is moderate for most of the relevant comparisons of these two bottom panels, rarely leading to changes of more than ~ 10 per cent in the regimes where data exist. We note that the good *global* agreement between TNG300 satellites and Wetzel et al. (2012) (top right panel) is driven by the broad host mass bin of $M_{200m} = 10^{12-15.2} \, M_\odot$. After we separate satellites into different host mass bins (bottom right panel), the level of agreement clearly depends on host mass, being best for hosts that are common though not the most sampled in SDSS (see Appendix B).

Finally, we recall that we have adopted instantaneous SFRs based on the gas in the TNG simulations. On the other hand, the SFRs in Wetzel et al. (2012) are derived from a combination of different tracers: Hα emission (for sSFR $> 10^{-11} \, \text{yr}^{-1}$), multiple emission lines (for $10^{-12} \, \text{yr}^{-1} < \text{sSFR} < 10^{-11} \, \text{yr}^{-1}$), or the amplitude of the 4000 Å Balmer break D_n4000 (for sSFR $< 10^{-12} \, \text{yr}^{-1}$), coupled to AGN contamination and fibre aperture corrections. Even more detailed future mocks will be required to assess the impact of such specific tracers.

4.3 TNG quenched fractions versus other models

We conclude with a comparison of several theoretical models, contrasting TNG versus the original Illustris simulation (Genel et al. 2014; Torrey et al. 2014; Vogelsberger et al. 2014a,b; Nelson et al. 2015; Sijacki et al. 2015), as well as EAGLE (Crain et al. 2015; Furlong et al. 2015; Schaye et al. 2015), RomulusC and Romulus25 (Tremmel et al. 2019), Hydrangea (a set of zoom-in clusters simulated with the EAGLE model; Bahé et al. 2017), and the semi-analytical model L-GALAXIES, which has been run directly

on the TNG dark matter-only simulations (Ayromlou et al. 2021; see Table 1 for more details). Similarly, the results we present for EAGLE are based on a re-analysis with the TNG version of SUBFIND, enabling an apples-to-apples comparison by ensuring that galaxies and their properties are identified and measured in exactly the same way.

For the remaining simulations (Romulus and Hydrangea), we note that results for quenched fractions are extracted from the literature, and thus may invoke different definitions and methodologies at a number of levels. Specifically, the TNG, Illustris, and EAGLE curves all use our fiducial definitions exactly, whereby e.g. the galaxy stellar masses and SFRs are evaluated within twice the stellar half-mass radius; however, for the other simulation results we do not have any information about the aperture within which the mass and SFR are evaluated. The L-Galaxies and Hydrangea results adopt the same quenched galaxy selection as that used for the Illustris, TNG, and EAGLE curves, while in Romulus the definition of quenched is based on a fit of the median values of the SFR within 0.1 dex bins of stellar mass in the range of 10^{8-10} $M_\odot$. We have shown, however (Section 3.1), that at low redshift, and particularly at $z = 0$, the quenched fractions do not change significantly between these two definitions.

Fig. 10 shows the quenched fractions of centrals (left-hand panel) and satellites (right-hand panel) at $z = 0$ (top) and $z = 2$ (bottom) as a function of galaxy stellar mass. For the TNG simulations, we show the Poissonian errors for TNG300 (denoted with shaded areas) and the cosmic variance uncertainties for TNG100 at $z = 0$ (denoted with striped areas). In principle, the jackknife error computed for TNG100 gives a lower estimate for the sample/cosmic variance error in the boxes smaller than TNG100, namely TNG50, Romulus, and Hydrangea. We compare all simulation results, where available, with different line colours, grouped by underlying galaxy formation model. For all models, we include satellites residing in hosts with $z = 0$ masses of $10^{14-14.5}$ $M_\odot$, which enables a comparison between the different models that otherwise have significantly different volumes and statistical sampling of massive hosts. As we have explored in our companion paper (Donnari et al. 2021), low-mass satellites are overall more quenched than centrals at fixed stellar mass. This holds true regardless of redshift, highlighting the role of the environment in the quenching process. Conversely, high-mass galaxies are quenched regardless of whether they are centrals or satellites, mainly due to SMBH feedback, at least in the TNG model. These two features are broadly common to all the simulations we explore here.

Specifically, at $z = 0$ and for all the models, the quenched fraction of centrals (upper left panel) is a monotonic function of the stellar mass, less than 10 per cent below $\sim 10^{10}$ $M_\odot$, increasing to ~ 50 per cent in EAGLE, Hydrangea, and RomulusC, and up to 80–100 per cent in TNG, Illustris, and L-Galaxies, particularly for galaxies more massive than $\sim 10^{11}$ $M_\odot$. Illustris, TNG, and L-Galaxies all show a relatively distinct transition between star-forming and quenched centrals, albeit at a different transitional mass scale.[4, 5] The origin of the different 'quenching mass scales' between TNG and

L-Galaxies, for instance, ultimately lies in the differing treatments of supermassive black hole feedback (Ayromlou et al. 2021), and current low-redshift data can discriminate on both the locus and steepness of the transition (see top left panel of Fig. 9). On the other hand, EAGLE, RomulusC, and Hydrangea show a weaker trend with galaxy stellar mass, with quenched fraction rising more slowly.

The quenched fractions of satellites versus mass at $z = 0$ (upper left panel) are more diverse. In Donnari et al. (2021), we have shown that this is a non-monotonic trend with stellar mass in TNG, which occurs due to the combination of environmental and mass quenching processes. Here we see that TNG100, TNG300, and TNG50 are all in good agreement, notwithstanding the different numerical resolution (see also Appendix C), with satellite quenched fractions of almost ~ 90 per cent, largely independent of stellar mass at this host mass scale. This is ~ 30 percentage points higher than that in Illustris. The three other hydrodynamical simulations also rise towards low mass: L-Galaxies is the only exception, with a monotonically increasing quenched fraction from ~ 60 to ~ 90 per cent across this mass range. RomulusC, Hydrangea, and EAGLE all have the opposite trend: lower quenched fractions for higher stellar masses.

At the high-mass end, when satellite galaxies exceed $\gtrsim 10^{10.5}$ $M_\odot$, differences start to become attributable to relative (in)efficiencies of SMBH feedback between the models. On the other hand, mismatches below this mass trace different efficiencies of environmental effects, which could be due to differences in background gas properties as discussed earlier, or to more subtle effects for instance in the numerical treatment of the hydrodynamics and resulting changes in gas stripping (Agertz et al. 2007; Sijacki et al. 2012).

Turning to higher redshifts, the models become much more varied in their predictions. Redshift $z = 2$ is a particularly illuminating regime, as current models often fail to reproduce the observed quenched fractions of massive galaxies, or similarly the observe space density of massive quenched galaxies (e.g. Valentino et al. 2020). For example, the original Illustris simulation produces essentially no quenched centrals already by $z = 2$ or 3, whereas the TNG model does (Donnari et al. 2019a).

The bottom panels of Fig. 10 show the fractions of quenched centrals and satellites at $z = 2$ for the available models. To take the time evolution of the star formation MS normalization with redshift, we change the definition of quenched galaxies to $\log{(\mathrm{sSFR/yr})} < -10$. We also change the host halo mass selection for satellites to a lower bin of $10^{12.5-13}$ $M_\odot$ to accommodate the available statistics. The differences in central quenched fractions at the massive end, above 10^{11} $M_\odot$, are clearly highly in the TNG model, and the same is also true for satellites. This suggests that both secular quenching through supermassive black hole feedback and environmentally driven quenching in high-density regions are processes that are both already in place in the early $z \sim 2$ Universe, as also inferred in recent observations (e.g. Glazebrook et al. 2017; Forrest et al. 2020).

5 SUMMARY AND CONCLUSIONS

Together with our companion paper (Donnari et al. 2021), we have used the IllustrisTNG simulations to investigate the diverse physical pathways of galaxy quenching and their observational signatures. In that work, we demonstrate how the TNG model produces a picture of both mass-driven and environmentally driven quenching. Here we focus on the large-volume TNG300 cosmological hydrodynamical

---

[4]We refer the reader to Weinberger et al. (2018) and Donnari et al. (2019a, 2021), for the main differences in quenching properties between TNG100 and TNG300, and between TNG and Illustris. The differences at high-mass end in TNG50 centrals in comparison to TNG100 and TNG300 are due to a combination of resolution effects and sample variance, as the volume of TNG50 is considerably smaller than the other TNG runs – see Appendix C for more details.

[5]Even if not shown, we have checked that the difference between TNG and L-Galaxies is not due to the different adopted stellar mass definition, whereby stellar masses are measured within twice the stellar half-mass radius for the TNG model and in practice accounting for all gravitationally bound stellar
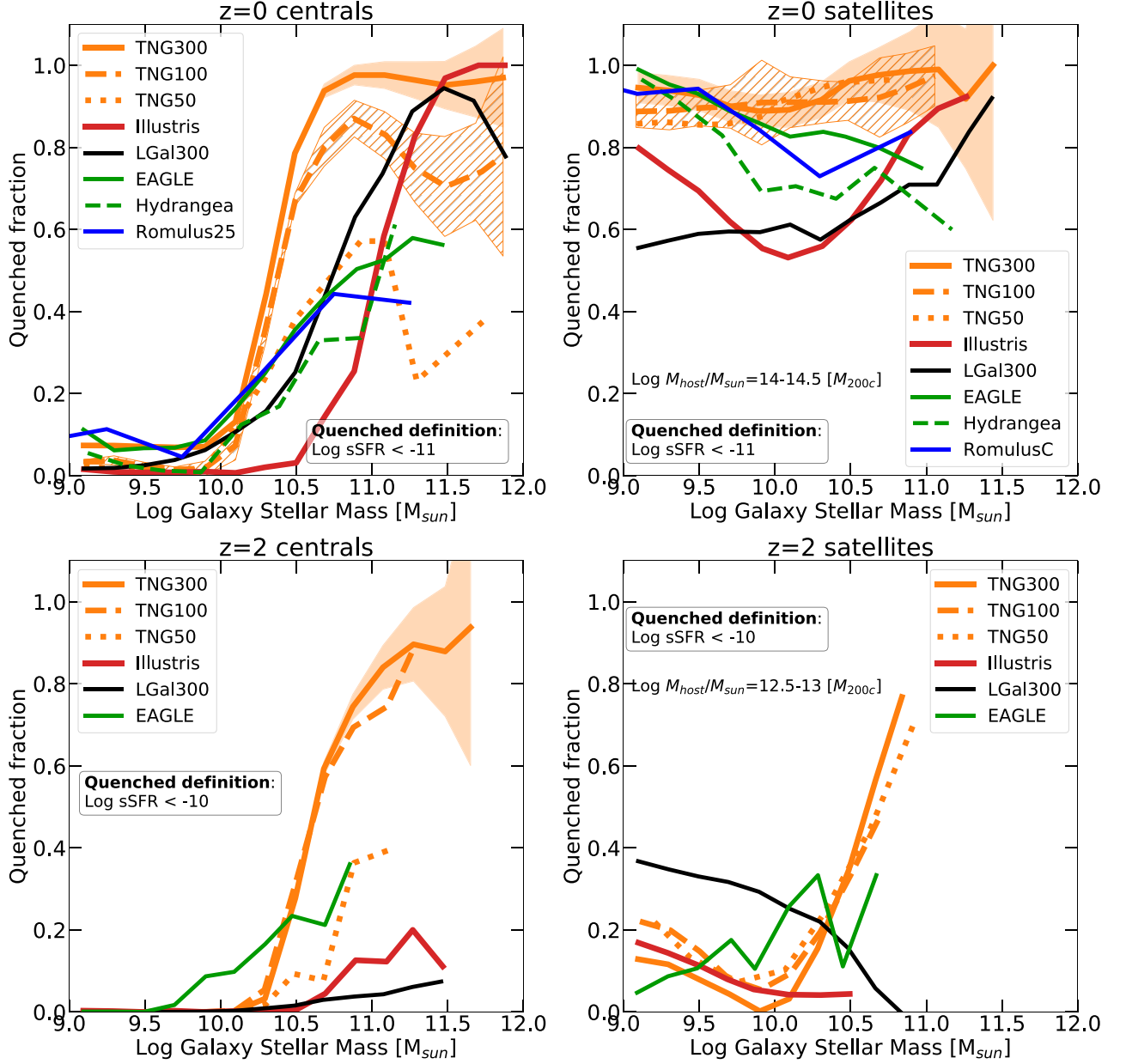
mass for L-Galaxies. See Appendix A for a quantification of such a choice on TNG300.

Figure 10. Comparison to other models. Quenched fractions of central galaxies (left) and satellite galaxies (right) at $z = 0$ (top) and $z = 2$ (bottom) in TNG300 (orange solid), TNG100 (orange dashed), TNG50 (orange dotted), Illustris (red), EAGLE and Hydrangea (green solid and dashed, respectively), RomulusC and Romulus25 (blue), and L-Galaxies (black). For TNG, Illustris, EAGLE, and Hydrangea curves, SFRs and stellar masses are measured within twice the stellar half-mass radius. For the satellite selections, we uniformly take all galaxies above a stellar mass of 10^9 $M_\odot$ within hosts of mass $10^{14-14.5}$ $M_\odot$ at $z = 0$ and $10^{12.5-13}$ at $z = 2$. Orange shaded areas denote estimates of the Poisson errors, shown for TNG300 for reference, whereas the striped areas around TNG100 give estimates of the cosmic variance uncertainties. While the trends of quenched fraction with mass for $z = 0$ centrals show the same qualitative trends between models, they differ in the details, particularly in the rapidity of the transition towards the quenched population. Satellite galaxy quenched fractions are even more disparate, with different models having a diversity of trends with mass. High redshift ($z = 2$) offers one of the most interesting regimes for comparison, as some models generate substantiative populations of quenched galaxies at these early epochs, while others do not.

simulation to: (i) study systematic issues in inferring quenched fractions, (ii) quantitatively compare TNG quenched fractions against low-redshift observational data, and (iii) contrast TNG with other theoretical models. Our key results are as follows:

(i) *Measurement choices*: We study the role played by different measurement choices, definitions, and sample selection issues in shaping the quenched fractions of galaxies (centrals and satellites in the 10^{9-12} $M_\odot$ stellar mass range) from $0 \leq z \leq 3$. Different

criteria to define 'quenched' translate into population-wide quenched fractions that differ by up to 70 (30) percentage points for centrals (satellites), particularly at the high-mass end of $M_\star \gtrsim 10^{10.5}$ $M_\odot$. Adopting a fixed threshold in sSFR at different redshifts fails to identify quenched galaxies due to the evolution of the star formation MS with time (Fig. 1). At high redshift $z = 2$–3, the averaging time-scale (Fig. 2) and the physical aperture in which SFR is measured have strong impacts on the inferred quenched fractions. For massive galaxies, larger apertures produce lower quenched fractions, by up to

∼ 50 per cent (Fig. 3). We show that adding systematic uncertainties to theoretical values for stellar mass and SFR is important when comparing to observational data, as it changes not only the values of quenched fractions for low-mass galaxies, but also the trend and/or slope with mass (Fig. 4).

(ii) *Sample selection*: Adopting a satellite selection that accounts for the two-dimensional projection effects and geometry inherent in observations, rather than theory-space choices more common in simulations, reduces the quenched fractions as a function of halocentric distance by ∼10–30 percentage points, especially for low-mass satellites ($\lesssim 10^{10.5}$ M$_\odot$) and hosts ($\lesssim 10^{14}$ M$_\odot$): Fig. 6. Quenched fractions as a function of redshift evaluated for galaxies above a minimum stellar mass threshold depend sensitively on the shape of the sample mass distributions. For example, a flat in log mass, rather than volume-limited, sample leads to inferred quenched fractions higher by ∼20 percentage points, at all redshifts, for all but the most massive galaxy samples (Fig. 7).

(iii) *Comparison to observations*: We make a quantitative comparison to observational results at low redshift, with a particular focus on the SDSS catalogue, for which we have created detailed mocks by matching the key analysis choices of Wetzel et al. (2012, 2013). We find that the TNG quenched fractions compare well with a selection of observational estimates, for both centrals and satellites (Fig. 8), being in striking quantitative agreement for global central and satellite quenched fractions at $z = 0.1$. For satellites, as a function of halocentric distance, the level of (dis)agreement depends on the host halo and galaxy stellar mass regime under consideration, being best for intermediate, group-mass scale hosts ($\lesssim 10^{13-14}$ M$_\odot$), and worst for the lowest mass satellites in either the lowest ($\sim 10^{12}$ M$_\odot$) or highest mass hosts ($\sim 10^{14.5-15}$ M$_\odot$, Fig. 9).

(iv) *Comparison to other models*: We contrast the fraction of quenched TNG galaxies versus other theoretical models: Illustris, Eagle, Hydrangea, Romulus25/C, and the semi-analytical model L-Galaxies. Although the trends of quenched fraction with galaxy mass for $z = 0$ centrals have qualitatively similar behaviour, the models differ in how rapidly the population transitions to predominantly quenched. Satellite quenched fractions are even more diverse – although all models produce a large fraction of quenched satellites in hosts with masses of $10^{14-14.5}$ M$_\odot$, the normalization and even trend with galaxy mass can differ. At $z = 2$, some models produce a significant population of quenched galaxies, while others do not. This fraction, for massive centrals, is higher in TNG than other models, reaching ∼ 80 per cent already at this early epoch. Together with satellite quenching at $z = 2$, this shows how environment and mass quenching in the early Universe are a model-dependent, and thus informatively constraining, phenomenon (Fig. 10).

Our study highlights the importance of carefully considering the many observational effects and choices that can play a role in any comparison between theory and data, particularly for a quantity as complex as the quenched fractions of galaxies. Our results suggest that the only way to make meaningful comparisons across galaxy samples, in fact of any type and origin, is by matching the operational definitions and choices for quenched galaxies, galaxy apertures, SFR tracers, sample mass distributions, etc. Overall, we have also validated the outcome of the TNG model, demonstrating its broad agreement with observation data, although here we have focused on the local Universe at low redshift. The signatures of different feedback models and the ways in which different simulations generate the dichotomy between the star-forming and quenched galaxy populations are encoded into the details of galaxy quiescence, which offers a powerful probe on theoretical models for galaxy evolution.

Future, careful, and targeted comparisons to other observational data sets will be invaluable, particularly at intermediate and high redshifts.

DATA AVAILABILITY

Data directly related to this publication and its figures are available on request from the corresponding author. The IllustrisTNG simulations themselves, including the more recent TNG50 data, are publicly available and accessible at www.tng-project.org/data.

REFERENCES

Agertz O. et al., 2007, MNRAS, 380, 963
Anand A., Nelson D., Kauffmann G., 2021, MNRAS, 504, 65
Appleby S., Davé R., Kraljic K., Anglés-Alcázar D., Narayanan D., 2020, MNRAS, 494, 6053
Asquith R. et al., 2018, MNRAS, 480, 1197
Ayromlou M., Nelson D., Yates R. M., Kauffmann G., Renneby M., White S. D. M., 2021, MNRAS, 502, 1051
Bahé Y. M. et al., 2017, MNRAS, 470, 4186
Baugh C. M. et al., 2019, MNRAS, 483, 4922
Bisigello L., Caputi K. I., Grogin N., Koekemoer A., 2018, A&A, 609, A82
Bluck A. F. L. et al., 2016, MNRAS, 462, 2559
Bravo M., Lagos C. D. P., Robotham A. S. G., Bellstedt S., Obreschkow D., 2020, MNRAS, 497, 3026
Brennan R. et al., 2015, MNRAS, 451, 2933
Campbell D., van den Bosch F. C., Hearin A., Padmanabhan N., Berlind A., Mo H. J., Tinker J., Yang X., 2015, MNRAS, 452, 444
Chabrier G., 2003, PASP, 115, 763
Crain R. A. et al., 2015, MNRAS, 450, 1937
Croton D. J. et al., 2006, MNRAS, 365, 11
Croton D. J. et al., 2016, ApJS, 222, 22
Darvish B., Mobasher B., Martin D. C., Sobral D., Scoville N., Stroe A., Hemmati S., Kartaltepe J., 2017, ApJ, 837, 16
Davies L. J. M. et al., 2019, MNRAS, 483, 5444
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Donnari M. et al., 2019a, MNRAS, 485, 4817
Donnari M. et al., 2019b, MNRAS, 489, 3036
Donnari M. et al., 2021, MNRAS, 500, 4004
Engler C. et al., 2021, MNARS, 500, 3957
Fang J. J. et al., 2018, ApJ, 858, 100
Fontanot F., De Lucia G., Monaco P., Somerville R. S., Santini P., 2009, MNRAS, 397, 1776
Fontanot F. et al., 2020, MNRAS, 496, 3943
Forrest B. et al., 2020, ApJ, 890, L1

Förster Schreiber N. M. et al., 2018, ApJS, 238, 21
Fossati M. et al., 2017, ApJ, 835, 153
Franx M., van Dokkum P. G., Förster Schreiber N. M., Wuyts S., Labbé I., Toft S., 2008, ApJ, 688, 770
Furlong M. et al., 2015, MNRAS, 450, 4486
Gao J., Zou H., Zhou X., Kong X., 2020, PASP, 132, 024101
Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, ApJ, 757, 85
Genel S. et al., 2014, MNRAS, 445, 175
Genel S. et al., 2018, MNRAS, 474, 3976
Glazebrook K. et al., 2017, Nature, 544, 71
Gunn J. E., Gott J. R. III, 1972, ApJ, 176, 1
Hahn C. et al., 2019, ApJ, 872, 160
Haines C. P., Gargiulo A., Merluzzi P., 2008, MNRAS, 385, 1201
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Henriques B. M. B., Yates R. M., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P. A., White S. D. M., 2020, MNRAS, 491, 5795
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Martini P., Robertson B., Springel V., 2005, ApJ, 630, 705
Jian H.-Y. et al., 2018, PASJ, 70, S23
Katsianis A. et al., 2020, MNRAS, 492, 5592
Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692
Leja J. et al., 2019, ApJ, 877, 140
Leja J., Speagle J. S., Johnson B. D., Conroy C., van Dokkum P., Franx M., 2020, ApJ, 893, 111
Lim S. H., Mo H. J., Lu Y., Wang H., Yang X., 2017, MNRAS, 470, 2982
Lin L. et al., 2014, ApJ, 782, 33
Lovell M. R. et al., 2018, MNRAS, 481, 1950
McGee S. L., Balogh M. L., Wilman D. J., Bower R. G., Mulchaey J. S., Parker L. C., Oemler A., 2011, MNRAS, 413, 996
Marinacci F. et al., 2018, MNRAS, 480, 5113
Merritt D., 1983, ApJ, 264, 24
Muzzin A. et al., 2013, ApJ, 777, 18
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nelson D. et al., 2015, Astron. Comput., 13, 12
Nelson D. et al., 2018a, MNRAS, 475, 624
Nelson D. et al., 2018b, MNRAS, 477, 450
Nelson D. et al., 2019a, Comput. Astrophys. Cosmol., 6, 2
Nelson D. et al., 2019b, MNRAS, 490, 3234
Old L. et al., 2014, MNRAS, 441, 1513
Old L. et al., 2015, MNRAS, 449, 1897
Old L. et al., 2018, MNRAS, 475, 853
Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392
Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Pillepich A. et al., 2019, MNRAS, 490, 3196
Pintos-Castro I., Yee H. K. C., Muzzin A., Old L., Wilson G., 2019, ApJ, 876, 40
Poggianti B. M. et al., 2017, ApJ, 844, 48
Quadri R. F., Williams R. J., Franx M., Hildebrandt H., 2012, ApJ, 744, 88
Salim S. et al., 2007, ApJS, 173, 267
Salpeter E. E., 1955, ApJ, 121, 161
Schaefer A. L. et al., 2019, MNRAS, 483, 2851
Schaye J. et al., 2015, MNRAS, 446, 521
Sherman S. et al., 2020, MNRAS, 499, 4239
Sijacki D., Vogelsberger M., Kereš D., Springel V., Hernquist L., 2012, MNRAS, 424, 2999
Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Springel V., 2010, MNRAS, 401, 791
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2018, MNRAS, 475, 676
Stevens A. R. H. et al., 2019, MNRAS, 483, 5334

Stevens A. R. H. et al., 2021, MNARS, 502, 3158
Terrazas B. A. et al., 2020, MNRAS, 493, 1888
Tinker J. L., 2020, preprint (arXiv:2007.12200)
Tinker J., Wetzel A., Conroy C., 2011, preprint (arXiv:1107.5046)
Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, MNRAS, 438, 1985
Torrey P. et al., 2018, MNRAS, 477, L16
Tremmel M. et al., 2019, MNRAS, 483, 3336
Valentino F. et al., 2020, ApJ, 889, 93
Vogelsberger M. et al., 2014a, MNRAS, 444, 1518
Vogelsberger M. et al., 2014b, Nature, 509, 177
Vogelsberger M. et al., 2018, MNRAS, 474, 2073
Wagner C. R., Courteau S., Brodwin M., Stanford S. A., Snyder G. F., Stern D., 2017, ApJ, 834, 53
Wang E., Wang H., Mo H., van den Bosch F. C., Lim S. H., Wang L., Yang X., Chen S., 2018, ApJ, 864, 51
Weinberger R. et al., 2017, MNRAS, 465, 3291
Weinberger R. et al., 2018, MNRAS, 479, 4056
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
Whitaker K. E. et al., 2010, ApJ, 719, 1715
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879
Wuyts S. et al., 2011, ApJ, 742, 96
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153

APPENDIX A: EFFECTS OF APERTURE ON THE GALAXY STELLAR MASS

Whereas throughout the paper we pay great attention on the way the quenched fractions are measured, also differences in the way galaxy stellar masses are estimated may a priori have an impact on how the quenched fractions from different samples compare. In Fig. A1, we quantify this possibility by showing the quenched fractions from TNG300, at fixed definition, versus galaxy stellar masses: The latter
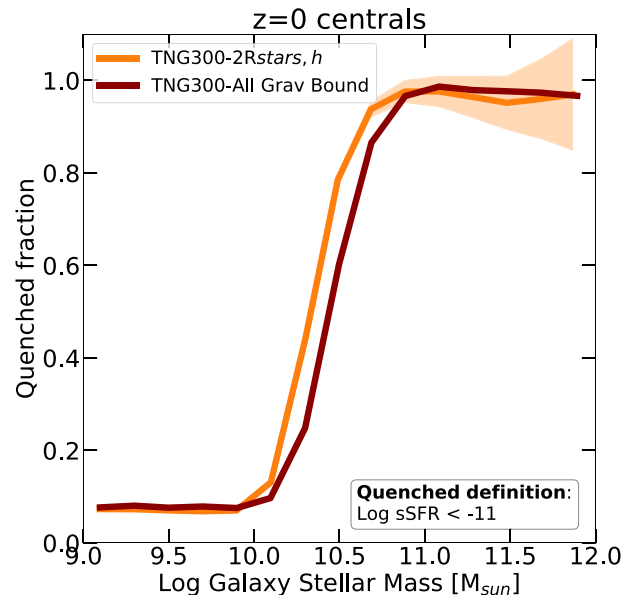

Figure A1. Effect of aperture on the galaxy stellar mass. Quenched fractions of TNG300 centrals at $z = 0$ for two different stellar mass definitions: within twice the stellar half-mass radius (orange curve) and accounting for all gravitationally bound particles (dark red curve). Different stellar mass definitions imply a slightly shifted quenched fraction in the *x*-axis direction.

are obtained within twice the stellar half-mass radius (orange curve, as in the main body of the paper) and accounting for all the stellar particles that are gravitationally bound to a galaxy's subhalo (red curve). A different stellar mass choice implies a shift in the *x*-axis direction, but this is noticeable only in the steep portion of the trend, where it can be mistaken for a difference in quenched fraction of up to 30 percentage points at fixed galaxy mass.

APPENDIX B: HOST AND STELLAR MASS DISTRIBUTIONS: TNG VERSUS SDSS SAMPLE

As discussed in Section 4.2, in order to properly compare the outcome of TNG300 with the SDSS data we create a mock catalogue of TNG300 satellites and centrals, by matching measurement choices and sample selections as adopted in Wetzel et al. (2012, 2013) (see Fig. 9). Importantly, since the SDSS catalogues used to compare with TNG include only galaxies with stellar mass in the range $M_{stars} = 10^{9.7-11.6}$ M$_\odot$ and haloes in the range $M_{200m} = 10^{11.4-14.9}$ M$_\odot$, we match the stellar mass distribution of both centrals and satellites, separately, as well as the host mass distribution and halocentric distance (see Tinker et al. 2011, for more details). Practically, for each SDSS galaxy with given properties, we randomly select a TNG300 galaxy satisfying: a stellar mass within 0.2 dex, a host mass M_{200m} within 0.1 dex, and a 2D projected distance/R_{200m} below 10 per cent from the corresponding observed values. We note that the host mass distribution and distances from the cluster centre have been matched only for satellite galaxies.

In Fig. B1, we show the resulting mass distributions of centrals (left-hand panels), satellites (middle panels), and host masses (right-hand panels), for TNG300 (orange curves) and SDSS data (black curves). The top panels show the intrinsic distribution in the TNG300 simulation. The bottom panels show the mocked distributions, which are used for the observational comparisons.

APPENDIX C: NUMERICAL RESOLUTION EFFECT ON QUENCHED FRACTIONS

We show here the resolution effect for the different resolution realizations of TNG50, namely TNG50-2 and TNG50-3 characterized by a baryonic mass of 6.8×10^5 and 5.4×10^6 M$_\odot$, respectively, in comparison to 8.5×10^4 M$_\odot$ for the flagship run.

In Fig. C1, we show the quenched fractions of centrals (left) and satellites (right) at $z = 0$ according to the three resolution levels of TNG50 (green), represented with different curve thicknesses. In all simulations, we select satellites residing in hosts with masses of 10^{13-14} M$_\odot$. To separate quenched and star-forming galaxies, we use log (sSFR/yr)< -11. For comparison, we also show results of TNG100, TNG100-2 (blue), and TNG300 (orange). We remind here that TNG100 has a mass resolution in between TNG50-2 and TNG50-3, while TNG100-2 and TNG300 have a mass resolution somewhat similar to TNG50-3.

For centrals, as already shown in Donnari et al. (2021), a lower resolution returns a higher quenched fraction at fixed galaxy stellar mass, by which amount depending on mass scale. For satellites, we see here that, while the quenched fractions agree very well
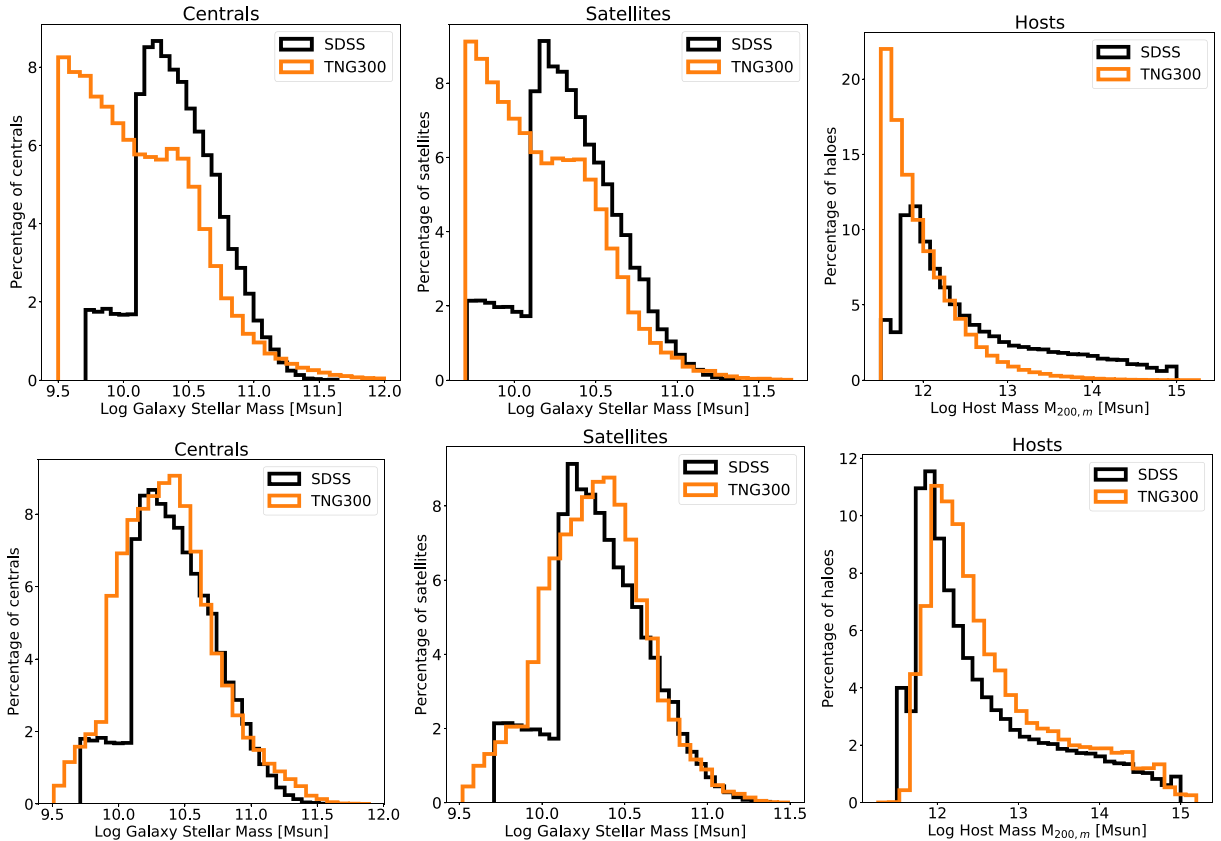

Figure B1. Host and stellar mass distribution of TNG300 and SDSS. The mass distribution of centrals (left), satellites (middle), and hosts (right) in TNG300 (orange) and SDSS (black). The distribution of SDSS galaxies is the same in the top and bottom rows, while for TNG300 we show the distribution of the raw data, i.e. the output of the simulation (top), and the distribution of the mock catalogue (bottom).
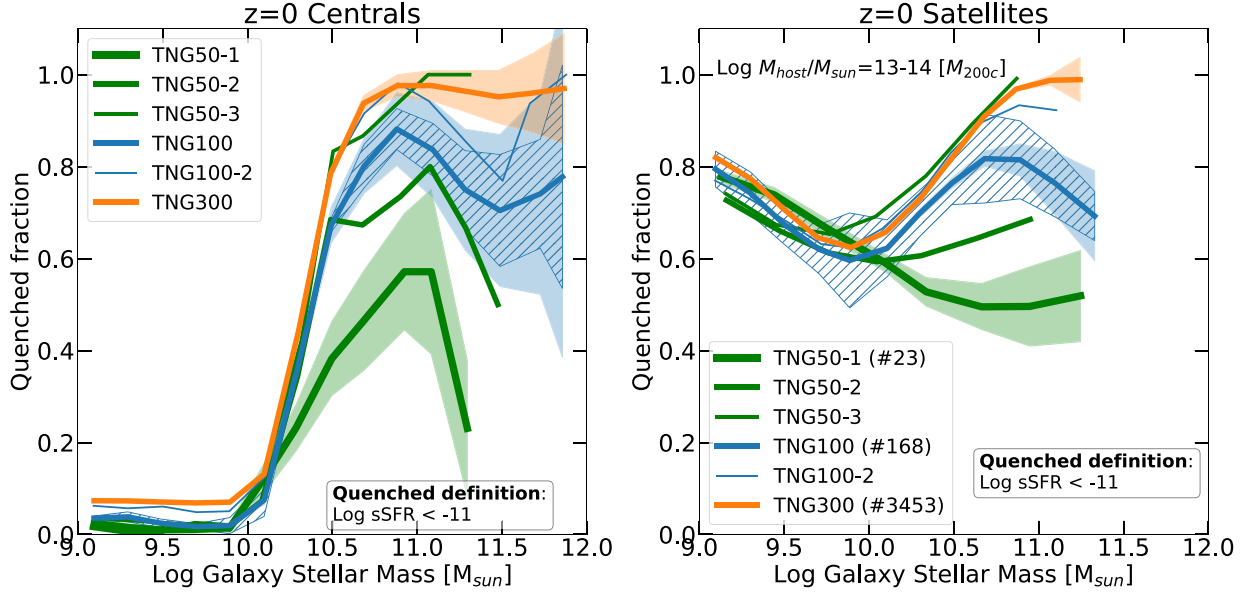
Figure C1. Resolution study of the quenched fractions as a function of stellar mass. Fraction of quenched galaxies as a function of stellar mass for $z = 0$ centrals (left) and satellites (right) according to the three different resolution levels of TNG50 (green), two resolution levels of TNG100 (blue), and TNG300 (orange). On the right-hand panel, in parenthesis we show the number of hosts with masses of 10^{13-14} $M_\odot$.
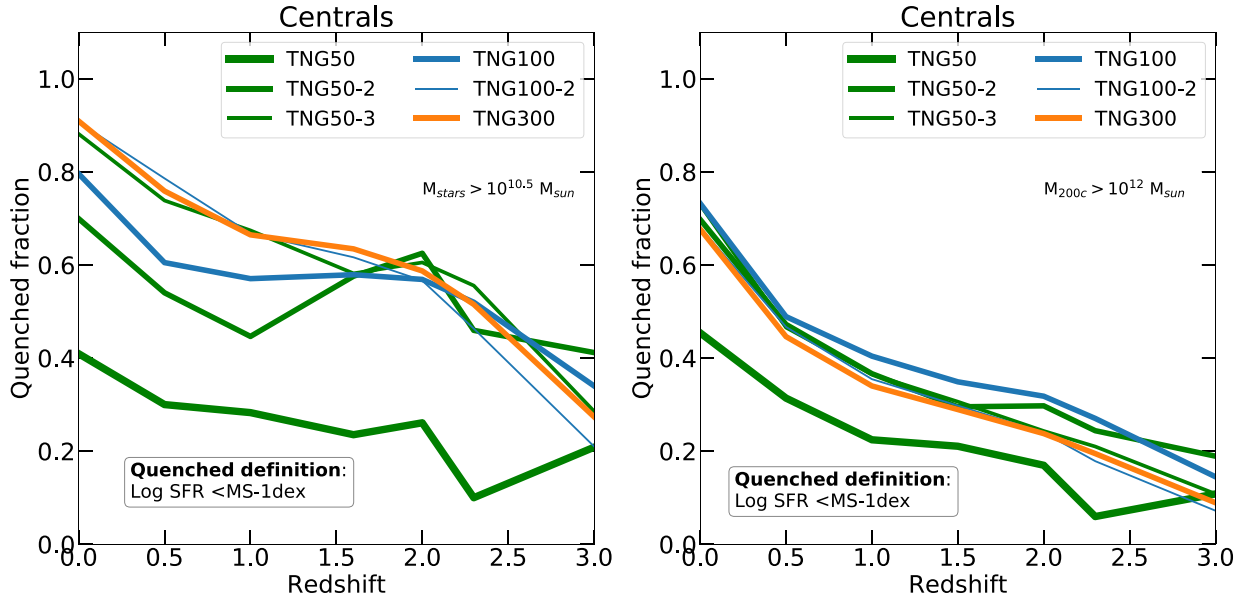

Figure C2. Resolution study of the quenched fractions as a function of redshift. Fraction of quenched centrals as a function of redshift for all central galaxies with $M_{\rm stars} > 10^{10.5}$ $M_\odot$ (left) and for all central galaxies residing in haloes with $M_{\rm 200c} > 10^{12}$ $M_\odot$ (right), according to the three different resolution levels of TNG50 (green), the two resolution levels of TNG100 (blue), and the highest resolution level of TNG300 (orange).

(within 10 per cent) across all the TNG resolution levels and across all satellite masses for more massive hosts (see top right panel of Fig. 10), for group-mass hosts (10^{13-14} $M_\odot$) the TNG simulations agree to a similar degree only at the satellite low-mass end, while exhibiting a resolution-dependent trend for massive galaxies, $M_{\rm stars} \gtrsim 10^{10.5}$ $M_\odot$. This would seem to suggest that, at least in TNG, the effects of environmental quenching processes are less susceptible to numerical resolution than the effects of SMBH feedback. However, besides the resolution, the different statistical sampling across the different TNG-simulated volumes plays a role in determining the differences we see at the high-mass end of both centrals and satellites, since the volume of TNG50 is considerably smaller than TNG100 and TNG300. At $z = 0$ in TNG50 there are 23 group-mass hosts, compared to, e.g. 168 in TNG100. We have verified that when we randomly downsample the TNG100 host population to 23 objects as is the case for TNG50, in about 30–40 per cent of the cases also TNG100 returns quenched fractions similar or even lower than the average ones of TNG50 at the $M_{\rm stars} \sim 10^{11}$ $M_\odot$ mass scale. This suggests that the differences between TNG50 and TNG100/TNG300 that we see in Fig. C1 are due to a combination of resolution effects and sample variance, the former probably to a larger degree.

Fig. C2 shows the quenched fractions as a function of redshift for all centrals with stellar mass larger than $10^{10.5}$ M$_\odot$ (left-hand panel) and for all central galaxies in haloes with $M_{200c} > 10^{12}$ M$_\odot$ (right-hand panel). Similarly to Fig. C1, we show the three resolution runs of TNG50 (green), two runs of TNG100 (blue) and TNG300 (orange) spanning a range of particle mass resolutions across a factor larger than 100.

Overall, when galaxies are selected according to their stellar mass (left-hand panel), the lower resolution runs return higher quenched fractions, with the actual amount depending on the redshift and simulation. Indeed, while TNG50 shows quenched fractions lower than all the other runs regardless of redshift ($\sim$10 percentage points at $z = 3$ and up to 40–50 percentage points at $z = 0$), at $z < 3$ TNG300, TNG100-2, and TNG50-3 agree well with TNG100, showing a quenched fraction of about 40 per cent on average at, e.g. $z = 2$–3. Interestingly, while all runs return a quenched fraction that is generally monotonically decreasing with redshift, individual runs may exhibit fluctuations at specific times, possibly due to sample variance.

On the right-hand panel of Fig. C2, central galaxies are taken with a cut in their host halo mass instead of galaxy stellar mass. Here, the resolution trends seen so far (higher fractions for lower resolution runs) are much less pronounced. TNG100, TNG300, TNG50-2, and TNG50-3 return almost the same quenched fractions at $z \lesssim 3$ within 10 percentage points. As in the left-hand panel, TNG50 shows lower fractions than all the other runs at all redshifts, but, in this case, the difference is smaller, capped at 20 percentage points of differences at most.

The results presented in Figs C1 and C2 suggest that, within the IllustrisTNG model, numerical resolution impacts the results for the quenched fraction versus galaxy stellar mass or versus redshift also because it impacts the stellar mass of galaxies that form within haloes (see discussions in Pillepich et al. 2018b; Engler et al. 2021). Furthermore, an improved implementation of the subgrid scheme for SMBH feedback that is less susceptible to numerical resolution, in fact, may lessen the differences. We remind here that, besides the softenings, also the BH kernel-weighted neighbour number (Weinberger et al. 2017) that defines the BH accretion and feedback region is changed with resolution, with a scaling based on the baryonic target mass (Pillepich et al. 2018a). A different scaling might help enhance the convergence of the quenched fractions of high-mass galaxies in the TNG model across the whole range of resolution levels probed by the TNG runs.

This paper has been typeset from a T$_{\rm E}$X/LAT$_{\rm E}$X file prepared by the author.