Testing the key role of the stellar mass–halo mass relation in galaxy merger rates and morphologies via DECODE, a novel Discrete statistical sEmi-empiriCal mODEl

Hao Fu, Francesco Shankar, Mohammadreza Ayromlou, Max Dickson, Ioanna Koutsouridou, Yetli Rosas-Guevara, Christopher Marsden, Kristina Brocklebank, Mariangela Bernardi, Nikolaos Shiamtanis, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **516,** 3206–3233 (2022) https://doi.org/10.1093/mnras/stac2205

Testing the key role of the stellar mass–halo mass relation in galaxy merger rates and morphologies via DECODE, a novel Discrete statistical sEmi-empiriCal mODEl

Hao Fu,[1★] Francesco Shankar,[1] Mohammadreza Ayromlou,[2,3] Max Dickson,[1] Ioanna Koutsouridou,[4,5] Yetli Rosas-Guevara,[6] Christopher Marsden,[1] Kristina Brocklebank,[1] Mariangela Bernardi,[7] Nikolaos Shiamtanis,[1] Joseph Williams,[1] Lorenzo Zanisi,[1] Viola Allevato,[8] Lumen Boco,[9] Silvia Bonoli,[6] Andrea Cattaneo,[10] Paola Dimauro,[11] Fangzhou Jiang,[12] Andrea Lapi,[9] Nicola Menci,[13] Stefani Petropoulou[1] and Carolin Villforth[14]

[1]*School of Physics and Astronomy, University of Southampton, Highfield SO17 1BJ, UK*
[2]*Universität Heidelberg, Zentrum für Astronomie, Institut für theoretische Astrophysik, Albert-Ueberle-Str 2, D-69120 Heidelberg, Germany*
[3]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str 1, D-85741 Garching bei München, Germany*
[4]*Dipartimento di Fisica e Astronomia, Universita degli Studi di Firenze, Via G. Sansone 1, I-50019 Sesto Fiorentino, Italy*
[5]*INAF/Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*
[6]*Donostia International Physics Centre (DIPC), Paseo Manuel de Lardizabal 4, E-20018 Donostia-San Sebastian, Spain*
[7]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[8]*INAF – Osservatorio di Astrofisica e Scienza delle Spazio di Bologna, OAS, Via Gobetti 93/3, I-40129 Bologna, Italy*
[9]*SISSA, Via Bonomea 265, I-34135 Trieste, Italy*
[10]*Observatoire de Paris/LERMA, PSL University, 61 av. de l'Observatoire, F-75014 Paris, France*
[11]*Observatório Nacional, Ministério da Ciencia, Tecnologia, Inovação e Comunicações, São Cristóvão, 20921-400 Rio de Janeiro, Brazil*
[12]*TAPIR, California Institute of Technology, Pasadena, CA 91125, USA*
[13]*INAF – Osservatorio Astronomico di Roma, via di Frascati 33, I-00078 Monte Porzio Catone, Italy*
[14]*Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK*

Accepted 2022 July 27. Received 2022 June 22; in original form 2022 April 19

ABSTRACT

The relative roles of mergers and star formation in regulating galaxy growth are still a matter of intense debate. We here present our DECODE, a new Discrete statistical sEmi-empiriCal mODEl specifically designed to predict rapidly and efficiently, in a full cosmological context, galaxy assembly, and merger histories for any given input stellar mass–halo mass (SMHM) relation. DECODE generates object-by-object dark matter merger trees (hence discrete) from accurate subhalo mass and infall redshift probability functions (hence statistical) for all subhaloes, including those residing within other subhaloes, with virtually no resolution limits on mass or volume. Merger trees are then converted into galaxy assembly histories via an input, redshift-dependent SMHM relation, which is highly sensitive to the significant systematics in the galaxy stellar mass function and on its evolution with cosmic time. DECODE can accurately reproduce the predicted *mean* galaxy merger rates and assembly histories of hydrodynamic simulations and semi-analytical models, when adopting in input their SMHM relations. In this work, we use DECODE to prove that only SMHM relations implied by stellar mass functions characterized by large abundances of massive galaxies and significant redshift evolution, at least at $M_\star \gtrsim 10^{11}\,\mathrm{M}_\odot$, can simultaneously reproduce the local abundances of satellite galaxies, the galaxy (major merger) pairs since $z \sim 3$, and the growth of Brightest Cluster Galaxies. The same models can also reproduce the local fraction of elliptical galaxies, on the assumption that these are strictly formed by major mergers, but not the full bulge-to-disc ratio distributions, which require additional processes.

Key words: galaxies: abundances – galaxies: evolution – galaxies: general – galaxies: haloes.

1 INTRODUCTION

The field of galaxy formation and evolution is still far from settled, with several open and still hotly debated issues. For example, it is still unclear what are the relative amounts of stellar mass that galaxies grow '*in situ*', via star formation, and acquire '*ex situ*' from, e.g. mergers with other galaxies (e.g. Guo & White 2008; Oser et al. 2010; Cattaneo et al. 2011; Lackner et al. 2012; Lee & Yi 2013; Pillepich et al. 2014; Rodriguez-Gomez et al. 2016; Qu et al. 2017; Clauwens et al. 2018; Pillepich et al. 2018b; Monachesi et al. 2019; Davison et al. 2020). In a Lambda cold dark matter

★ E-mail: h.fu@soton.ac.uk

(ΛCDM) Universe, galaxies are in fact believed to live at the centre of dark matter (DM) haloes, which grow their mass via mergers with other DM haloes along with smooth mass accretion from their environments (e.g. Murali et al. 2002; Conselice & Arnold 2009; Genel et al. 2010; L'Huillier, Combes & Semelin 2012). Each merger between two DM haloes could in principle trigger a merger between their central galaxies and, therefore, galaxy mergers should indeed be frequent in a DM-dominated Universe. However, cosmological models suggest that in many instances the dynamical friction timescale, i.e. the time that two galaxies take to merge, is longer than the age of the Universe, resulting in the smaller galaxy orbiting as an unmerged satellite of the most massive central galaxy (e.g. Khochfar & Burkert 2006; Fakhouri, Ma & Boylan-Kolchin 2010; McCavana et al. 2012). A prominent case is the orphan satellite galaxies, whose DM subhaloes can no longer be resolved in the simulations, but they continue orbiting the central galaxy. It is thus clear that to impose more stringent constraints on the role of mergers in shaping galaxies in a ΛCDM Universe, it is first of all essential to correctly predict the merger histories of the host DM haloes. After this, the following vital step is to identify the correct mapping between galaxies and host DM haloes to translate DM merger trees into a galaxy assembly history, a task far from trivial (e.g. Hopkins et al. 2010a; Grylls et al. 2019, 2020a).

It has been noted that widely distinct merger histories could lead to similar morphologies and kinematic properties in the remnant galaxies (Bournaud, Jog & Combes 2007). Moreover, different hierarchical models often predict strongly divergent balances between the stellar mass formed *in situ* during the early epoch, highly star-forming and dust-enshrouded phase, and the fraction of stellar mass acquired *ex situ* via mergers. For example, the SAM presented in González et al. (2011) suggests that only a few per cent of the final stellar mass is formed in a typical massive galaxy during its initial burst, while at the other extreme, several groups suggest that most of the stellar mass was acquired in a moderate-to-strong burst of star formation at high redshifts (e.g. Granato et al. 2004; Chiosi, Merlin & Piovan 2012; Merlin et al. 2012; Lapi et al. 2018).

Semi-empirical models (SEMs) have been introduced in the last decades as a powerful, *complementary* tool to probe galaxy evolution (see e.g. Conroy & Wechsler 2009; Hopkins et al. 2009b; Cattaneo et al. 2011; Zavala et al. 2012; Shankar et al. 2014; Rodríguez-Puebla et al. 2017; Moster, Naab & White 2018; Behroozi et al. 2019; Grylls et al. 2019; Drakos et al. 2022). By design, SEMs avoid the modelling of galaxy growth and assembly within DM haloes from first principles, unlike more traditional modelling approaches. In their simplest form, SEMs adopt abundance matching techniques (e.g. Kravtsov et al. 2004; Vale & Ostriker 2004; Yang et al. 2004; Shankar et al. 2006; Behroozi, Conroy & Wechsler 2010; Moster et al. 2010), based on the matching between the cumulative number densities of the measured stellar mass functions (SMF) and the host DM halo mass functions (HMF), to generate a monotonic stellar mass–halo mass (SMHM) relation through which they statistically assign galaxies to host DM haloes at different redshifts. Starting from this mapping, SEMs can then focus on specific questions, such as the merger rates of galaxies implied by a specific SMHM relation, or the role played by mergers in forming bulges in galaxies (e.g. Behroozi et al. 2010; Moster et al. 2010; Hopkins et al. 2010b; Moster, Naab & White 2013; Grylls et al. 2019). SEMs are based on minimal input assumptions and associated parameters, allowing for a high degree of transparency in the results whilst avoiding degeneracies.

Additional assumptions can be gradually included in the modelling, e.g. varying the major merger mass ratio threshold for forming ellipticals (as further discussed below), but always allowing for extreme flexibility and transparency.

In the STatistical sEmi-Empirical modeL (STEEL; Grylls et al. 2019, 2020a, hereafter referred to as G19 and G20, respectively) showed how a mean SMHM relation can convert mean DM halo assembly histories into galaxy merger histories, characterized by a total mean accretion track and cumulative mass accreted by merging satellites. From these quantities, galaxy star formation histories can then be computed subtracting from the total galaxy growth the contribution via mergers. The resulting star formation histories can then be compared with independent observational data. The number of surviving satellites can also be compared with relevant data at different redshifts and host halo masses. By including an empirically motivated linear relation between galaxy size and host halo size (Kravtsov 2013; Stringer et al. 2014; Zanisi et al. 2020, 2021a, b) were able to reproduce the strong size evolution of massive galaxies and their size functions up to redshift $z = 0$. Marsden et al. (2021) (see also Ricarte & Natarajan 2018) have then extended these SEMs by including empirical estimates of the evolution of the stellar mass profile of galaxies (e.g. Shankar et al. 2018 and references therein) to predict the full velocity dispersion profiles of central galaxies via detailed Jeans modelling. SEMs are thus a powerful tool to explore mean trends in the assembly, structural, dynamical, and star formation histories of galaxies. Grylls, Shankar & Conselice (2020b) have, however, recently highlighted (see also O'Leary et al. 2021) that even relatively moderate differences in the 'mapping' between galaxy stellar mass and host halo mass, i.e. in the input SMHM relations, can generate significantly distinct galaxy pairs and ultimately galaxy merger rates (along with their associated star formation histories). In a SEM framework, differences in the SMHM relation are mostly induced by systematics in the input galaxy SMFs (e.g. G20), but the loophole identified by G20 can in fact be extended to all theoretical models predicting different SMFs and thus SMHM relations. This strong dependence of the merger rates on the underlying SMHM relation severely limits the effectiveness of the comparison between data on merger rates and hierarchical models developed largely independently of the fitted data used to measure the merger rates (or pair fractions). For example, the answer to the (still open) question whether galaxy major mergers with a mass ratio above, say, 1/4, can generate the right abundances of ellipticals at different epochs (see e.g. Hopkins et al. 2009b, 2010b; Shankar et al. 2013; Grylls et al. 2020a), will strongly depend on which SMHM relation is employed in the hierarchical model at hand (either semi-empirical or not), as we will further prove in this work.

The aim of the present work is twofold. (1) We first present our new Discrete statistical sEmi-empiriCal mODEl (DECODE) specifically designed to efficiently and rapidly predict the merger histories, star formation histories, and satellite abundances of galaxies of any stellar mass at any redshift $z < 3$, for a given set of input SMF. DECODE, as detailed in Section 3, further improves on its predecessor STEEL by replacing *statistical* distributions with catalogues of *distinct* objects, similarly to an *N*-body simulation, and by a more accurate treatment of the subhaloes. Nevertheless, it still retains the flexibility and higher computational performance of STEEL, severely reducing (and in some cases completely eliminating) the limitations imposed by resolution problems in mass and volume, which can heavily impact the modelling of galaxies in a full cosmological setting (see e.g. discussion in van den Bosch et al. 2014). (2) We then use DECODE to study how different renditions of the measured SMF at different epochs impacts the number of galaxy mergers, the formation of ellipticals, and the mean bulge fraction of galaxies in the local Universe. We will show that major mergers may be sufficient to

account for all local ellipticals, but additional processes such as disc instabilities and disc regrowth mechanisms must be invoked to simultaneously explain the mean bulge-to-total distributions of local galaxies.

This paper is organized as follows. In Section 2, we describe the data set we use in our work. In Section 3, we introduce our model DECODE, provide an overview on its numerical implementation, and test the performance of our model against available observational data sets, SAMs, and hydrodynamic simulations. In Section 4, we present our model's predictions for the satellite abundances, merger histories, morphology, and bulge-to-total (B/T) ratios of central galaxies, as well as the mean growth history of brightest cluster galaxies (BCGs). Finally, in Sections 5 and 6 we discuss our results and draw our conclusions. In this paper, we adopt the ΛCDM cosmological model with parameters from Planck Collaboration VI (2018) best-fitting values, i.e. $(\Omega_{\rm m}, \Omega_\Lambda, \Omega_{\rm b}, h, n_{\rm S}, \sigma_8) = (0.31, 0.69, 0.049, 0.68, 0.97, 0.81)$, and use throughout a Chabrier (Chabrier 2003) stellar initial mass function.

2 DATA

In this work, we make heavy use of both numerical/theoretical and observational data sets. We use the former mostly for validation tests of DECODE, while the latter are used both as inputs for DECODE, as well as outputs to test DECODE's predictions. More specifically, we make use of: (1) the Millennium simulation to test the accuracy of the abundances in the surviving/unmerged subhaloes in DECODE; (2) the TNG100 simulation, to compare the performance of DECODE to a hydrodynamic simulation in terms of galaxy properties (in this work mostly fraction of ellipticals and B/T mass ratios); (3) the GALICS SAM, to compare how DECODE compares to a full ab initio analytical model of galaxy formation; (4) *SDSS* and MaNGA to compare DECODE's predictions on satellite abundances, fraction of ellipticals, and B/T ratios with large data sets of local galaxies. Below we provide relevant details on each of these comparison data sets. In Appendix C, we discuss how DECODE can faithfully reproduce the galaxy assembly histories of other cosmological models when it receives in input their mean SMHM relations. We will also compare with another cosmological SEM (EMERGE; see Moster et al. 2018).

2.1 The Millennium simulation

We use the Millennium DM-only simulation (Springel et al. 2005) scaled to the Planck cosmology (Planck Collaboration XIII 2016) employing the method of Angulo & White (2010) and Angulo & Hilbert (2015). All DM haloes are identified using a Friends-Of-Friends (FOF) algorithm (Davis et al. 1985). Furthermore, all DM subhaloes are detected using the SUBFIND algorithm (Springel et al. 2001), based on which each FOF halo has one central subhalo, and the rest of the subhaloes are labelled as satellite subhaloes. The SUBFIND algorithm considers a minimum number of particles $n_{\rm min} = 20$ for identifying subhaloes. We consider the infall time of each satellite subhalo as the time when it last changes its type from central to satellite. This information is taken from the publicly available L-Galaxies semi-analytical model (Ayromlou et al. 2021)[1] that runs on top of the Millennium simulation.

We note that although the halo virial mass M_{200} is reported as the mass within the halo virial radius R_{200}, the FOF halo could extend beyond this scale. Therefore, satellite subhaloes of an FOF halo can exist beyond the R_{200} as well. Nevertheless, this does not constitute a limitation in the comparison of the subhalo mass function (SHMF) with DECODE since we base our subhaloes generation on the unevolved total SHMF from Jiang & van den Bosch (2014, 2016), which has already been shown to be in good agreement with the one from the Millennium simulation as presented in Li & Mo (2009). Furthermore, recently Green, van den Bosch & Jiang (2021) presented a more accurate SHMF computed with an updated version of their model SatGen, where they are able to follow subhalo orbits and effects of numerical disruption. The SHMF that we use slightly differs with the one from Green et al. (2021) by a factor of ∼0.1 dex only at $M_{\rm h,sub}/M_{\rm h,par} \lesssim 10^{-3}$, where their contribution to the galaxy mergers is completely negligible. Nevertheless, we checked that the unevolved total subhalo distribution given by the updated Green et al. (2021) version of SatGen is statistically unchanged with respect to that presented in Jiang & van den Bosch (2016). This is also guaranteed by the fact that the unevolved SHMF is a manifestation of the progenitor mass function used in the extended-Press–Schechter formalism, which depends only on the cosmological parameters.

2.2 The TNG simulation

We make use of the public data release[2] from the TNG100 hydrodynamical simulation of the IllustrisTNG project (Nelson et al. 2019). The IllustrisTNG simulation (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018) is a set of cosmological hydrodynamical simulations performed using the code AREPO (Springel 2010). Employing subgrid physics, TNG implements astrophysical processes relevant to galaxy evolution, such as the cooling of the hot gas, star formation, the evolution of stars, supernova feedback, supermassive black hole formation (seeding), and AGN feedback (see Pillepich et al. 2018a; Weinberger et al. 2017, for full model description). The TNG model has been presented in three different cosmological boxes so far ($l_{\rm box} \sim 50, 100, 300$ Mpc). Here, we take the 100-Mpc box (TNG100) for our analysis because this is the simulation box used to calibrate the TNG model against observations. Therefore, TNG100 outputs the most reliable results among the other TNG simulations.

The galaxy morphologies are calculated as in Genel et al. (2015) and Marinacci, Pakmor & Springel (2014). The kinematic decomposition of galaxies is based on the distribution of the circular parameter of individual stellar particles, which is defined as $\epsilon = J_z/J(E)$ (Du et al. 2019). Here, J_z corresponds to the specific angular momentum in the symmetric axis of the galaxy and $J(E)$ the maximum specific angular momentum possible at the specific binding energy (E) of the stellar particle. The bulge component of each galaxy comprises the stellar particles with the $\epsilon < 0$ and a fraction of stellar particles with a positive circular parameter that mirrors around zero. The mass of the bulge is twice the mass of the stellar particles, with a circular parameter $\epsilon < 0$. The elliptical galaxies are defined as the galaxies with bulge-to-total stellar mass ratio $B/T > 0.7$.

2.3 GALICS

We make use of the data computed via GALICS 2.2 (Koutsouridou & Cattaneo 2022). GALICS 2.2 is the latest version of the GALICS (Galaxies In Cosmological Simulations) SAM of galaxy formation (Hatton et al. 2003; Cattaneo et al. 2006, 2017, 2020). The main differences between the GALICS 2.2 version used for this article and

[1] https://lgalaxiespublicrelease.github.io

[2] https://www.tng-project.org/data

the latest published version GALICS 2.1 (Cattaneo et al. 2020) are the presence of feedback from active galactic nuclei (AGNs) in GALICS 2.2 (GALICS 2.1 did not include AGN feedback) and the model for morphological transformations in galaxy mergers (in GALICS 2.1, major mergers destroyed the disc component completely; GALICS 2.2 adopts a more realistic model based on the numerical results of Kannan et al. 2015). Since Cattaneo et al. (2020), there have also been small improvements in the modelling of supernova feedback and disc instabilities.

The implementation of the disc instabilities is based on the results of Devergne et al. (2020) who studied the growth of pseudo-bulges in isolated thin exponential stellar discs embedded in static spherical haloes. They found that discs with $v_{\rm d}/v_{\rm c} > 0.6$ are unstable (see also Efstathiou, Lake & Negroponte 1982), where $v_{\rm c}$ is the circular velocity of the galaxy and $v_{\rm d}$ is the circular velocity considering only the disc's gravity. In what follows, we adopt the following fitting formula (Devergne et al. 2020) for the B/T mass ratio

$$B/T = 0.5 f_{\rm d}^{1.8}(3.2 r_{\rm d}), \quad (1)$$

where $f_{\rm d} = (v_{\rm d}/v_{\rm c})^2$ is the contribution of the disc to the total gravitational acceleration, and $r_{\rm d} = R_{\rm d}/R_{\rm vir}$ is the dimensionless exponential scale length in units of the virial radius $R_{\rm vir}$.

The mass ratio of the mergers is defined as $\mu = M_2/M_1$, where M_1 and M_2 are the total (baryonic and DM) matter within the half-mass radii of the primary and secondary galaxies, respectively. We base our assumptions for the mergers on Kannan et al. (2015) which can be summarized as follows:

(i) a fraction μ of the stars in the disc of the primary galaxy is transferred to the central bulge;
(ii) another fraction 0.2μ of the same stars is scattered into the DM halo, which acquires a stellar component in GALICS;
(iii) a fraction $\mu(1 - f_{\rm gas\text{-}disc})$ of the gas in the disc of the primary galaxy is transferred to the central cusp, where $f_{\rm gas\text{-}disc}$ is the gas fraction on the primary galaxy's disc;
(iv) even if the gas that remains in the disc undergoes a starburst in the case of major merger $\mu > 0.25$, we assume that this gas has the same star formation time-scale as that in the central cusp (that is $t_{\rm SF} = M_{\rm gas}/{\rm SFR} = 0.2\,{\rm Gyr}$), see also Powell et al. (2013) who showed that most major mergers exhibit extended SF in their early stages;
(v) the stars and the gas in the bar of the primary galaxy are transferred to the bulge and cusp of the merger remnant, respectively (Bournaud & Combes 2002; Berentzen et al. 2007);
(vi) a fraction μ of all the stars of the secondary galaxy ends up in the bulge of the merger remnant, while the rest is added to the disc;
(vii) all the gas of the secondary galaxy ends up in the central cusp.

2.4 *Sloan Digital Sky Survey* and MaNGA

Our reference data from *SDSS* is the Data Release 7 (DR7; Abazajian et al. 2009) as presented in Meert, Vikram & Bernardi (2015, 2016), which has a median redshift of $z \sim 0.1$. Stellar masses are computed using the best-fitting `Sérsic-Exponential` or `Sérsic` photometry of *r*-band observations, and by adopting the mass-to-light ratios by Mendel et al. (2014). Furthermore, we adopt the truncation of the light profile as prescribed in Fischer, Bernardi & Meert (2017). The Meert et al. catalogues are matched with the Yang et al. (2007, 2012) group catalogues, which allow us to identify central and satellite galaxies and provide an estimate of the group halo mass. Neural-network based morphologies from Domínguez Sánchez et al. (2018) are adopted in the following.

The first observable against which we test our model is the stellar mass function of satellites, which is computed using standard $V_{\rm max}$ weighting. We further test the model against the fraction of elliptical galaxies, $f_{\rm ellipticals}$ as a function of stellar mass. The Domínguez Sánchez et al. (2018) catalogue provides estimates of T-Types, as well as the probability for early type Galaxies (i.e. T Type ≤ 0) of being S0, $P_{\rm S0}$. We compute $f_{\rm ellipticals}$ by considering only early type Galaxies for which $P_{\rm S0}$ falls below a certain threshold. We have accurately tested that the dependence of the ellipticals fraction on such threshold is extremely mild, showing variation within 10 per cent even for $P_{\rm S0} \lesssim 0.3$. In this work, we adopt $P_{\rm S0} < 0.5$ according to the results from Domínguez Sánchez et al. (2018). We estimate the error bars on the satellite stellar mass functions and on $f_{\rm ellipticals}$ using the Poisson statistics.

Later in this work we will be interested in bulge-to-total ratios (B/T) of galaxies. Determining B/T values from observed surface brightness profiles is slightly involved, since not all objects are well fit by two-component (SerExp) profiles. While a careful analysis of B/T values for all *SDSS* galaxies is not yet available, reliable values of B/T have recently been provided by Bernardi et al. (in preparation) for the objects in the MaNGA survey (Bundy et al. 2015; Drory et al. 2015; Law et al. 2015). MaNGA is a component of the *Sloan Digital Sky Survey* IV (Gunn et al. 2006; Smee et al. 2013; Blanton et al. 2017; hereafter *SDSS* IV) and uses integral field units (IFUs) to measure spectra across nearby galaxies. The MaNGA final data release (DR17; Abdurro'uf et al. 2021) includes observations of about 10 000 galaxies: the `DR17 MaNGA Morphology Deep Learning` Value Added catalogue (DR17-MMDL-VAC) provides morphological classifications and the `DR17 MaNGA PyMorph photometric` Value Added Catalogue (DR17-MPP-VAC) provides Sérsic (Ser) and Sérsic + Exponential (SerExp) fits to the 2D surface brightness profiles of these objects, along with a detailed flagging system for using the fits (see Domínguez Sánchez et al. 2022 for details). Bernardi et al. (in preparation) describe how to combine the photometric and flagging information to determine reliable B/T values for these objects, and how the B/T correlate with morphology. Because the MaNGA selection function is complicated, and because the MaNGA sample is much smaller than the *SDSS*, we use the *SDSS* to determine the shape of the stellar mass function and how $f_{\rm ellipticals}$ varies with stellar mass, but MaNGA to determine how B/T correlates with stellar mass.

3 THE DECODE IMPLEMENTATION

In this Section, we present our state-of-the-art discrete semi-empirical model, DECODE. DECODE is designed as a flexible, fast, and accurate tool to predict the average merger and star formation histories of central galaxies at any given epoch without the need of resorting to a full SAM, hydrodynamical simulation, or even a complex, multiparameter cosmological SEM. In this work, we will mostly focus on the mean mass assembly and merger histories of central galaxies, along with their satellite abundances. We will leave the study of star formation histories to a separate study.

The main steps of DECODE can be summarized as follows:

(i) generation of the central DM halo population (Section 3.1);
(ii) generation of the DM subhalo population (Section 3.2);
(iii) evolution of subhaloes after infall (Sections 3.3 and 3.4);
(iv) populating haloes with galaxies (Section 3.5).

Fig. 1 depicts our general framework to generate the population of DM haloes and subhaloes which we further detail in the sections below. In brief, our methodology relies on generating large catalogues of parent haloes extracted from the HMF and endowed with a mean halo growth history as derived from *N*-body numerical simulations and analytical models (e.g. van den Bosch, Tormen & Giocoli 2005). Subhaloes are subsequently extracted from the unevolved SHMFs, also based on accurate studies performed on *N*-body simulations (e.g. Jiang & van den Bosch 2016). In its implementation of the evolutionary tracks for central galaxies, DECODE is similar in concept to its predecessor statistical semi-empirical model STEEL (G19), although it crucially differs from it in its implementation, beyond the performance itself. DECODE in fact avoids continuous statistical weights of central and satellite galaxies, but it directly works on discrete objects, similarly to an *N*-body simulation, making it easier to handle object-by-object variations, but still preserving STEEL's extreme flexibility and broad independence on volume and/or mass resolution limitations.

In addition, as detailed below, DECODE distinguishes between satellites and satellites of satellites, a key feature that was absent in STEEL. DECODE's main objective is to still predict *mean* galaxy growth histories, as in G19's statistical model STEEL, but instead of using statistical weights, it relies on the generation of stochastic samples of haloes and galaxies that, on average, grow in mass as predicted by STEEL, as further detailed and demonstrated in Appendix A. By working with discrete sources, DECODE avoids the need to assign weights to each evolutionary step, which is a far from trivial task when propagated through different layers of complexities in the galaxy modelling. We also note that, although we adopt a Planck cosmology throughout, as specified in Section 1, our results are insensitive to the exact choice of input cosmological parameters within reasonable ranges. This independence on cosmological parameters is mostly induced by the heavy use of the SHMF, which has been shown several times in the literature to be of very similar shape in different simulations (e.g. Jiang & van den Bosch 2016; Green et al. 2021). We will further reiterate on this point in Section 3.6.

3.1 Generating the population of parent haloes

Our first step consists in generating a large catalogue of parent DM haloes extracted from the HMF. In this work, we adopt the definition and parametrization of the HMF according to Tinker et al. (2008), which accounts only for central haloes.[3] We extract haloes and their masses in our catalogue from the cumulative HMF multiplied by an input cosmological volume. We choose here to use a reference box of 250 Mpc on a side, which allows throughout to balance speed with accuracy, although we stress that DECODE is flexible enough to generate even larger volumes with ease. This method is extremely rapid and allows to simulate even large samples of massive cluster-sized haloes.

The average mass accretion history of each central halo is then computed using the methodology described in van den Bosch et al. (2014). Instead of computing the growth of each single halo, we predefine a fine halo grid in mass of 0.1 dex width and assign the same mean history to all the DM haloes contributing to the same cell in the grid. This initial step provides the average growth history of the 'main progenitor' (when compared to a traditional merger tree).

[3]We make use of the `lss.mass_function` module in the PYTHON `colossus` package (Diemer 2018).

3.2 Generating the population of subhaloes of different orders

The average number of subhaloes of a given mass that falls on to the parent halo at any given time is given by the SHMF. Here, we adopt the definition of unevolved SHMF distinguished by 'order' of accretion in the merger tree (Jiang & van den Bosch 2014, 2016), with first-order subhaloes being the ones falling directly on to the main branch, second-order subhaloes the ones already satellites in first-order subhaloes at the time of accretion on to the main branch, and so on (panel B in Fig. 1).

The same methodology applied to parent haloes is used to calculate the number and mass of all the subhaloes that have ever fallen on to each given parent halo, by making use of the cumulative total unevolved SHMF. Once the number and mass of the subhaloes are known, the order of the subhaloes are assigned by considering the SHMF distinguished by order, for which we use the recipe from Jiang & van den Bosch (2016) for the first-order SHMF and equation (17) of Jiang & van den Bosch (2014) for higher orders. The probability that a given subhalo of a given mass is of first or higher order, is then simply given by the relative ratio between first/higher order SHMF and the total SHMF computed in the chosen bin of (sub)halo mass.

We stress that DECODE is a statistical SEM, where the mock catalogues of DM haloes are stochastic realizations of the input HMF and SHMF, taken from analytical fits to *N*-body simulation. The only free parameter in our model is the scatter in stellar mass at given halo mass, which is included in the abundance matching relation in equation (8), and it only contributes to the shape of the mean SMHM relation, as we will further detail in Section 3.5.

3.3 Infall redshifts

To predict a robust merger history of central galaxies, clear knowledge of infall redshifts ($z_{\rm inf}$) of their satellite galaxies is required. In our model, we adopt the definition of infall redshift as the time when a DM halo was accreted for the *first* time as subhalo, or in other words when it entered the virial radius of another halo. We assign infall redshifts to the subhaloes in a statistical way distinguishing between first order and higher order subhaloes. For first order subhaloes, we apply the redshift probability distribution dictated directly by the SHMF, as follows. In practice, at any given redshift z, for a parent halo growing by $\mathrm{d}M_{\rm h,par}(z)$ at any redshift interval $z + \mathrm{d}z$ to z, the probability density function (PDF) of infall redshifts of subhaloes of infall mass $M_{\rm h,sub}$ is given by the derivative the SHMF with respect to the redshift, formally

$$\mathrm{PDF}(z_{\rm inf}) \propto \frac{\mathrm{d}}{\mathrm{d}z}\phi(M_{\rm h,sub}, M_{\rm h,par}(z)) \, . \tag{2}$$

As mentioned above, the unevolved SHMF $\phi(M_{\rm h,\,sub})$ provides the total number and mass of the subhaloes that have ever merged with the parent halo at any epoch. The mass of the parent halo $M_{\rm h,par}(z)$ will grow with cosmic time when moving from $z + \mathrm{d}z$ to z. The change in the associated SHMF $\mathrm{d}\phi(M_{\rm h,sub}, M_{\rm h,par}(z))$, will provide the number and mass of the subhaloes of mass within $M_{\rm h,sub}$ and $M_{\rm h,sub} + \mathrm{d}M_{\rm h,sub}$, that have merged with the parent halo and contributed to its mass growth in the redshift interval $\mathrm{d}z$. Thus, the (normalized) derivative with redshift of the SHMF at a fixed subhalo mass $M_{\rm h,sub}$ given in equation (2), will provide the PDF for the subhaloes of mass within $M_{\rm h,sub}$ and $M_{\rm h,sub} + \mathrm{d}M_{\rm h,sub}$, merging with the parent halo $M_{\rm h,par}(z)$ at any given redshift z. Redshifts of infall for subhaloes of any given mass $M_{\rm h,sub}$ are then generated by randomly extracting them from the PDF given in equation (2).

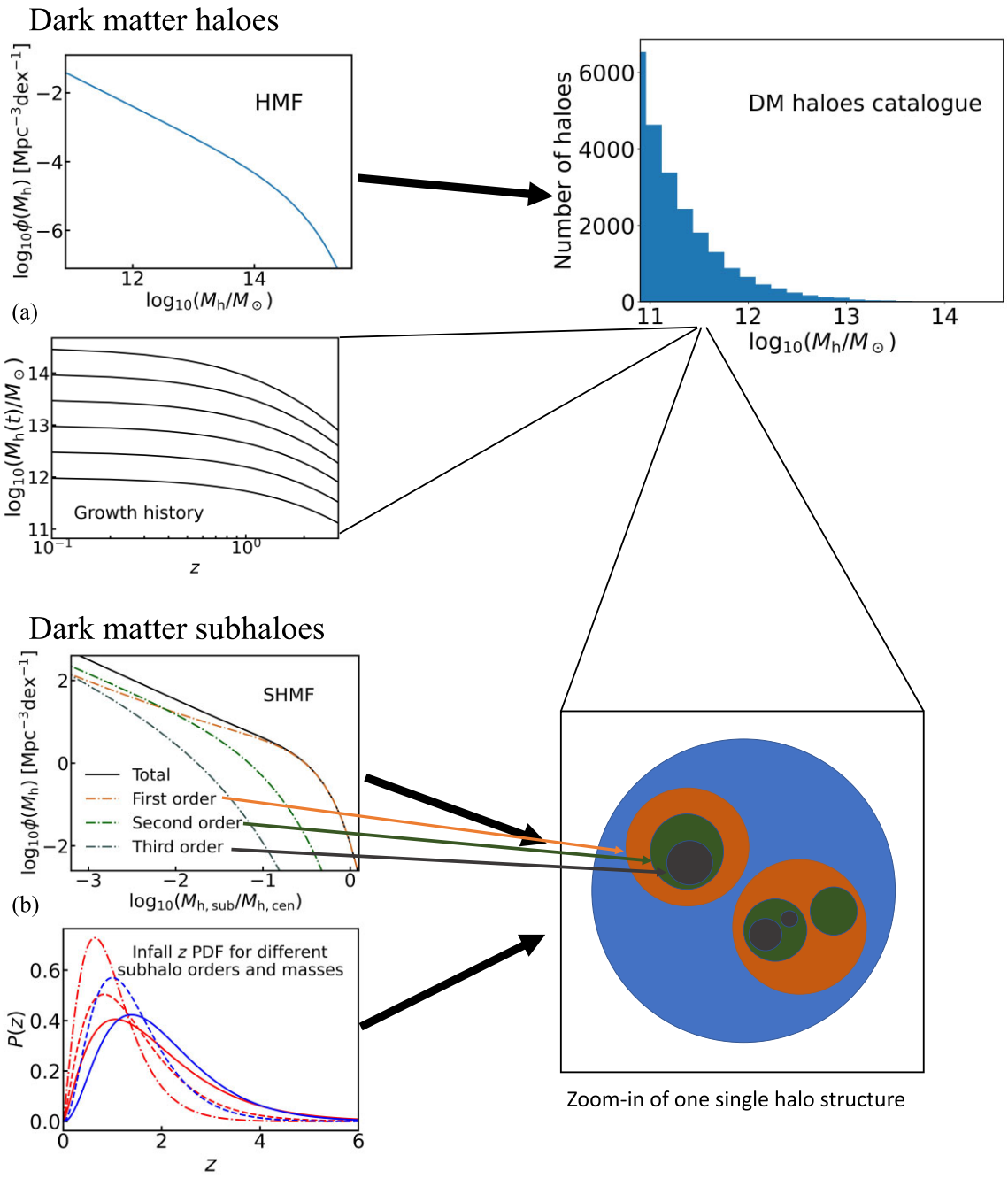

Figure 1. Scheme of the backbone of DECODE for the dark matter side as described in Section 3. Panel A: functions used to generate the parent haloes catalogue (represented by the histogram in mass bins). The HMF is used as probability density distribution to generate the masses of the dark matter haloes, and a mean accretion history is assigned to each of them through an analytical fit. This contributes to build a set of main progenitors discretely, each of them characterized by a mean accretion history. The histogram in the top right panel represents a stochastic realization of the HMF. Panel B: statistical functions used to create the dark matter subhaloes. For each parent halo, we compute the SHMF for all subhaloes as well as for each order, and use it as probability density distribution for generating the subhalo population. The order of the subhaloes in the merger tree is assigned using the SHMFs distinguished by order (coloured dashed–dotted lines), and in this work we limit our attention up to the second order. Finally, the redshift of infall is assigned to subhaloes via fitted analytical equations, depending on their order and mass (see left-hand panel of Fig. 3 for the distinction for different orders and masses). In this way, the merging structure of each halo is known, i.e. the infalling subhaloes' order, mass and time at infall.

The methodology described so far to assign redshifts of infall to subhaloes can only be applied to first-order subhaloes, as the SHMF only provides information on the subhaloes merging with the parent halo. For second-order subhaloes, we adopt a similar, but not identical, recipe. We first generate the full merger tree associated to a given parent halo P0 by following its mass accretion history backwards in time and, using the recipe described above, computing the population of first-order subhaloes S1 and their redshifts of infall from the first-order SHMF (Jiang & van den Bosch 2014, 2016). For each S1 we then follow its mass and satellite accretion history using

again the first-order SHMF. The satellites of S1 will be second order, S2, with respect to P0. We repeat this loop up to the third order[4] as orders higher than the third have an insignificant contribution to the total number density of satellites (Jiang & van den Bosch 2014). In order to speed up the computational time, we first run fine merger histories for all relevant subhaloes S2 and S3, and then compute analytic fits to the PDF(z) of their redshifts of infall. The parametrization for the infall redshift distribution we adopt is given by the following formula:

$$\mathcal{P}(z) = Az^{\alpha}\frac{1}{\delta e^{\beta z} - \gamma}\,, \tag{3}$$

where A, α, β, γ, and δ are dimensionless free fitting parameters, with best-fitting values reported in Section 3.6. We use the analytical PDF of equation (3) to assign the infall redshifts statistically[5] to all second- and third-order subhaloes, especially when simulating large boxes and cluster-sized parent haloes.

The procedure described above generates a stochastic merger tree of subhaloes for a given *mean* parent halo mass accretion track $M_{\rm h,par}(z)$. In other words, DECODE produces a stochastic distribution of subhaloes merging on a mean halo. As quantitatively proven and discussed in Appendix A, when averaged over a large population of subhaloes, this approach is equivalent to an average one in which discrete subhaloes are replaced by statistical weights given by the SHMF, as carried out in STEEL. We stress that the main advantage of building halo assembly histories via discrete sources resides in the extreme flexibility of working with discrete objects and not with statistical weights, especially when transitioning to galaxies and the modelling of their evolutionary properties.

3.4 Merging time-scales and surviving subhaloes

Once a subhalo first falls into its host halo, it is affected by tidal stripping and dynamical friction, resulting in an overall net mass-loss. Many works have carefully studied via numerical simulations these processes (see e.g. van den Bosch et al. 2005; Giocoli, Tormen & van den Bosch 2008; Angulo et al. 2009; Jiang & van den Bosch 2016; van den Bosch et al. 2016, 2017; Green & van den Bosch 2019), and have found that the average mass-loss rate of satellite subhaloes can be analytically expressed as

$$\dot{M}_{\rm h,sub} = -\mathcal{A}\frac{M_{\rm h,sub}}{\tau_{\rm dyn}}\left(\frac{M_{\rm h,sub}}{M_{\rm h,host}}\right)^{\zeta}, \tag{4}$$

where $\mathcal{A} = 1.54$, $\zeta = 0.07$, and $M_{\rm h,sub}$ and $M_{\rm h,host}$ are, respectively, the masses of the subhalo and halo that hosts the former subhalo. $\tau_{\rm dyn}$ is the halo dynamical time-scale given by

$$\tau_{\rm dyn}(z) = 1.628h^{-1}{\rm Gyr}\left[\frac{\Delta_{\rm vir}(z)}{178}\right]^{-1/2}\left[\frac{H(z)}{H_0}\right]^{-1}, \tag{5}$$

with $H(z)$ being the Hubble's parameter at redshift z and $\Delta_{\rm vir}$ the virial parameter taken from equation (6) of Bryan & Norman (1998). The typical time-scale that a subhalo needs in order to fully merge with its progenitor from the time of first accretion is well described by the merging time-scale formula given by equation (5) of Boylan-Kolchin, Ma & Quataert (2008), which we report below

$$\tau_{\rm merge} = \tau_{\rm dyn}A\frac{(M_{\rm h,host}/M_{\rm h,sub})^{b}}{\ln(1+M_{\rm h,host}/M_{\rm h,sub})}\exp\left[c\frac{J}{J_{\rm c}(E)}\right]\left[\frac{r_{\rm c}(E)}{R_{\rm vir}}\right]^{d}, \tag{6}$$

where $J/J_c(E)$ the orbital energy and (A, b, c, d) are free parameters that govern the dependence of the merging time-scale on the mass ratio. Here, we adopt the fitting parameters provided by McCavana et al. (2012). In particular, in order to apply equation (6), we assign an orbital circularity ξ to galaxies according to Khochfar & Burkert (2006), by extracting a random value from a Gaussian distribution centred in $\bar{\xi} = 0.5$ and with standard deviation $\sigma_{\xi} = 0.23$, and compute the ratio between the average radius of the orbit $r_{\rm c}$ and the host halo virial radius $R_{\rm vir}$

$$\frac{r_{\rm c}}{R_{\rm vir}} = \frac{\xi^{2.17}}{1-\sqrt{1-\xi^2}}. \tag{7}$$

The analytic recipes described above are an approximation to the complex dynamics of DM subhaloes, and also the numerical simulations from which they are extracted may themselves suffer from resolution and/or incompleteness effects. To allow for some flexibility in the merging time-scales, following G19, we also include a fudge factor $f_{\rm dyn}$ in equation (6), $\tau_{\rm merge} \rightarrow f_{\rm dyn}\tau_{\rm merge}$, which we assume to be slightly dependent on parent mass, as detailed in Section 3.6.

To test the validity of the methodology described in the previous paragraph, we analyse the population of the surviving subhaloes at present day via the unevolved surviving SHMF. For the latter we adopt the definition of Jiang & van den Bosch (2016), which is the number density of the surviving subhaloes still present today as a function of their mass at the time of first accretion. We again assume that subhaloes of third order or higher are not statistically significant to the surviving population (Jiang & van den Bosch 2014).

The next step is to assign merging time-scales to all subhaloes of different ranking, which we implement in DECODE in the following way. For any parent halo of mass $M_{\rm P0}$, with a first-order subhalo of mass $M_{\rm S1}$ and a second-order subhalo of mass $M_{\rm S2}$:

(i) we first calculate the merging time-scale of the first-order subhalo which depends on the ratio $M_{\rm P0}/M_{\rm S1}$;

(ii) depending on the first accretion epoch and the merging time-scale, we consider and implement in DECODE the three following possibilities: (1) the first-order subhalo has survived today and we assume at this step that its higher order subhaloes inside still exist; (2) the first-order subhalo has not survived and it releases all its higher order subhaloes to the parent[6] (Jiang & van den Bosch 2016); (3) the higher order subhaloes have been tidally disrupted before their first-order subhalo has merged;

(iii) we assign the merging time-scale to the second-order subhalo which depends on the ratio $M_{\rm S1}/M_{\rm S2}$;

[4]We explore and provide results for the distribution of infall redshifts up to the third order for completeness. However, for the purposes of this paper discussed in Section 4 we limit our investigation to the second-order subhaloes. We have tested that orders higher than the second have negligible contribution to the amount of mergers.

[5]In reality, for each single parent halo, a generic subhalo of the ith order must have fallen at a redshift higher than the infall redshift of the $(i-1)$th subhalo. We therefore set the infall redshift of the $(i-1)$th order subhalo as a lower bound for the ith order subhalo redshift.

[6]We investigate also different ways of treating the evolution of subhaloes after the time of infall. In particular, we explored two additional possibilities: (1) every higher order subhaloes merge together with their host first-order subhalo, (2) higher order subhaloes have a dynamical friction longer than the age of the Universe and never merge. In both cases, there is not any appreciable difference in terms of satellite abundances and mergers.

(iv) finally, to the second-order subhaloes that are released from a first-order to the parent we assign a new merging time-scale using the ratio $M_{\rm P0}/M_{\rm S2}$,[7] when released to the parent halo P0.

In Section 3.6, we will compare DECODE's predicted abundances of local unmerged subhaloes, described in terms of the surviving SHMF, with the analytical model of Jiang & van den Bosch (2016) and with the resolved SHMF of the Millennium simulation, showing very good agreement when adopting a fudge factor of $f_{\rm dyn} \sim 0.64$. We stress already here that simply adopting $f_{\rm dyn} \sim 1$ does not alter any of our main results.

3.5 Building the mapping between galaxy stellar mass and host dark matter halo mass

One of the key components of our modelling is the relation between stellar mass and host halo mass, the SMHM relation. The latter is computed via the formalism put forward in Aversa et al. (2015, see equation 37 therein), which allows to calculate the mean stellar mass at given halo mass

$$\int_{\log M_*}^{+\infty} \phi(M'_*, z) {\rm d}\log M'_* = \int_{-\infty}^{+\infty} \frac{1}{2}{\rm erfc}\left\{\frac{\log M_{\rm h}(M_*) - \log M'_{\rm h}}{\sqrt{2}\sim\sigma_{\log M_*}}\right\} \times \cdot\, \phi(M'_{\rm h}, z){\rm d}\log M'_{\rm h}, \quad (8)$$

where $\tilde{\sigma}_{\log M_*} = \sigma_{\log M_*}/\mu$, with $\sigma_{\log M_*}$ being the Gaussian scatter at fixed halo mass and $\mu = {\rm dlog}\, M_*/{\rm dlog}\, M_{\rm h}$ the derivative of the SMHM relation. Equation (8)[8] provides a fast and flexible methodology to compute the SMHM relation numerically, without the need for a pre-defined analytical fit, but requiring in input only one parameter, the scatter in stellar mass at fixed halo mass (see also Kravtsov, Vikhlinin & Meshcheryakov 2018). When applying equation (8), the other two main ingredients are the observational SMF and the HMF including the subhalo term. We make use of the total HMF which accounts for both parent haloes and subhaloes $\phi(M_{\rm h,tot}) = \phi(M_{\rm h,par}) + \phi(M_{\rm h,sub})$, where $\phi(M_{\rm h,sub}) = k \cdot \phi(M_{\rm h,par})$ with k being a correction factor (as described in Appendix B). The additional term $\phi(M_{\rm h,sub})$ allows to include all unstripped subhaloes and unmerged up to redshift z, as predicted by DECODE following the recipes detailed in Section 3.4. We find that the new total HMF $\phi(M_{\rm h,tot})$, inclusive of the surviving satellites, is similar to the parent HMF $\phi(M_{\rm h,\,par})$ but, as expected, with a steeper low mass end. A similar approach was adopted by Behroozi, Wechsler & Conroy (2013) who, in their appendix G provide a redshift-dependent analytical formula to correct the HMF for the abundances of surviving satellites. We adopt their analytical formula which we fit to reproduce our realizations of halo + subhalo mass functions with DECODE. Our best-fitting parameters are given in Appendix B. The SMHM relation generated at each redshift via equation (8) is then given as input in DECODE to assign galaxies to all parent haloes and to all subhaloes of any rank at the time of infall. We also test our SMHM relations computed via equation (8) with the SMHM relations computed using the halo peak velocity function and the same SMF as input (see Appendix E).

Throughout we assume that satellites at infall follow the same SMHM relation of centrals at that epoch. We note that if we were to relax this assumption by allowing for a somewhat different SMHM relation for satellites, our main results would be unaltered in the stellar mass range of interest here $M_* \gtrsim 10^{10}\,{\rm M}_\odot$, in line with the findings of other SEMs that suggest similar distributions of centrals and satellites at the high-mass end (see e.g. Rodríguez-Puebla, Drory & Avila-Reese 2012; Dvornik et al. 2020; Contreras, Angulo & Zennaro 2021; Engler et al. 2021). Furthermore, we consider only 'frozen' models in this work (i.e. the mass of the satellite is assumed to be constant after the infall). As also shown by G19, allowing for some star formation and stellar stripping in the satellites after infall, following standard recipes in the literature, does not alter any of our conclusions on the abundances of satellites, at least for galaxies of stellar mass above $10^{10}\,{\rm M}_\odot$. We will study the full impact of stellar stripping and latent star formation in satellites in a separate work. We point out that galaxies are assigned to DM haloes via the *mean* SMHM relation, derived via equation (8), because DECODE at this level of development is mostly sensitive to the mean galaxy growth and mean merger histories. Therefore, in this paper we show only the mean predictions for the satellite abundances, ellipticals, and B/T ratios.

3.6 Validating the dark sector

Before presenting the power and flexibility of DECODE in efficiently probing crucial aspects of galaxy evolution, such as merger pairs and bulge formation, we test the accuracy of DECODE in matching the number densities and $z_{\rm inf}$ distributions of unevolved and unmerged subhaloes of first and second order as predicted by N-body simulations and SAMs. We first note that the unevolved total SHMF fitted by Jiang & van den Bosch (2016) from the MultiDark simulation (Klypin, Trujillo-Gomez & Primack 2011), is well consistent, we verified, with the total unevolved SHMF extracted from the Millennium simulation, at least above the resolution limit of the latter. This is quite significant as it further proves the universality of the SHMF with respect to the underlying cosmological model and also other aspects of the simulations, such as the halo finder algorithm.

The left-hand panel of Fig. 2 shows the unevolved *surviving* SHMF (i.e. composed by subhaloes not yet merged or completely disrupted) for two different values of parent halo masses, as labelled. The results are compared with those from the Millennium simulation and from the SAM of Jiang & van den Bosch (2016), which reproduces well the results from the Millennium simulation. The predictions from DECODE on the number of surviving, unstripped subhaloes, are plotted with dashed lines, and become indistinguishable from the grey solid line of Jiang & van den Bosch (2016), when a small correction is applied to the merging time-scales of McCavana et al. (2012), as also pointed out by G19. As shown in the right-hand panel of Fig. 2, the fudge factor $f_{\rm dyn}$ in the dynamical friction time-scale, we find, is well represented by the following linear relation with the halo-to-subhalo mass ratio $\psi = (M_{\rm h,host}/M_{\rm h,subhalo})$

$$f_{\rm dyn} = a\psi + b, \quad (9)$$

where the best-fitting values for a and b are 0.000 35 and 0.65, respectively.

In order to be used as a flexible tool to model, e.g. galaxy merger rates, it is essential for DECODE to not only generate the correct abundances of subhaloes of different orders, but also to reproduce the correct probability distributions of their infall redshifts $z_{\rm inf}$. The left-hand panel of Fig. 3 shows the predicted $z_{\rm inf}$ PDFs predicted by DECODE (histograms) for different subhalo masses and order, as labelled. As described in Section 3.3, these PDFs have been calculated by generating full merger trees for each parent halo and subhalo. Such a procedure is of course not practical and

[7] Here, $M_{\rm S2}$ is the evolved mass of the second-order subhalo S2, that following a mass evolution according to equation (4)

[8] We test that equation (8) provides accurate results for $\sigma_{\log M_*} \lesssim 0.3$.

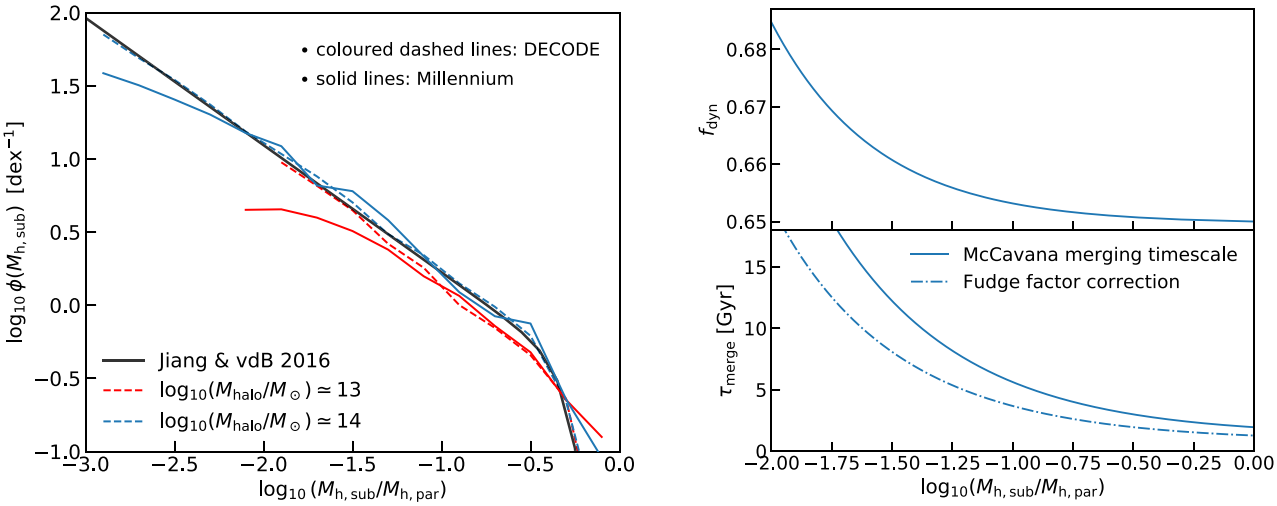

Figure 2. Left-hand panel: Comparison between the surviving unevolved subhalo mass function for two different parent halo masses at redshift $z = 0$. The coloured dashed lines are the results from DECODE, the solid lines are the results extracted from the Millennium simulation (as described in Section 2) and the solid black line the analytical form taken from Jiang & van den Bosch (2016). Right upper panel: fudge factor as function of the mass ratio according to equation (9). Right lower panel: merging time-scale from McCavana et al. (2012) (solid line) compared with that computed by applying the fudge factor correction (dashed–dotted line).

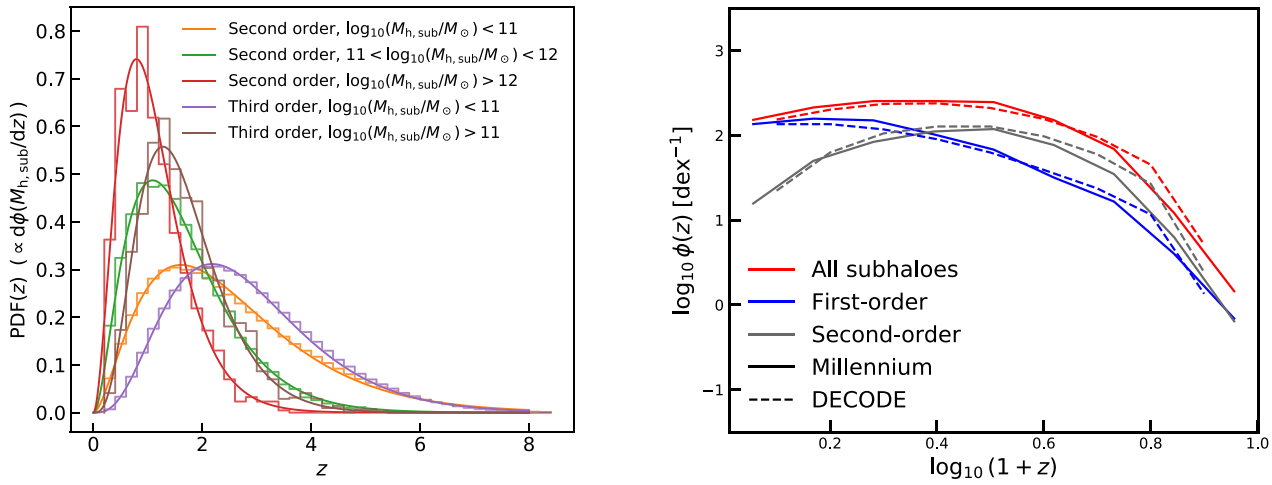

Figure 3. Left-hand panel: analytical (normalized) probability distributions of the infall redshifts adopted in this work to generate the mock catalogues. The results are organized for different subhalo orders and mass ranges. The histograms show the results from the merger tree and the curves show the best fits. Right-hand panel: comparison between number densities of the infall redshifts from our model DECODE (dashed lines) and from Millennium simulation (solid lines) for parent haloes of mass selected between 10^{14} and $10^{14.1}$ $M_\odot$. Results are shown for subhaloes of all orders (red lines), first order (blue lines), and second order (grey lines). Similar results are found for other parent halo masses.

time-consuming. For this reason, we fit analytical functions to the PDFs of all second- and third-order subhaloes of different masses of interest here, and we report the best-fitting parameters for equation (3) in Table 1. We recall that the PDFs for first-order subhaloes are instead directly computed by the change of the SHMF along the parent halo mean mass accretion track (Section 3.4). The right-hand panel of Fig. 3 compares the number densities of the infall redshifts of first- and second-order subhaloes accreting on to a parent haloes of mass between 10^{14} and $10^{14.1}$ $M_\odot$ as predicted by DECODE (dashed lines) and the Millennium simulation (solid lines). The agreement is good, further validating the accuracy of our modelling.

4 RESULTS

As introduced in the previous Sections, DECODE is a flexible statistical SEM that can predict, for a given input SMHM relation, the implied star formation and merger histories of central galaxies. It is in concept similar to its predecessor, STEEL, but it has relevant new features, including the separation in subhalo accretion order and a more refined treatment of infall time-scales.

In this work we will mostly focus on the ability of DECODE to rapidly predict the mean merger histories of central galaxies of different stellar mass for different input SMHM relations. In a separate work, we will extend DECODE to predict star formation

Table 1. Best-fitting parameters of the infall redshift distribution parametrization from equation (3), for different subhalo orders and mass intervals.

Subhalo order	Mass range	A	α	β	γ	δ
Second	$10 < \log_{10}(M_{\rm h,sub}/{\rm M}_\odot) < 11$	$10.06^{+3.51}_{-3.86}$	$1.07^{+0.16}_{-0.14}$	$13.78^{+3.51}_{-5.78}$	$1.86^{+9.25}_{-7.80}$	$1.77^{+0.44}_{-0.37}$
Second	$11 < \log_{10}(M_{\rm h,sub}/{\rm M}_\odot) < 12$	$24.93^{+6.24}_{-10.31}$	$1.77^{+0.22}_{-0.18}$	$8.87^{+3.39}_{-3.64}$	$0.42^{+6.22}_{-6.71}$	$1.94^{+0.43}_{-0.35}$
Second	$\log_{10}(M_{\rm h,sub}/{\rm M}_\odot) > 12$	$24.81^{+5.64}_{-9.19}$	$2.67^{+0.26}_{-0.21}$	$2.31^{+0.88}_{-0.87}$	$-2.60^{+2.52}_{-4.08}$	$1.86^{+0.45}_{-0.34}$
Third	$10 < \log_{10}(M_{\rm h,sub}/{\rm M}_\odot) < 11$	$8.37^{+2.76}_{-3.36}$	$1.50^{+0.21}_{-0.18}$	$13.50^{+3.74}_{-5.79}$	$3.00^{+10.33}_{-8.75}$	$3.29^{+0.53}_{-0.51}$
Third	$\log_{10}(M_{\rm h,sub}/{\rm M}_\odot) > 11$	$23.71^{+7.73}_{-12.63}$	$2.54^{+0.31}_{-0.31}$	$3.41^{+1.57}_{-1.70}$	$-3.56^{+4.59}_{-4.65}$	$3.12^{+0.70}_{-0.49}$

histories and the amount of intracluster light generated from the stellar stripping of infalling satellites.

G20 identified a major role played by the SMHM relation in shaping the merger history of a central galaxy, especially in the case of more massive galaxies. In brief, if the high-mass end of the SMHM is flatter, it will imply that a halo increasing with mass via mergers will correspond to progressively lower growth in the stellar mass of its central galaxy. In the extreme condition of a perfectly flat SMHM relation, any increase in halo mass would not be followed by any increase in stellar mass, i.e. the (especially major) merger rate of those type of central galaxies will be drastically reduced.

In this Section, we further expand on G20 taking into account the different SMHM relations as derived from abundance matching using the latest data on the SMF at low and high redshifts (Section 4.1). We will then discuss the implied merger rates (Section 4.2), number densities of 'unmerged galaxies', i.e. satellites, in the local Universe for different SMHM relations (Section 4.3), the implied fraction of ellipticals originating from major mergers (Section 4.4), and the distribution of B/T stellar mass ratios of galaxies in the local Universe induced by major mergers and some models of disc instabilities (Section 4.5). We will also present the prediction of the growth histories of BCGs in Section 4.6. We will show that, as expected, the aforementioned quantities are highly dependent on the input SMHM relation in a hierarchical DM-dominated framework of galaxy evolution.

4.1 Stellar mass–halo mass models

As anticipated above, and further discussed in detail by many groups (e.g. Wang & Jing 2010; Guo et al. 2011; Moster et al. 2010, 2013), the SMHM relation is strongly dependent on the shape and evolution of the measured SMF. Unfortunately, the latter is far from known with sufficient accuracy, not even in the local Universe. Bernardi et al. (2013, 2016, 2017), for example, showed that the SMFs in *SDSS* and CMASS at $z = 0.1$ and $z = 0.5$, respectively, are highly dependent on the light profile chosen to fit the photometry. When the same methodology is consistently applied to infer stellar masses, no apparent evolution is detected in the high-mass end of the SMF up to at least $z = 0.5$–0.8 (see also Shankar et al. 2014), whilst significant evolution at all masses is inferred when comparing Bernardi et al. (2017) and, e.g. Davidzon et al. (2017), who derived stellar masses from SED fitting. Additional systematic differences in SMFs have been claimed by, e.g. Leja et al. (2020), who support larger stellar masses at fixed SFR. The substantial systematic uncertainties in measured SMFs naturally propagate into the shape, scatter, and evolution of the SMHM relation. In what follows, we present four models for the SMHM relation derived from direct abundance matching, as detailed in Section 3.5, with different observed SMFs. Our aim is to estimate to what extent observationally informed systematic differences in the input SMHM relation impact the implied galaxy merger rates and bulge fractions, a task that is particularly suited to address with DECODE. The four SMHM models considered in this work can be summarized as follows:

(i) Model 1:

(a) Bernardi et al. (2017) SMF, assumed to be constant up to $z \approx 1.5$. This is of course an extreme assumption but worth exploring as it is still unclear whether apparent evolution in the SMF at $z > 0$ may be, at least in part, driven by non-ideal/inconsistent estimates of galaxy stellar masses, as mentioned above. Indeed, Kawinwanichakij et al. (2020) recently suggested that there is no measurable evolution in the SMF up to at least $z = 1.5$.

(b) Tinker et al. (2008) HMF.

(c) 0.15 dex constant scatter in stellar mass at fixed halo mass.

(ii) Model 2:

(a) Tomczak et al. (2014) SMF at $z > 0$ and Bernardi et al. (2017) at $z = 0$. This model is also somewhat extreme because, as detailed above, the $z > 0$ SMF estimates may be affected by systematic measurement errors and/or incomplete. Nevertheless Models 2 and 1 should bracket the range of possible evolutionary patterns of the SMF, at least based on the present data and on the assumption of a constant IMF.

(b) Tinker et al. (2008) HMF.

(c) 0.15 dex constant scatter in stellar mass at fixed halo mass.

(iii) Model 3:

(a) equivalent to Model 1 but assuming a linearly increasing scatter with redshift up to $z = 2$ as follows:

$$\begin{cases} \sigma_{\log M_*} = 0.15 + 0.1z & \text{for } z < 2 \\ \sigma_{\log M_*} = 0.25 & \text{for } z \geq 2 \end{cases} \quad (10)$$

(iv) Model 4:

(a) equivalent to Model 2 but with the same z-dependent scatter as Model 3.

All the reference Models listed above start from the same $z = 0$ SMF. The latter is built joining the Bernardi et al. (2017) SMF, which is valid down to $M_* \sim 10^9\,{\rm M}_\odot$, and the Baldry et al. (2012) SMF which extends down to $M_* \sim 10^6\,{\rm M}_\odot$ (see e.g. Shankar et al. 2006; Kravtsov et al. 2018). In Models 1 and 3, as detailed above, we strictly assume no evolution in the SMF up to $z = 1.5$. Beyond this redshift it becomes unrealistic to assume no further evolution in the SMF and thus we extend the SMF at higher redshift with a toy model that smoothly decreases the normalization of the SMF from $z = 1.5$ using the following log-linear correction, which simultaneously

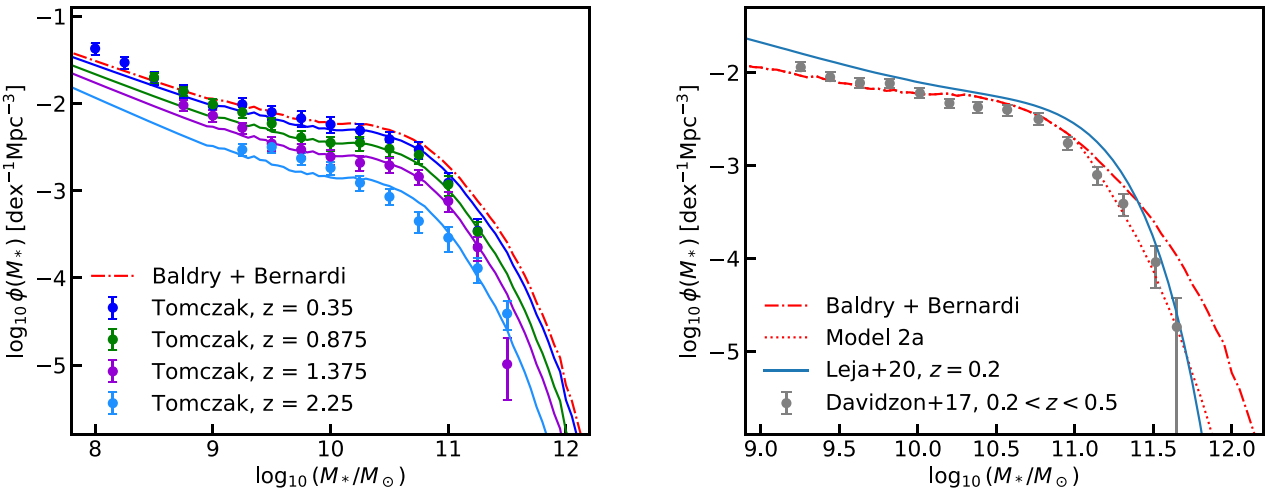

Figure 4. Stellar mass functions used for the abundance matching of the SMHM for the different models described in Section 4.1. In both panels, the red dashed–dotted line represents the combination of the SMF from Baldry et al. (2012) and Bernardi et al. (2017) at $z = 0.1$. In the left-hand panel, the dots with error bars represent the observational data from Tomczak et al. (2014) and the coloured solid lines the redshift-dependent normalization correction (equation 12) applied to the Baldry + Bernardi SMF to be consistent with the Tomczak + 14 data at $z > 0$. In the right-hand panel, the dots with error bars show the data from Davidzon et al. (2017), while the blue solid line and red dotted line show the SMF from Leja et al. (2020) and the correction (as described in Section 4.1) to the Baldry + Bernardi SMF for being consistent with the Davidzon + 17 data, respectively.

allows to track the Tomczak et al. (2014) SMF data at $z > 1.5$ and to be consistent with the Bernardi et al. (2017) SMF up to $z = 1.5$

$$\log_{10}\phi(M_*(z)) \simeq (0.99 + 0.13(z - 1.5)) \times \log_{10}\phi(M_*(z = 0.1)). \quad (11)$$

In Models 2 and 4, we instead assume the SMF to continuously evolve in normalization at $z > 0$ in a way to be broadly consistent with the Tomczak et al. (2014) SMFs at $0 < z < 3$, as shown in the left-hand panel of Fig. 4. When applying the Tomczak et al. (2014) evolution to the SMF, we make use of the following correction to the Bernardi et al. (2017) + Baldry et al. (2012) SMF at $z = 0$

$$\log_{10}\phi(M_*(z)) \simeq (0.99 + 0.13z) \times \log_{10}\phi(M_*(z = 0.1)). \quad (12)$$

Furthermore, we also explore a variant of Model 2, which we refer to as Model 2a, with an SMF matching the SMF calibrated by Davidzon et al. (2017) characterized by a less abundant number density of massive galaxies (dotted line in Fig. 4). We will discuss in the following Sections how the latter SMF will modify the implied SMHM relation and in turn the number of merging pairs and fraction of ellipticals. It is relevant here to clarify that our method relies on direct abundance matching between the SMF and HMF at any given epoch. However, while for the latter, analytical fits extracted from *N*-body simulations are available at all redshifts, this is not the case for the SMF, for which analytical fits are provided only in some pre-defined redshift bins. In addition, some of the high redshift data of interest to our work lack or have a poor determination of the high-mass end of the SMF (e.g. Tomczak et al. 2014). These are the main reasons why we need to define a full shape for the input SMF at all redshifts. We reiterate here that the aim of our work is to explore how different shapes and evolutionary trends in the input SMF, within the range allowed by current observations, impact the implied SMHM relation and related quantities such as the galaxy merger rates.

Given the SMFs at any given epoch $z < 4$ and for each Model, it is now necessary to specify the host halo mass functions to then apply the abundance matching routine given in equation 8. We choose the Tinker et al. (2008) HMF, which provides the abundances of host parent haloes (we note that switching to other forms of the HMF would yield very similar results throughout). As the SMFs at $z > 0$ contain both central and satellite galaxies (though the latter become progressively negligible in number densities at earlier epochs), we need to correct the Tinker et al. (2008) HMF by the abundances of surviving satellites at any epoch of interest, as specified in Section 3.5.

The SMHM relations, derived from our abundance matching algorithm for each Model, are shown in Fig. 5 in the redshift range $0 < z < 3$. Models 1 and 3 are characterized by a high-mass slope $\mathrm{dlog_{10}}(M_*/\mathrm{M_\odot})/\mathrm{dlog_{10}}(M_\mathrm{h}/\mathrm{M_\odot})$ (between $M_\mathrm{h} = 10^{13}$ and $M_\mathrm{h} = 10^{14}\,\mathrm{M_\odot}$) of 0.550, and a low-mass one (between $M_\mathrm{h} = 10^{11}$ and $M_\mathrm{h} = 10^{11.5}\,\mathrm{M_\odot}$) of 1.30, both with a normalization of 10.509 at $M_\mathrm{h} = 10^{12}\,\mathrm{M_\odot}$, at $z = 1$. Models 2 and 4 instead have a high-mass slope of 0.588 and 0.508, a low mass slope of 1.15, and a normalization of 10.138 and 10.163, respectively. As shown in Fig. 5, our Model 2 at $z = 0$ is very close to both Moster et al. (2018) and Behroozi et al. (2019), at least at high stellar masses. The main difference between Model 2a and the others is the significantly flatter slope of 0.414 at the high-mass end. We will show below that such apparently small differences in the shape of the SMHM relations, especially the differences in the slopes at the high-mass end of the SMHM relations, are sufficiently large to generate significant systematic differences in, e.g. the major merger rates and implied elliptical fractions of up to a factor of 2–4 in some stellar mass bins. We notice that by keeping the SMF constant in time (Models 1 and 3) induces a weak evolution at low masses and a more pronounced one at larger masses. On the other hand, the models characterized by an evolving SMF (Models 2 and 4), generate an SMHM relation with evident redshift evolution at low stellar masses and a weak one at higher stellar masses (see also Shankar et al. 2006; Moster et al. 2018, G20).

Finally, we also investigated the possibility of a mass-dependent scatter in stellar mass at fixed halo mass as input in our abundance matching. To this purpose, we assume a halo mass-dependent scatter, similarly to what suggested by other SEMs (e.g. Moster et al. 2018;

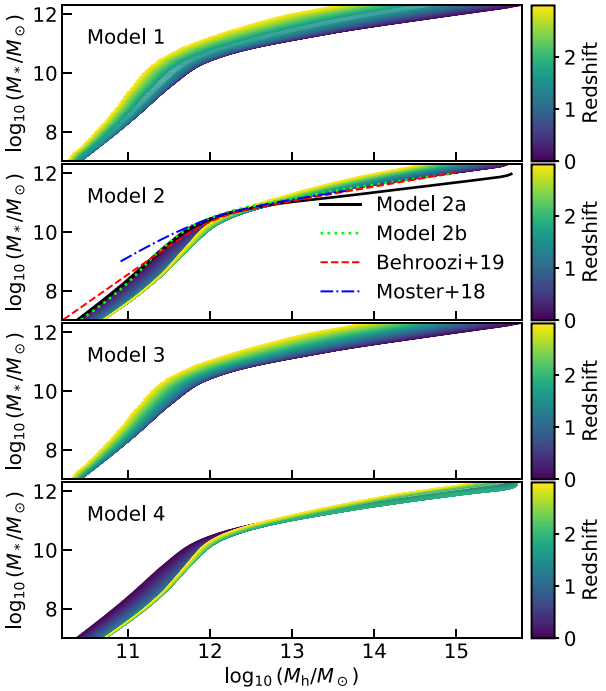

Figure 5. Stellar mass–halo mass relations computed via abundance matching. The four panels show the relations for the four models described in Section 4.1, respectively, for the range of redshift denoted by the colour code. In the second panel, we show also Models 2a and 2b, along with two SMHM relations from other works in the literature (Moster et al. 2018; Behroozi et al. 2019) for comparison (black solid, green dotted, red dashed, and blue dashed–dotted lines, respectively).

Behroozi et al. 2019), constant at $M_{\rm h} > 10^{12}\,{\rm M}_\odot$ and increasing linearly with $\log_{10}(M_{\rm h}/{\rm M}_\odot)$ below that mass

$$\begin{cases} \sigma_{\log M_*} = 2.95 - 0.23\log_{10}(M_{\rm h}/{\rm M}_\odot) & \text{for } M_{\rm h} < 10^{12}\,{\rm M}_\odot \\ \sigma_{\log M_*} = 0.15 & \text{for } M_{\rm h} \geq 10^{12}\,{\rm M}_\odot\,. \end{cases} \quad (13)$$

We show the SMHM relation at $z = 0$ implied by the scatter in equation (13) in the second panel of Fig. 5, labelled as Model 2b. The comparison shows that Model 2b is fully equivalent to Model 2 at high masses and slightly steeper at low masses, and these differences we checked are similar at all redshifts. We have tested that, since Model 2b is equivalent to Model 2 in the mass range where major mergers are significant, i.e. $M_\star \gtrsim 10^{11}\,{\rm M}_\odot$, the predicted amount of major mergers remains unchanged and, hence, also all the other quantities analysed in this work. Therefore, we do not show the results for Model 2b any further, and concentrate on the two models with redshift-dependent scatter (Models 3 and 4), which alters also the high-mass end of the SMHM relation and consequently the major merger rates.

4.2 Merger rates

The input SMHM relation has a direct impact on the merger rate of each galaxy that DECODE produces (e.g. Stewart et al. 2009b; Hopkins et al. 2010b; Grylls et al. 2020b; O'Leary et al. 2021), which in turn influences the implied satellite abundances, fraction of ellipticals, and B/T ratios. In this Section, we focus on galaxy merger rates and other predictions will be discussed in the following Sections.

First of all, we checked that our halo–halo merger rates are consistent with the halo–halo merger rates derived by Fakhouri et al. (2010) from the Millennium simulation at $z \gtrsim 0.35$. We note that our fits drop slightly faster than those of Fakhouri et al. (2010) at $z < 0.35$, as also previously noted by G20, which we checked is mostly induced by our adopted halo mass mean accretion histories from van den Bosch et al. (2014) which are somewhat steeper than those presented in Fakhouri et al. (2010) at these redshifts. Fig. 6 shows the cumulative number of total and major mergers from DECODE (upper and lower panels, respectively) below $z < 4$ predicted from DECODE, expected in a hierarchical ΛCDM Universe, as a function of parent halo mass. We found that the number of mergers, with both first- and second-order subhaloes, is roughly constant with parent halo mass (see also Shankar et al. 2014). However, such an invariance in halo mass is broken when mapping DM halo mergers to galaxy mergers via the double power-law shaped SMHM relation. We show this result in Fig. 7, where we plot the average number of major mergers, along with its 1σ uncertainty, as a function of the (final) galaxy stellar mass. The red lines represent the contribution of mergers from first-order satellites and the blue lines the contribution of mergers from second order. The plot clearly shows that, for all the four SMHM models, the contribution from the second-order satellites is in good approximation relatively negligible, at least for massive

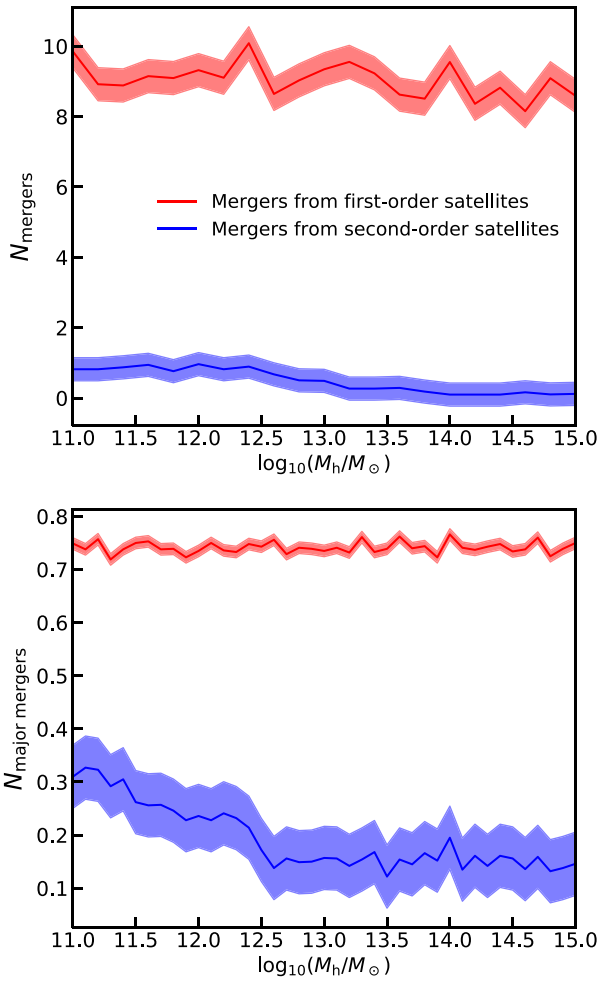

Figure 6. Upper panel: number of dark matter mergers from the contribution of first- and second-order subhaloes as function of the final parent halo mass at redshift $z = 0$. The solid lines represent the mean value, while the shaded areas show the 1σ uncertainty. Lower panel: same as upper panel, but for major mergers, for which we assume a mass ratio $M_{\rm h,sub}/M_{\rm h,par} > 0.25$.

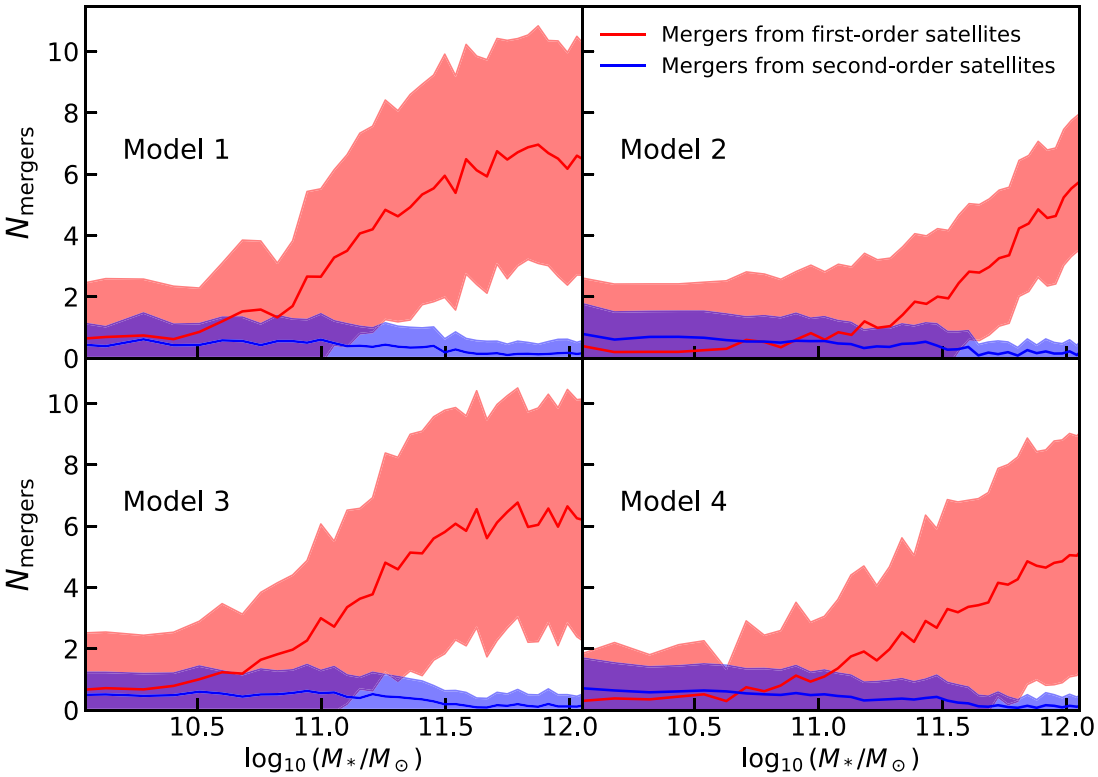

Figure 7. Number of galaxy major mergers from first- and second-order satellites, with mass ratio $M_{*,\rm sat}/M_{*,\rm cen} > 0.25$, as a function of the central galaxy mass. The solid lines represent the mean value, while the shaded areas show the 1σ uncertainty. Results are shown for the four different models described in Section 4.1, as labelled.

central galaxies. On the other hand, the number of mergers from the first-order satellites increases significantly when the central galaxy mass increases.

The sharp increase of major mergers with galaxy stellar mass and the fact that major mergers for massive galaxies outnumbers the major mergers of the parent haloes can be both explained from the shape of the SMHM relation. Indeed, while the total number of mergers is the same for the DM halo and the host galaxy (and it is independent of the halo or stellar mass), the classification as major or minor merger depends on the mass ratio between halo or stellar mass ($M_{\rm h,1}/M_{\rm h,2} > 0.25$ is the major merger threshold for halo mergers and $M_{\star,1}/M_{\star,2} > 0.25$ for galaxy mergers). Since the SMHM relation sets $M_{\rm h} \propto M_\star^\alpha$, the halo mass ratio is related to the stellar mass ratio as $M_{\rm h,1}/M_{\rm h,2} = (M_{\star,1}/M_{\star,2})^\alpha$. If the SMHM relation was linear ($\alpha = 1$) the halo mass ratio would be equal to the stellar mass ratio and, consequently, a halo major merger would correspond to a galaxy major merger. On the other hand, a steep power-law relation with $\alpha > 1$ would decrease the stellar mass ratio with respect to the halo mass ratio, while a flat power law with $\alpha < 1$ would increase it. As a consequence, the number of galaxy major mergers is reduced for a steep power law, while it is enhanced for a flat one. In other words it is less (more) likely to find similar mass galaxies for similar mass haloes for a steep (flat) SMHM. Since the SMHM relation in all the four considered models is a broken power law with a steep faint end and a flat bright end, the number of galaxy major mergers tends to naturally increase towards higher stellar masses. However, the level of increase is different for the four models since even small variations in the SMHM relation produce strong effects in the merger rates. For example, Model 2 with a high-mass slope of 0.558 predicts less than half of the mergers at $M_* \sim 1$–$3 \times 10^{11}\,{\rm M}_\odot$ compared to Model 1 and Model 3 with a high-mass slope of 0.550. On the other hand, Model 4, despite having the lowest high-mass end slope at $z = 0$, produces slightly more mergers than Model 2 but less than Models 1 and 3, a trend that can be explained by the lower slope in the SMHM relation at higher redshifts in Model 4 than in Model 2, being e.g. 0.498 and 0.551 the average slope up to $z = 1.5$, respectively. For more massive galaxies, Models 1 and 3 predict on average a higher number of major mergers compared to Models 2 and 4. This effect is also directly reflected in the fraction of ellipticals and B/T ratios, as we will see in the following Sections. We also note that, although the normalization of the SMHM does not directly impact the number of major mergers, it has an important role in determining the total merger history, which in turns will impact the star formation history, as we will discuss in future work. Furthermore, the amount of scatter in the SMHM relation instead seems to play a relatively less significant role in controlling the fraction of major mergers when assuming a constant SMF.

Furthermore, we show in the upper panel of Fig. 8 the major merger pair fraction as predicted by DECODE for Models 1, 2, and 2a. We compare our results with the data from UKIDSS UDS, VIDEO/CFHT-LS, UltraVISTA/COSMOS, and GAMA surveys, presented in Mundy et al. (2017). The pair fraction in DECODE is calculated as the number of infalling satellites with stellar mass ratio above 1/4 living within 5 and 30 kpc from the centre of the central galaxy. We assume that the distance of the satellite galaxies scales proportionally to its dynamical friction time-scale (Guo et al. 2011). We make use of the projected 2D distances computed following the recipe in Mundy et al. (2017) and Simons et al. (2019), assigning

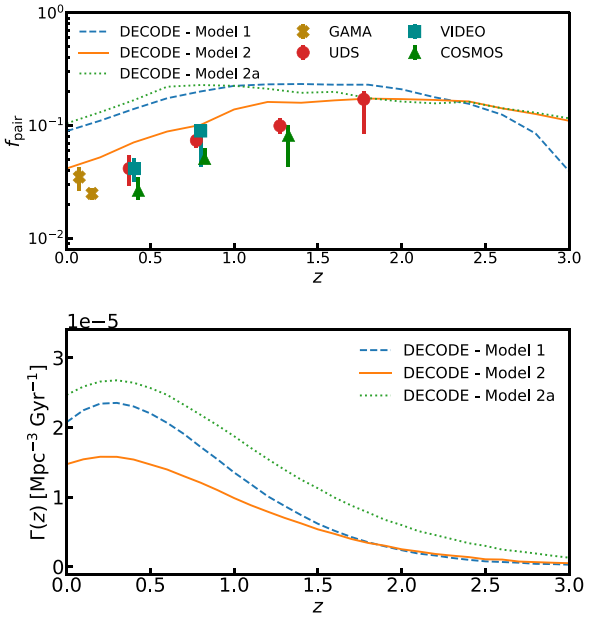

Figure 8. Upper panel: major merger pair fraction, with mass ratio $\mu \gtrsim 0.25$, for galaxies at $M_* \gtrsim 10^{11}$ M$_\odot$ as predicted by DECODE for Models 1, 2, and 2a (blue dashed, orange solid, and green dotted lines, respectively). The points with error bars show the observational data of UDS, VIDEO, COSMOS, and GAMA surveys, as presented in Mundy et al. (2017). Lower panel: major merger rates as a function of redshift, as predicted by DECODE's Models 1, 2, and 2a.

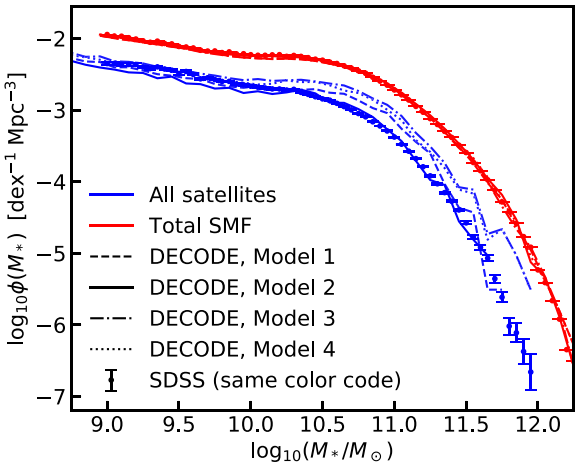

Figure 9. Total and satellite galaxies stellar mass function predicted by DECODE compared to the data from *SDSS*. The solid, dashed, dashed–dotted, and dotted lines show the prediction for Models 1, 2, 3, and 4, described in Section 4.1, respectively. The dots and error bars represent the data from *SDSS*.

stochastically a polar angle in spherical coordinates and projecting the 3D distances on to the z-axis. Analogously to what inferred for the number of major mergers, Model 1 tends to produce more pair fractions with $\mu > 1/4$ than Model 2. The available data on pairs is still sparse, and still subject to systematics in the determination of the stellar masses, but overall Model 2 tends to be more aligned with the data, at least at $z \geq 0.5$. On the other hand, Models 1 and 2a, characterized by a flatter slope in the SMHM relation at the high-mass end, predict a higher pair fraction with respect to the data and to Model 2. This prediction is in line with the fact that a flatter SMHM relation produces a higher number of major mergers, as discussed above.

Finally, in the lower panel of Fig. 8 we show a prediction of the major merger rates from DECODE's Models 1, 2, and 2a. As we can see, the flatter the SMHM relation (Models 1 and 2a) the higher the rate of implied major mergers, in line with what shown in the upper panel. On the other hand, Model 2, characterized by a steeper SMHM high-mass end, predicts a much lower major merger rate, at least at lower redshifts. We do not show the comparison with observational data or any other model because the merging time-scales that we use (McCavana et al. 2012) are different to those adopted in other theoretical and observational works.

In the next Sections we will show and discuss how the different shapes and evolution of the input SMHM relation play a crucial role in determining the satellite abundances, fraction of ellipticals, and B/T ratios.

4.3 Predicting the abundances of satellite galaxies

Having defined the mapping between galaxy stellar mass and host DM halo mass, we can start to predict galaxy observables that can be used to validate DECODE and the input SMHM relation. As a very first test, we compute the number density of surviving satellite galaxies in the local Universe and compare it with *SDSS* data. Satellites can be effectively considered as the other side of the same coin with respect to mergers. In fact, surviving satellites in a hierarchical DM-dominated Universe, can be interpreted as 'failed mergers', i.e. all those infalling satellite galaxies that have not yet had the time to merge with their central galaxies at the time of observation. Therefore satellites, just like mergers, represent a pivotal test of hierarchical models and of the input SMHM relation. Although the total SMF is an input in DECODE, the satellite SMF is an actual prediction of the model, as it depends both on the satellite evolution after infall and on the rate of galaxy mergers, which in turn depend on the high-z SMHM relation and on the dynamical friction time-scales.

Fig. 9 shows the results of the SMF for all galaxies (centrals + satellites) and only satellites (red and blue lines, respectively). The different types of line distinguish the four SMHM models, and the blue and red dots with error bars are, respectively, the satellite[9] and total galaxy SMF as measured in *SDSS* using, for consistency with our Models, the data from Bernardi et al. (2017). As one can see from the red lines, the total SMF is well reproduced by DECODE, as expected by construction via the abundance matching relation given in equation (8). For all SMHM models reported in Fig. 5, we assume a 'frozen' scenario in which satellite galaxies are assumed to retain the same stellar mass after infall with no further growth via star formation or loss via, e.g. stellar stripping. G19 showed that the frozen model was able to reproduce the bulk of the observed satellite population in their SEM STEEL. We do find a similar result with DECODE in Fig. 9, despite using a significantly more accurate distributions of satellites of first and second order and dynamical friction time-scales. We will explore in separate work the impact of star formation and stellar stripping on the satellite population and star formation histories. We anticipate here that including standard recipes for stellar stripping as given by, e.g. Cattaneo et al. (2011), and for star formation after infall following the analytic recipes by G20 (and references therein), we obtain very similar results to those shown in Fig. 9. Fig. 9 shows that only Model 2, the one

[9]Centrals and satellites classification is from Yang et al. (2007).

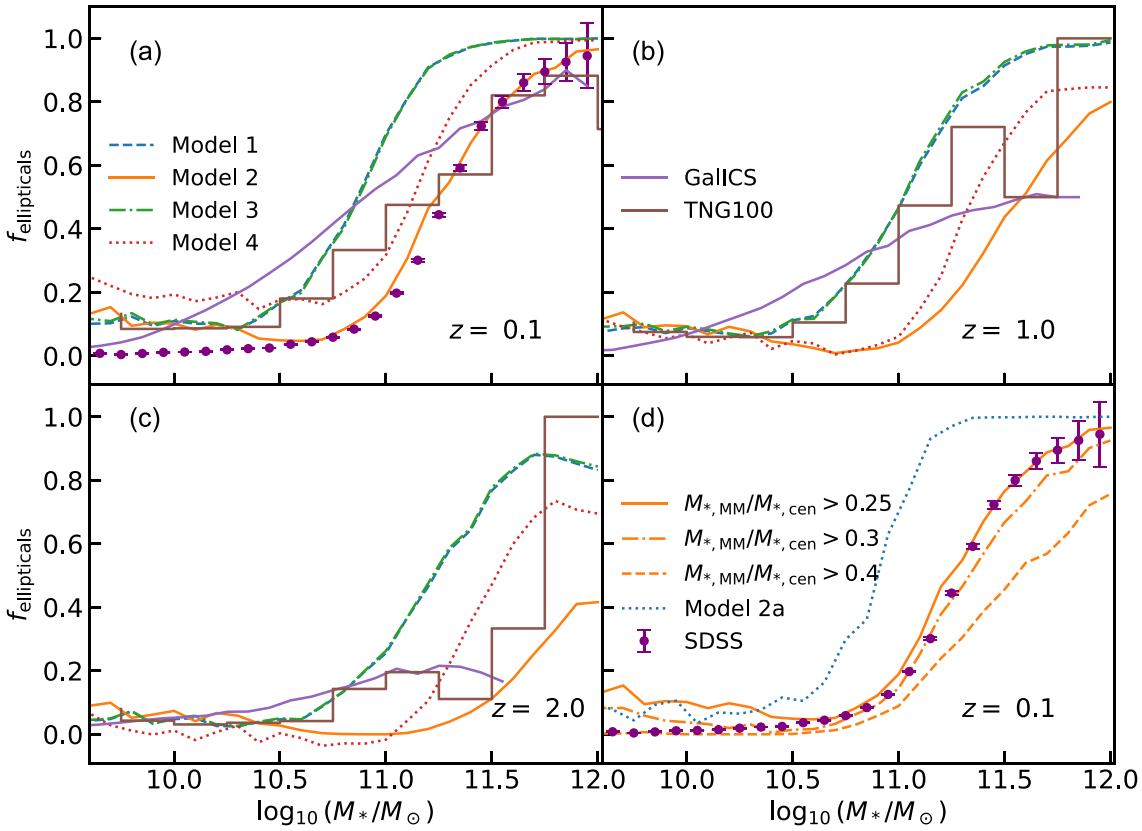

Figure 10. Fraction of elliptical galaxies as a function of the galaxy stellar mass as predicted by DECODE. Panels A, B, and C show the predictions for the four models described in Section 4.1 at redshifts $z = 0.1$, 1, and 2, respectively. Panel D shows the predictions for Model 2 at redshift $z = 0.1$, along with the impact of assuming different mass ratio for the major mergers (denoted as MM in the legend), as well as the prediction for Model 2a. The data from the *SDSS* survey, GALICS semi-analytical model and the TNG100 hydrodynamical simulation are included, as labelled, for comparison.

characterized by an evolving SMF and constant scatter, better lines up with the observational data of satellites, especially at stellar masses $M_* \gtrsim 3 \times 10^{10}\,\mathrm{M}_\odot$. We will see below that Model 2 also performs better against other observational constraints.

4.4 Morphology of central galaxies

In this section, we apply DECODE to study another key aspect of galaxy evolution: the role of mergers in shaping galaxies, and in particular in generating elliptical type galaxies. Many works have in fact suggested a close link between the number of major mergers and the number of ellipticals (e.g. Bournaud et al. 2007; Hopkins et al. 2009a, 2010a; Shankar et al. 2013; Fontanot et al. 2015). Major mergers (for which we assume $f = M_{*,\mathrm{sat}}/M_{*,\mathrm{cen}} \gtrsim 0.25$) may be capable of destroying pre-existing galactic discs and form stellar spheroids, as suggested by hydrodynamical simulations and analytical models (see e.g. Baugh 2006; Malbon et al. 2007; Bower et al. 2010; Tacchella et al. 2019; Lagos et al. 2022). However, as discussed by G20 and above, as the number of major mergers is strongly dependent on the shape of the input SMHM relation, we expect the fraction of ellipticals to be similarly impacted by the choice of SMHM relation. We will investigate this possibility in this Section.

For each SMHM model, we follow the merger history of galaxies in the mock from redshift $z = 4$, and we label each galaxy that has undergone a major merger as an 'elliptical', and then compute the fraction of ellipticals at different redshifts. We show the results in Fig. 10. Panels A, B, and C show the fraction of ellipticals predicted by DECODE using the four Models of SMHM described in Section 4.1 for redshifts $z = 0.1$, 1, and 2, respectively, along with the predictions from the TNG simulation and from the semi-analytical model GALICS, as well as with the *SDSS* data (see Section 2.4), only shown in the panels at $z = 0.1$. We note that, as discussed in Section 2, GALICS and TNG define elliptical galaxies in a somewhat different way than a simple cut in μ, but we still include their predictions in Fig. 10 for completeness.

It is immediately clear from Fig. 10 that Model 2 is the only one among our four chosen SEMs that can faithfully reproduce the *SDSS* data at $z = 0.1$, on the assumption that ellipticals are strictly formed from major mergers above a mass ratio of $\mu > 0.25$, the standard limit adopted in state-of-the-art SAMs (e.g. Guo et al. 2011; Fontanot et al. 2015; Lacey et al. 2016). Model 4, which only differs from Model 2 on the assumed scatter around the SMHM relation, is moderately close, but still higher than the data for the same cut in merger ratio $\mu > 0.25$, further proving that even a modest variation in the scatter of the SMHM relation can significantly impact galactic outputs. In particular, an increase of the scatter at fixed slope at high stellar masses tends to increase the number of major mergers. On the other hand, Models 1 and 3 produce a significantly higher fraction of ellipticals. The main reason why models including evolution in the SMF predict less elliptical galaxies than models with constant SMF is a direct consequence of the number of mergers that they predict, as discussed in Section 4.2. Fig. 10 also shows that the TNG100 simulation provides a decent match to the *SDSS* at

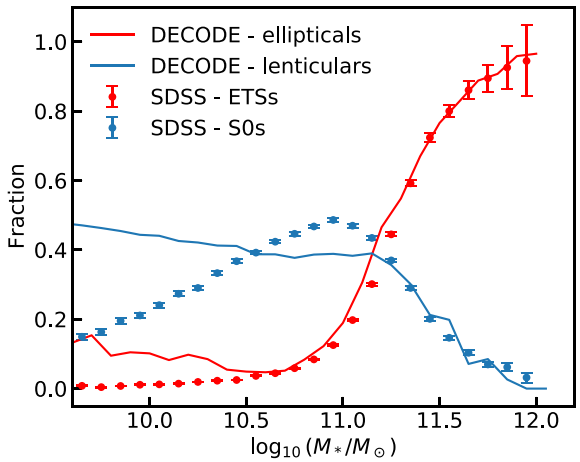

Figure 11. Fraction of lenticular and elliptical galaxies as predicted by DECODE for Model 2 compared with the data from the *SDSS* Survey, as labelled.

$z = 0.1$, especially at galaxy masses $M_* > 2 \times 10^{11}$ M$_\odot$, although we should caution that some internal self-inconsistencies naturally arise when comparing with the TNG100 outputs with these data as the TNG100 simulation does not exactly reproduce the same SMF of Bernardi et al. (2017) (see e.g. Pillepich et al. 2018b), on which the elliptical fractions are based, and thus has an SMHM relation that slightly differs from the one in Model 2. The SAM GALICS[10] also well matches the *SDSS* elliptical fractions at high stellar masses, but it overproduces them at lower stellar masses. Lastly, we show in panel D of Fig. 10 the fraction of ellipticals predicted by Model 2a. This model, characterized by a flatter high-mass end slope, produces a much larger number of mergers, as expected from the discussion in Section 4.2, and thus it increases the implied fraction of ellipticals at all stellar masses.

In conclusion, the question whether major mergers are the triggers for the formation of local ellipticals, is strongly dependent on the SMHM relation in input or generated by the model in use, and, at fixed SMHM relation, on the exact threshold chosen for being classified as a major merger, as shown in panel D.

Finally, we show for completeness the fraction of lenticular-type galaxies as a function of stellar mass in Fig. 11 predicted from DECODE for Model 2, compared with the observational data from *SDSS*. Several works have suggested that lenticular galaxies might be created by mergers as well (e.g. Christlein & Zabludoff 2004; Laurikainen, Salo & Buta 2005; Blanton & Moustakas 2009). Here, we label the lenticulars as the galaxies that have had at least one merger with mass ratio $0.05 < \mu < 0.25$, whilst above $\mu = 0.25$ they would end up as ellipticals. This simple merger recipe is capable of reproducing the data for stellar masses $M_* \gtrsim 10^{11}$ M$_\odot$, but it fails at lower stellar masses, suggesting that additional processes may be at work in forming less massive lenticulars, such as disc instabilities and/or disc regrowth.

4.5 Bulge-to-total ratios

In the previous sections, we showed that the shape of the SMHM relation drives the number and thus the rate of mergers galaxies undergo through cosmic time (at fixed dynamical friction time-scale). Only specific SMHM relations, such as the one defined in Model 2, which is characterized by a larger number density of massive galaxies and a significant evolution in normalization at $z > 0.5 - 1$, are able to simultaneously reproduce the number of local satellites and the number of local ellipticals, on the assumptions that the latter are formed out of mergers between galaxies with a mass ratio $M_{*,1}/M_{*,2} > 0.25$. We now move a step forward in our modelling and test how well our Models 1 and 2 reproduce the B/T ratios of local galaxies, as measured in MaNGA (see Section 2).

To perform a meaningful and instructive comparison between models and data, we make use of two simple but theoretically well-motivated toy models for the formation of bulges in hierarchical models:

(i) In Model BT1, we assume that when a major merger occurs, with $M_{*,1}/M_{*,2} > 0.25$, the descendant galaxy is strictly an elliptical with $B/T = 1$. This is a common assumption made in semi-analytical models of galaxy evolution (e.g. Cole et al. 2000; Hatton et al. 2003; Bower et al. 2006; De Lucia & Blaizot 2007; Guo et al. 2011; Croton et al. 2016; Lacey et al. 2016; Cattaneo et al. 2017). We then assume that in minor mergers the mass of the satellite can be accreted either on to the bulge or on to the disc component.

(ii) In Model BT2, we instead assume that the remnant galaxy has a surviving disc with $B/T = 0.5$. In other words, we assume that the disc is not entirely disrupted in a major merger, irrespective of the gas fraction in the progenitor galaxies, but in fact a significant fraction of it survives and/or is rapidly reaccreted (e.g. Hopkins et al. 2009b; Puech et al. 2012). We then explore the impact on the final B/T of assuming the satellite mass in minor mergers to be added systematically to the disc or to the bulge component. We note that, in what follows, we consider the B/T at fixed galactic stellar mass averaged over *all* central galaxies that enter that bin in stellar mass.

Another popular route to form stellar bulges in galaxies is via disc instabilities, which are usually implemented in SAMs in broadly two ways. A first-type envisions that when the circular velocity of the disc becomes larger than a given reference circular velocity, then the disc is considered unstable and a mass is transferred from the disc to the bulge. The amount of mass transferred from the disc to the bulge varies significantly from model to model (e.g. Cole et al. 2000; Bower et al. 2006; Monaco, Fontanot & Taffoni 2007; Guo et al. 2011; Lacey et al. 2016; Izquierdo-Villalba et al. 2019; Henriques et al. 2020). The disc instabilities of the second type are triggered by high redshift cold flows of gas, which favour the formation of (possibly) long-lived gas clumps that migrate towards the centre via dynamical friction in the gaseous disc (e.g. Dekel & Birnboim 2006; Dekel, Sari & Ceverino 2009; Bournaud et al. 2011; Di Matteo et al. 2012; Oklopčić et al. 2017; Dekel, Lapiner & Dubois 2019). To include an example of the second type of disc instabilities in DECODE to generate stellar bulges, we adopt the parametrization of the baryonic inflow rate from Bournaud et al. (2011)

$$\dot{M}_{\rm b} = 25 \frac{M_{\rm disc}}{10^{11}\,{\rm M}_\odot} \left(\frac{1+z}{3}\right)^{3/2} {\rm M}_\odot\,{\rm yr}^{-1}, \quad (14)$$

with $M_{\rm disc}$ the mass of the disc at redshift z. Equation (14) assumes that most of the mass inflow rate, which is in gaseous form, will form clumps that via dynamical friction will end up forming a stellar bulge at the galaxy centre. We apply this recipe only to galaxies which have a gas fraction $f_{\rm gas} \geq 0.5$ which are more likely to have undergone disc instabilities, since large amount of gas and mass densities inevitably lead to disc fragmentation (see Lang et al. 2014). Gas masses are not present in DECODE. To this purpose, following other SEMs (e.g.

[10]We remind the reader that elliptical galaxies in GALICS are identified as those galaxies with $B/T > 0.7$.

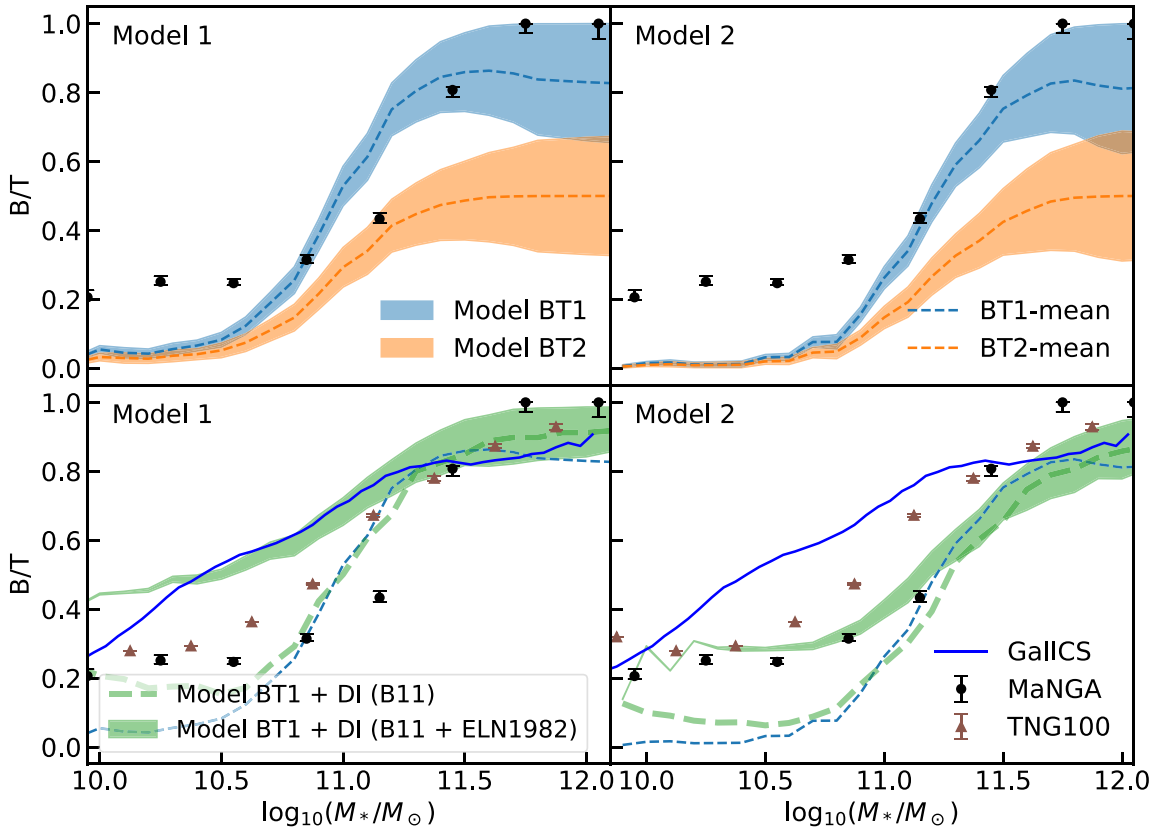

Figure 12. Mean bulge-to-total ratios for the two different toy models described in Section 4.5 and SMHM relationships of Models 1 and 2. The blue and orange shaded areas in the upper panels show the results for Models BT1 and BT2 (for Models 1 and 2, respectively), while the green areas in the lower panels show the results for Model BT1 but by including disc instabilities of Efstathiou et al. (1982) (equation 15). The dashed tick green lines show the B/T ratios for Model BT1 by including only Bournaud et al. (2011) disc instabilities. The shaded areas are constrained by the two limit cases, where all the mass from minor mergers goes into the bulge and disc, respectively, and the thin dashed lines show the mean. The black error bars represent the 1σ bound from the MaNGA survey. Finally, the blue solid line shows the prediction from GALICS semi-analytical model and the brown triangles with error bars the results from the TNG simulation. The uncertainties for the MaNGA survey and the TNG simulation are estimated using the standard deviation on the median.

Hopkins et al. 2009b; Shankar et al. 2014), we assign gas fractions to any galaxy in the mock via the empirical mean relations derived by Stewart et al. (2009a) from a number of galaxy samples out to $z \sim 3$. We find that equation (14) does not generate enough large stellar bulges at $z = 0$, as expected by simple direct integration of equation (14).[11] Therefore, some slightly stronger disc instabilities need to be implemented to better match local data. To this purpose, we follow the recipe from Efstathiou et al. (1982) building on the analytical modelling of many previous works (e.g. Cole et al. 2000; Monaco et al. 2007)

$$\epsilon \sqrt{G M_{\rm disc}/R_{\rm disc}} > V_{\rm ref}, \tag{15}$$

where $M_{\rm disc}$ is the mass of the disc, $R_{\rm disc}$ is the half-mass radius calculated via the redshift-dependent analytical fit by Shen et al. (2003), $V_{\rm ref}$ is the reference velocity calculated assuming an exponential profile (see e.g. Tonini et al. 2006), and ϵ is a factor of order unity (see e.g. Shankar et al. 2014, and references therein). When the condition in equation (15) is verified in galaxies that still have a disc-dominated structure, we assume that the disc transfers a sufficient stellar mass to the bulge to reestablish dynamical equilibrium (e.g. Hatton et al. 2003; Shen et al. 2003).

The top panels of Fig. 12 report our results for the merger Models BT1 and BT2, for the SMHM relation in Model 1 (left-hand panel) and Model 2 (right-hand panel), shown with cyan- and orange-shaded areas, respectively, where the upper and lower bounds mark the limiting cases where all the stellar mass of the minor merger is transferred to the bulge and to the disc, respectively, and the dashed lines represent the mean of the two cases. We compare our predictions with data on the mean B/T as a function of galaxy stellar mass from MaNGA (black dots with error bars), as detailed in Section 2.4. The fact that the shaded areas in SMHM Model 2 are slightly broader at high stellar masses than those in Model 1 is an artefact of the number of mergers predicted by these Models, as shown in Figs 6 and 7. In particular, at fixed number of DM mergers, Model 2 predicts less major mergers, i.e. more minor mergers than Model 1. This leads to a higher bulge/disc mass in the cases where all the mass from minor mergers goes to the bulge/disc, leading, therefore, to a larger bound.

The MaNGA data clearly indicate that all local galaxies have a B/T ratio that is confined within 0.2–1, with a mean value slightly rising with increasing stellar mass, from ∼0.2 reaching an average value of $B/T \sim 1$ only in the most massive galaxies with $M_* \gtrsim 10^{11}$ M$_\odot$. Models that, like our BT1, assume a strict $B/T = 1$ during major mergers, can better reproduce the data at high stellar masses,

[11] We note that the growth of stellar bulges via clumpy accretion can be further increased by lowering the $f_{\rm gas}$ threshold, though it tends to still be moderate, reaching a value of up to $B/T \sim 0.2$ at low stellar masses, $M_\star \lesssim 3 \times 10^5$ M$_\odot$.

especially when assuming that all the minor mergers contribute to the bulge component. On the other hand, models that, like BT2, assume that a significant fraction of the disc survives and/or is rapidly reaccreted after the merger, fall short in matching the data at high stellar masses, suggesting that these kinds of model are not extremely suitable to describe the average B/T ratios, at least at high stellar masses.

Irrespectively of what we assume for the B/T in major (or even minor) mergers, both our BT1 and BT2 Models drastically fail in reproducing the observed B/T in galaxies with $M_* \lesssim 2 \times 10^{11}\,\mathrm{M}_\odot$ (top panels of Fig. 12). An *additional* component/process should be included in DECODE to reproduce the data. Here, again we can witness the usefulness of a semi-empirical approach that, as discussed in the Section 1, can include additional layers of complexities wherever necessary, as guided by the data, in a transparent and efficient way. The bottom panels of Fig. 12 show the prediction of Model BT1 by including disc instabilities as discussed above. We do not show Model BT2 as it fails to model the B/T ratios even with the addition of disc instabilities. The main effect of including disc instabilities in our Model BT1 is that it tends to increase the bulge masses at lower, but not at higher, galaxy stellar masses, where the condition in equation (15) is more easily met mainly because lower mass galaxies retain their disc morphologies for a longer time. In particular, we find that, irrespective of the exact input SMHM relation, to broadly match the data we need to assume that a fraction of the disc mass is transferred to the bulge at each disc instability event, in line with some previous cosmological models (e.g. Bower et al. 2006). In the bottom panels of Fig. 12, we show both the cases where clumpy accretion (following Bournaud et al. 2011) and 'classical' disc instabilities (Efstathiou et al. 1982) are implemented. We find that Model BT1,[12] the one without disc regrowth, with SMHM Model 2 broadly matches the *SDSS* local average B/T ratios when some level of disc instabilities is included in the model, while it fails to match the data at low masses with SMHM Model 1. In particular, disc instabilities implemented following Bournaud et al. (2011) appear not to be sufficient to provide enough boost to the average B/T at the low-mass end to match the data, even if the baryonic inflow rate in equation (14) is doubled. On the other hand, disc instabilities as in Efstathiou et al. (1982) allow to broadly reproduce the observational data of *SDSS*, when the factor ϵ in equation (15) is roughly 0.5. Intriguingly, one could argue that strong disc instabilities, and not necessarily major mergers, are responsible for forming most stellar bulges, even at the highest stellar masses. In fact, the strength of a disc instability in forming a bulge closely depends on the D/T ratio, the more disc mass there is, the more potential there is for the bulge to grow in mass after a disc instability. We however checked that even on the assumption of ineffective major mergers preserving a $B/T \sim 0.2$, the disc instabilities would still fall short in boosting the B/T up to unity at high stellar masses.

In the bottom panels of Fig. 12, for completeness, we compare our predictions with the outputs from the TNG100 simulation and the GALICS SAM (brown triangles with error bars and blue solid line, respectively). Interestingly, both TNG100 and GALICS predict large mean $B/T \gtrsim 0.9$ at high stellar masses in reasonable agreement with our predictions and the data. In addition, they also include bulge formation via disc instabilities (see e.g. Tacchella et al. 2019; Cattaneo et al. 2020), predicting indeed $B/T > 0.2$ at low masses in line with the data, though the mean B/T predicted by GALICS tends to be larger than the one measured in MaNGA towards lower stellar masses (we provide in Appendix D a detailed comparison between the B/T ratio found in MaNGA and those from other samples/studies). Our Model 1 is roughly consistent with the predictions of GALICS at all masses but not with the observational data and the TNG simulation at low masses, while Model 2 behaves in nearly the opposite fashion, highlighting once again the dependence of galactic properties, this time the mean B/T ratios, on the input SMHM relation.

Finally, in Fig. 13 we show the mean mass growth of bulges and discs for four galaxy masses at $z = 0$ for Model BT1 with and without disc instabilities, as labelled. We note that, when disc instabilities are not included, at low stellar masses (e.g. $M_*(z = 0) \sim 10^{10.5}\,\mathrm{M}_\odot$) the disc dominates the overall mass growth of the galaxies and the bulge component is almost negligible, unless disc instabilities are included (long-dashed orange lines). Moving towards higher stellar masses, the bulge begins to be gradually more dominant, as a direct consequence of the increasing number of mergers, and adding disc instabilities does alter this trend noticeably.

4.6 Brightest cluster galaxies history

As a final application of DECODE, we study the stellar mass growth history of BCGs, which are massive elliptical galaxies that constitutes an additional source of information for understanding the evolution of galaxies and large-scale structure. Several studies have addressed already this issue, both from observations (e.g. Whiley et al. 2008; Collins et al. 2009; Stott et al. 2010; Lidman et al. 2012; Bellstedt et al. 2016; Lin et al. 2017; Zhang et al. 2017) and numerical works, such as SAMs (e.g. De Lucia & Blaizot 2007; Contini et al. 2014), SEMs (Shankar et al. 2015), and hydrodynamic simulations (e.g. Pillepich et al. 2018b; Ragone-Figueroa et al. 2018). Here, we show the predictions of DECODE for the stellar mass assembly of BCGs and how these compare to the results from other works.

The results are shown in Fig. 14. The red dashed and solid lines show the total stellar mass fractional growth of BCGs, selected in haloes with present-day mass $M_\mathrm{h}(z = 0) > 8 \times 10^{14}\,\mathrm{M}_\odot$, predicted by DECODE with Models 1 and 2, along with their 1σ uncertainties (shaded areas). The data are compared with the COSMOS data (Cooke et al. 2019), results from the hydrodynamic simulations of Ragone-Figueroa et al. (2018) and SAM of Contini et al. (2014) (as labelled in the figure), selected in the same mass region. According to Model 1 BCGs have already formed most of their mass at redshifts $z > 1.5$, because of the assumed constant SMF up to that redshift which maps the DM halo mass accretion history into a higher average stellar mass growth. On the other hand, assuming Model 2 BCGs have only formed roughly 50 per cent of their mass by $z = 1.5$ and have grown the remaining mass at later epochs.

The BCG stellar masses in Ragone-Figueroa et al. (2018) are computed with the mass within the spherical radius of 50 kpc (M50). Our Model 2 is in relatively good agreement on the mass formation history of BCGs with M50 selected galaxies, predicting a factor of $\sim$1.5 in the mass increase between $z = 1$ and $z = 0$ against the factor 1.4 of Ragone-Figueroa et al. (2018).

Fig. 14 shows that models characterized by a SMHM relation with a significantly evolving underlying SMF as in Model 2, are once

[12] We note that the narrower shaded areas of the Model with disc instabilities are a direct consequence of the transfer of mass from the disc to the bulge and tends to shrink into a single line in the case where disc instabilities take place at any time-step since the formation epoch of the galaxy, which explains also why the bounds are even thinner at low masses where the condition of equation (15) is more easily satisfied.

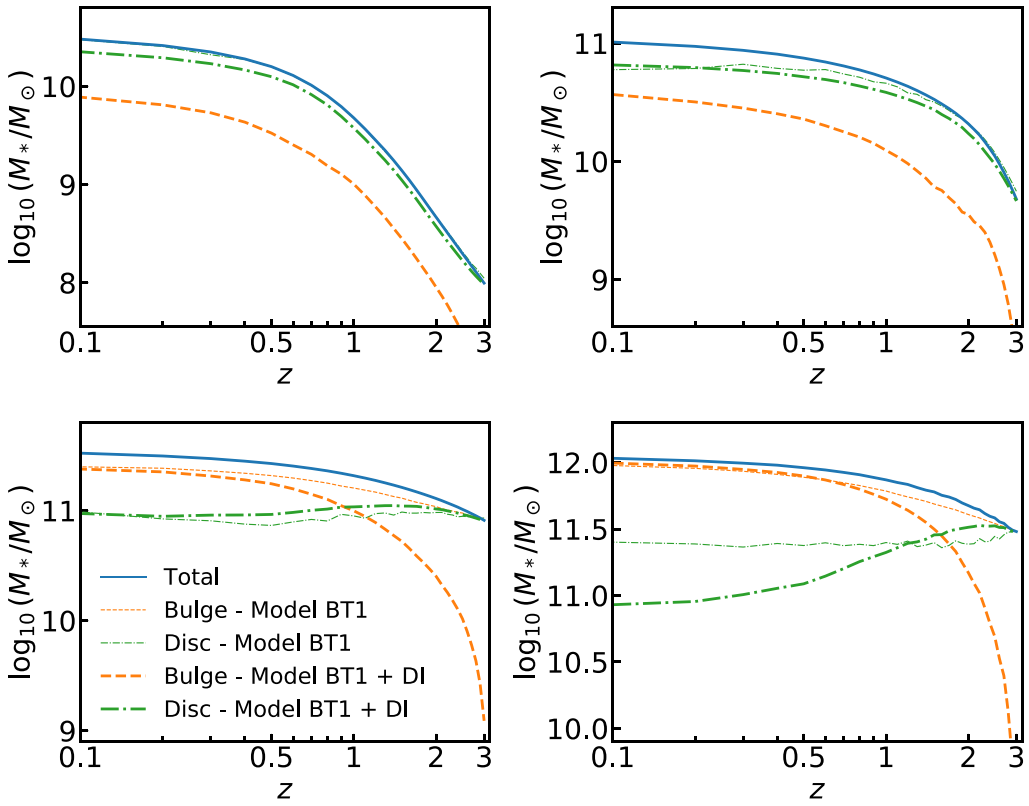

Figure 13. Average growth of total, bulge and disc stellar mass for four different galaxy mass bins at $z = 0$. The solid blue lines show the total stellar mass growth. The thin dashed and dashed–dotted lines show the mass growth of the bulges and discs, respectively, for Model BT1, while the thick lines show the mass growths for Model BT1 including disc instabilities.

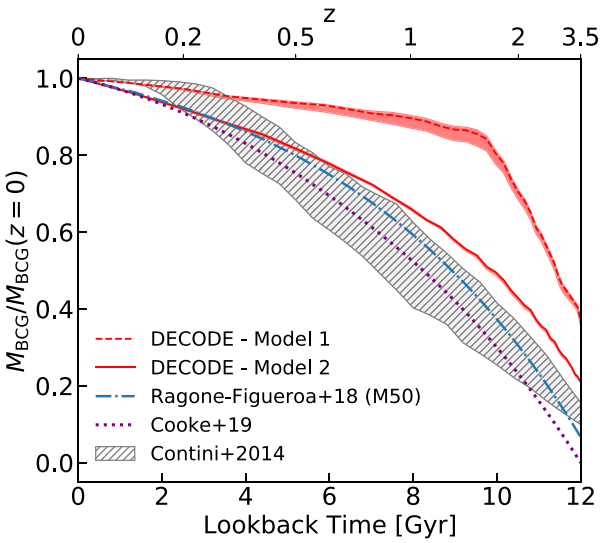

Figure 14. Fractional stellar mass growth of BCGs predicted by DECODE for Models 1 and 2 as a function of lookback time with their 1σ bounds (red lines and shaded areas). The blue dashed–dotted line corresponds to the fit M50 selected BCGs according to equation (2) in Ragone-Figueroa et al. (2018). The purple dotted line shows the observed median growth from COSMOS as presented in Cooke et al. (2019). The grey dashed area includes the evolutionary histories of the models in the Contini et al. (2014) semi-analytical model.

again better tuned to reproduce the current data, this time data on BCGs, in line with predictions from hydrodynamic simulations and SAMs. We note that in all our Models satellites are frozen. Increasing their masses via residual star formation after infall would possibly steepen the evolution for both Models, possibly slightly improving the match with the data for Model 1.

5 DISCUSSION

The discrete statistical and semi-empirical model, DECODE, presented in this work constitutes an invaluable complementary tool to the existing cosmological models for modelling galaxy evolution. These models, either analytical or numerical, are affected by significant volume/mass resolution limitations and/or a large amount of input assumptions and parameters. Also other existing works on SEMs, such as Behroozi et al. (2019) and Moster et al. (2018), based on abundance matching between galaxy–halo properties, still need large computational resources to run. Instead our model, based on statistical input distributions, allows to rapidly simulate a large volume box and investigate the mean properties of galaxies without relying on ab initio analytical models or simulations.

In the last decade several SEMs have been developed. We mention here some examples and discuss how DECODE differs from and complements them. Moster et al. (2018) showed an empirical relation between the mass growth of DM haloes and the galaxy star formation rate, from which they retrieve the mergers history and other properties. Moster et al. (2013) and Behroozi et al. (2019) proposed SEMs where they populate DM haloes via SMHM relations with

galaxies in vast catalogues. These models connect galaxy to DM halo properties via several analytical relations, which could involve a non-negligible number of free parameters. Our present model, instead, although more restricted in scope as it only focuses on mean galaxy assembly histories, essentially relies on only one input parameter, the scatter in the SMHM relation. From this single parameter and abundance matching, DECODE can then generate robust predictions on mean galaxy growth histories, merger rates, satellite abundances, and star formation histories, thus providing a flexible and transparent tool to probe galaxy evolution in a full cosmological context. G19 recently developed an SEM where they apply abundance matching to statistical distributions of DM haloes and galaxies, which is limited by the non-discreteness since it is based on statistical weights over continuous probability densities. Our model instead applies the SMHM relation from abundance matching directly to the mock universe generated stochastically by making a realization of the distributions themselves.

The fact that DECODE starts from the SMF to compute the SMHM relation, makes DECODE a powerful tool to set more stringent constraints on the SMF. As discussed in Section 4.1, the galaxy SMF is far from being well known, especially at high redshifts, and many works suggested different (and sometimes contrasting) results, in terms of shape and/or evolution in time (e.g. Bernardi et al. 2013, 2016, 2017; Tomczak et al. 2014; Davidzon et al. 2017; Huang et al. 2018; Kawinwanichakij et al. 2020; Leja et al. 2020). Our current tests show a preference for SMFs characterized by a larger number of massive galaxies and a significant evolution in time. These SMFs generate SMHM relations that, in turn, produce a sufficient number of mergers to match the local fraction of ellipticals, satellite abundances, and BCG growth. Our study thus highlights the strong dependence of galaxy stellar mass assembly histories on the input SMHM relation. Stewart et al. (2009b) and Hopkins et al. (2010b) also found that galaxy merger rates depend on the input SMHM relation. Their results, along with ours, imply that, for a fixed DM merger tree, the major merger rates and other quantities strongly depend on the mapping between stellar mass and halo mass, which in turns depends on the systematics, shape and evolution of the SMF. Following their path, a fraction of 40 per cent of mass-loss during mergers can alter significantly the merger rates and, therefore, also the fraction of elliptical galaxies and B/T ratios, up to more than a factor of 2, which would allow also flatter SMHM relations to perform well.

In this work, we also found evidence for the need of disc instabilities to boost the formation of bulges at lower stellar masses. This result is in line with the general notion of fast and slow rotators (e.g. Bernardi et al. 2019; Domínguez Sánchez et al. 2020), which suggest that the former dominates at $M_* \lesssim 10^{11.5}\,\mathrm{M}_\odot$ and the latter at $M_* \gtrsim 10^{11.5}\,\mathrm{M}_\odot$, or also with the distinction between pseudo- and classical bulges (see discussions in Gadotti 2009; Fisher & Drory 2010; Shankar et al. 2012, 2013, and references therein). Similar findings are retrieved in SAMs. Guo et al. (2011) found that at stellar masses $M_* \gtrsim 10^{11}\,\mathrm{M}_\odot$ mostly all galaxies have a $B/T > 0.7$, and mostly $B/T < 0.7$ below the same stellar mass threshold, in line with the results of this work. Similar conclusions are derived from GALICS SAM and the TNG100 simulation (Fig. 12), as well as from other works in the literature (e.g. Fontanot et al. 2015; Kannan et al. 2015; Rodriguez-Gomez et al. 2015; Tacchella et al. 2019). Shankar et al. (2014) found that if strong disc instabilities are included in the models, then these could also contribute to a stronger dependence of the mean galaxy size on environment (halo mass), which is in line with our results that point to relatively mild disc instabilities.

6 CONCLUSIONS

In this paper we have presented DECODE, the Discrete statistical sEmi-empiriCal mODEl, built to accurately simulate *mean* DM halo and galaxy growth and merger histories, for any input SMHM relation. DECODE generates discrete populations of DM haloes and assigns an average mass accretion history to each of them via an input mass function. It subsequently generates their merger trees via an input subhalo distribution function, and assigns to each subhalo the infall redshift and dynamical friction time-scale using statistical density functions that we fit and accurately test against the Millennium simulation. We then populate the (sub)haloes with central/satellite galaxies via diverse and observationally motivated SMHM relations, computed via numerical abundance matching techniques which are very sensitive to the shape of the input SMF. Thanks to its statistical nature, DECODE is flexible, rapid, and not affected by limitation in volume or mass resolution. In this work, we provide useful analytical recipes for the infall redshift distributions of subhaloes of first to third order (Section 2.1), along with the correction to the halo mass function for unmerged and unstripped subhaloes (Appendix B).

We apply DECODE to predict the galaxy–galaxy merger rates, satellite abundances (which can be considered as unmerged satellites), and BCG growth. We also explore how merging pairs can impact on the fraction of ellipticals and mean B/T ratios of local galaxies, by assuming that the former are formed in major mergers, and the latter are shaped by both major/minor mergers and disc instabilities.

Our main results on the galaxy evolution probed via DECODE can be summarized as follows:

(i) DECODE can generate accurate galaxy stellar mass assembly and merger histories starting from an input SMHM relation with only one input parameter, the scatter around the SMHM relation. DECODE can reproduce the average galaxy growth histories of hydrodynamic simulations and SAMs when inputting their SMHM relations.

(ii) Via DECODE, we showed how sensitive many galaxy observables are on the input SMHM relation, and thus on the input SMF, in particular galaxy merger rates, satellite abundances, and BCG growths.

(iii) A SMHM relation implied by an SMF characterized by a larger number of massive galaxies and a normalization significantly decreasing at high redshift, is more suitable to reproduce the correct abundances of satellite galaxies in the local Universe and the stellar mass growth of BCGs at $z < 1$, as well as the combined major merger pair fractions as inferred from GAMA, UDS, VIDEO, and COSMOS.

(iv) Our reference SMHM relation is also able to reproduce the fraction of local elliptical galaxies on the assumption that these are formed from major mergers with $\mu > 0.25$, as often assumed in cosmological SAMs. In other words, the validity of the $\mu > 0.25$ threshold is strongly dependent on the input SMHM relation.

(v) The same SMHM relation is also able to reproduce the mean B/T ratio of local MaNGA galaxies, with a contribution from disc instabilities at stellar masses below $M_* \lesssim 10^{11}\,\mathrm{M}_\odot$.

In conclusion, DECODE is a valuable, complementary tool for probing galaxy evolution and the relevant physical processes involved therein. It can indeed rapidly probe galaxy merger rates, satellites abundances, morphologies, star formation histories for any given input SMHM relation and with minimal input parameters. DECODE will also constitute a very precious instrument for generating robust galaxy mock catalogues for the upcoming large-scale extragalactic surveys such as *Euclid* (e.g. Scaramella et al. 2021; Euclid

Collaboration 2019, 2022) and LSST (e.g. Bridge et al. 2009; Covey et al. 2010; Ptak & LSST Galaxies Collaboration 2011; Gawiser et al. 2013; Riccio et al. 2021).

ACKNOWLEDGEMENTS

We warmly thank Philip J. Grylls for reading the manuscript, and for useful discussion and comments. We thank the referee for a careful reading of the manuscript and for useful inputs. We also thank Sergio Contreras for useful discussions. This work received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860744. HF acknowledges partial support from the 'Torno Subito' programme. MA acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant number NE 2441/1-1). YRG acknowledges the support of the 'Juan de la Cierva Incorporation' fellowship (IC2019-041131-I).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. DECODE is available for download at https://github.com/haofu94/DECODE.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abdurro'uf et al., 2022, ApJS, 259, 35
Angulo R. E., Hilbert S., 2015, MNRAS, 448, 364
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, MNRAS, 399, 983
Aversa R., Lapi A., de Zotti G., Shankar F., Danese L., 2015, ApJ, 810, 74
Ayromlou M., Kauffmann G., Yates R. M., Nelson D., White S. D. M., 2021, MNRAS, 505, 492
Baldry I. K. et al., 2012, MNRAS, 421, 621
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Bellstedt S. et al., 2016, MNRAS, 460, 2862
Berentzen I., Shlosman I., Martinez-Valpuesta I., Heller C. H., 2007, ApJ, 666, 189
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Bernardi M., Meert A., Sheth R. K., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016, MNRAS, 455, 4122
Bernardi M., Meert A., Sheth R. K., Fischer J. L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2017, MNRAS, 467, 2217
Bernardi M., Domínguez Sánchez H., Brownstein J. R., Drory N., Sheth R. K., 2019, MNRAS, 489, 5633
Blanton M. R., Moustakas J., 2009, ARA&A, 47, 159
Blanton M. R. et al., 2017, AJ, 154, 28
Bournaud F., Combes F., 2002, A&A, 392, 83
Bournaud F., Jog C. J., Combes F., 2007, A&A, 476, 1179
Bournaud F., Dekel A., Teyssier R., Cacciato M., Daddi E., Juneau S., Shankar F., 2011, ApJ, 741, L33
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, MNRAS, 407, 2017
Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93
Bridge C., Lotz J. M., Armus L., Ferguson H. C., LSST Galaxies Collaboration, 2009, in American Astronomical Society Meeting Abstracts, vol. 214, Pasadena, p. 407.05
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bundy K. et al., 2015, ApJ, 798, 7
Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, MNRAS, 370, 1651
Cattaneo A., Mamon G. A., Warnick K., Knebe A., 2011, A&A, 533, A5
Cattaneo A. et al., 2017, MNRAS, 471, 1401
Cattaneo A., Koutsouridou I., Tollet E., Devriendt J., Dubois Y., 2020, MNRAS, 497, 279
Chabrier G., 2003, PASP, 115, 763
Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, MNRAS, 460, 3100
Chiosi C., Merlin E., Piovan L., 2012, preprint (arXiv:1206.2532)
Christlein D., Zabludoff A. I., 2004, ApJ, 616, 192
Clauwens B., Schaye J., Franx M., Bower R. G., 2018, MNRAS, 478, 3994
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Collins C. A. et al., 2009, Nature, 458, 603
Conroy C., Wechsler R. H., 2009, ApJ, 696, 620
Conselice C. J., Arnold J., 2009, MNRAS, 397, 208
Contini E., De Lucia G., Villalobos Á., Borgani S., 2014, MNRAS, 437, 3787
Contreras S., Angulo R. E., Zennaro M., 2021, MNRAS, 508, 175
Cooke K. C., Kartaltepe J. S., Tyler K. D., Darvish B., Casey C. M., Le Fèvre O., Salvato M., Scoville N., 2019, ApJ, 881, 150
Covey K. R. et al., 2010, in American Astronomical Society Meeting Abstracts, vol. 215, Washington, p. 401.08
Croton D. J. et al., 2016, ApJS, 222, 22
Davidzon I. et al., 2017, A&A, 605, A70
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Davison T. A., Norris M. A., Pfeffer J. L., Davies J. J., Crain R. A., 2020, MNRAS, 497, 81
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Sari R., Ceverino D., 2009, ApJ, 703, 785
Dekel A., Lapiner S., Dubois Y., 2019, preprint (arXiv:1904.08431)
Devergne T. et al., 2020, A&A, 644, A56
Di Matteo T., Khandai N., DeGraf C., Feng Y., Croft R. A. C., Lopez J., Springel V., 2012, ApJ, 745, L29
Diemer B., 2018, ApJS, 239, 35
Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661
Domínguez Sánchez H., Bernardi M., Nikakhtar F., Margalef-Bentabol B., Sheth R. K., 2020, MNRAS, 495, 2894
Domínguez Sánchez H., Margalef B., Bernardi M., Huertas-Company M., 2022, MNRAS, 509, 4024
Drakos N. E. et al., 2022, ApJ, 926, 194
Drory N. et al., 2015, AJ, 149, 77
Du M., Ho L. C., Zhao D., Shi J., Debattista V. P., Hernquist L., Nelson D., 2019, ApJ, 884, 129
Dvornik A. et al., 2020, A&A, 642, A83
Efstathiou G., Lake G., Negroponte J., 1982, MNRAS, 199, 1069
Engler C. et al., 2021, MNRAS, 500, 3957
Euclid Collaboration, 2019, A&A, 627, A23
Euclid Collaboration, 2022, A&A, 657, A90
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Favole G., Montero-Dorta A. D., Artale M. C., Contreras S., Zehavi I., Xu X., 2022, MNRAS, 509, 1614
Fischer J. L., Bernardi M., Meert A., 2017, MNRAS, 467, 490
Fisher D. B., Drory N., 2010, ApJ, 716, 942
Fontanot F., Macciò A. V., Hirschmann M., De Lucia G., Kannan R., Somerville R. S., Wilman D., 2015, MNRAS, 451, 2968
Gadotti D. A., 2009, MNRAS, 393, 1531
Gawiser E. J. et al., 2013, American Astronomical Society Meeting Abstracts, vol. 221, Long Beach, p. 247.08
Genel S., Bouché N., Naab T., Sternberg A., Genzel R., 2010, ApJ, 719, 229
Genel S., Fall S. M., Hernquist L., Vogelsberger M., Snyder G. F., Rodriguez-Gomez V., Sijacki D., Springel V., 2015, ApJ, 804, L40

Giocoli C., Tormen G., van den Bosch F. C., 2008, MNRAS, 386, 2135
González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, ApJ, 735, L34
Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580
Green S. B., van den Bosch F. C., 2019, MNRAS, 490, 2091
Green S. B., van den Bosch F. C., Jiang F., 2021, MNRAS, 503, 4075
Grylls P. J., Shankar F., Zanisi L., Bernardi M., 2019, MNRAS, 483, 2506
Grylls P. J., Shankar F., Leja J., Menci N., Moster B., Behroozi P., Zanisi L., 2020a, MNRAS, 491, 634
Grylls P. J., Shankar F., Conselice C. J., 2020b, MNRAS, 499, 2265
Gunn J. E. et al., 2006, AJ, 131, 2332
Guo Q., White S. D. M., 2008, MNRAS, 384, 2
Guo Q., White S., 2014, MNRAS, 437, 3228
Guo Q. et al., 2011, MNRAS, 413, 101
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
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. et al., 2009a, MNRAS, 397, 802
Hopkins P. F., Cox T. J., Younger J. D., Hernquist L., 2009b, ApJ, 691, 1168
Hopkins P. F., Younger J. D., Hayward C. C., Narayanan D., Hernquist L., 2010a, MNRAS, 402, 1693
Hopkins P. F. et al., 2010b, ApJ, 715, 202
Huang S., Leauthaud A., Greene J. E., Bundy K., Lin Y.-T., Tanaka M., Miyazaki S., Komiyama Y., 2018, MNRAS, 475, 3348
Izquierdo-Villalba D., Bonoli S., Spinoso D., Rosas-Guevara Y., Henriques B. M. B., Hernández-Monteagudo C., 2019, MNRAS, 488, 609
Jiang F., van den Bosch F. C., 2014, MNRAS, 440, 193
Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848
Kannan R., Macciò A. V., Fontanot F., Moster B. P., Karman W., Somerville R. S., 2015, MNRAS, 452, 4347
Kawinwanichakij L. et al., 2020, ApJ, 892, 7
Khochfar S., Burkert A., 2006, A&A, 445, 403
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Koutsouridou I., Cattaneo A., 2022, MNRAS, Submitted
Kravtsov A. V., 2013, ApJ, 764, L31
Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, ApJ, 609, 35
Kravtsov A. V., Vikhlinin A. A., Meshcheryakov A. V., 2018, Astron. Lett., 44, 8
L'Huillier B., Combes F., Semelin B., 2012, A&A, 544, A68
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Lackner C. N., Cen R., Ostriker J. P., Joung M. R., 2012, MNRAS, 425, 641
Lagos C. d. P., Emsellem E., van de Sande J., Harborne K. E., Cortese L., Davison T., Foster C., Wright R. J., 2022, MNRAS, 509, 4372
Lang P. et al., 2014, ApJ, 788, 11
Lapi A. et al., 2018, ApJ, 857, 22
Laurikainen E., Salo H., Buta R., 2005, MNRAS, 362, 1319
Law D. R. et al., 2015, AJ, 150, 19
Lee J., Yi S. K., 2013, ApJ, 766, 38
Leja J., Speagle J. S., Johnson B. D., Conroy C., van Dokkum P., Franx M., 2020, ApJ, 893, 111
Li Y., Mo H., 2009, preprint (arXiv:0908.0301)
Lidman C. et al., 2012, MNRAS, 427, 550
Lin Y.-T. et al., 2017, ApJ, 851, 139
Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2007, MNRAS, 382, 1394
Marinacci F., Pakmor R., Springel V., 2014, MNRAS, 437, 1750
Marinacci F. et al., 2018, MNRAS, 480, 5113
Marsden C., Shankar F., Bernardi M., Sheth R. K., Fu H., Lapi A., 2022, MNRAS, 510, 5639
McCavana T., Micic M., Lewis G. F., Sinha M., Sharma S., Holley-Bockelmann K., Bland-Hawthorn J., 2012, MNRAS, 424, 361
Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943
Meert A., Vikram V., Bernardi M., 2016, MNRAS, 455, 2440
Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3
Merlin E., Chiosi C., Piovan L., Grassi T., Buonomo U., La Barbera F., 2012, MNRAS, 427, 1530
Monachesi A. et al., 2019, MNRAS, 485, 2589
Monaco P., Fontanot F., Taffoni G., 2007, MNRAS, 375, 1189
Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822
Mundy C. J., Conselice C. J., Duncan K. J., Almaini O., Häußler B., Hartley W. G., 2017, MNRAS, 470, 3507
Murali C., Katz N., Hernquist L., Weinberg D. H., Davé R., 2002, ApJ, 571, 1
Naiman J. P. et al., 2018, MNRAS, 477, 1206
Nelson D. et al., 2018, MNRAS, 475, 624
Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2
O'Leary J. A., Moster B. P., Naab T., Somerville R. S., 2021, MNRAS, 501, 3215
Oklopčić A., Hopkins P. F., Feldmann R., Kereš D., Faucher-Giguère C.-A., Murray N., 2017, MNRAS, 465, 952
Oser L., Ostriker J. P., Naab T., Johansson P. H., Burkert A., 2010, ApJ, 725, 2312
Pillepich A. et al., 2014, MNRAS, 444, 237
Pillepich A. et al., 2018a, MNRAS, 473, 4077
Pillepich A. et al., 2018b, MNRAS, 475, 648
Planck Collaboration VI, 2018, A&A, 641, A6
Planck Collaboration XIII, 2016, A&A, 594, A13
Powell L. C., Bournaud F., Chapon D., Teyssier R., 2013, MNRAS, 434, 1028
Ptak A., LSST Galaxies Collaboration, 2011, in American Astronomical Society Meeting Abstracts, vol. 217, Seattle, p. 252.16
Puech M., Hammer F., Hopkins P. F., Athanassoula E., Flores H., Rodrigues M., Wang J. L., Yang Y. B., 2012, ApJ, 753, 128
Qu Y. et al., 2017, MNRAS, 464, 1659
Ragone-Figueroa C., Granato G. L., Ferraro M. E., Murante G., Biffi V., Borgani S., Planelles S., Rasia E., 2018, MNRAS, 479, 1125
Ricarte A., Natarajan P., 2018, MNRAS, 474, 1995
Riccio G. et al., 2021, A&A, 653, A107
Rodriguez-Gomez V. et al., 2015, MNRAS, 449, 49
Rodriguez-Gomez V. et al., 2016, MNRAS, 458, 2371
Rodríguez-Puebla A., Drory N., Avila-Reese V., 2012, ApJ, 756, 2
Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651
Scaramella R. et al., 2021, preprint (arXiv:2108.01201)
Shankar F., Lapi A., Salucci P., De Zotti G., Danese L., 2006, ApJ, 643, 14
Shankar F., Marulli F., Mathur S., Bernardi M., Bournaud F., 2012, A&A, 540, A23
Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, MNRAS, 428, 109
Shankar F. et al., 2014, MNRAS, 439, 3189
Shankar F. et al., 2015, ApJ, 802, 73
Shankar F. et al., 2018, MNRAS, 475, 2878
Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978
Simard L., Trevor Mendel J., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11
Simons R. C. et al., 2019, ApJ, 874, 59
Smee S. A. et al., 2013, AJ, 146, 32
Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, ApJ, 833, 109
Springel V., 2010, MNRAS, 401, 791
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nature, 435, 629
Springel V. et al., 2018, MNRAS, 475, 676
Stewart K. R., Bullock J. S., Wechsler R. H., Maller A. H., 2009a, ApJ, 702, 307

Stewart K. R., Bullock J. S., Barton E. J., Wechsler R. H., 2009b, ApJ, 702, 1005
Stott J. P. et al., 2010, ApJ, 718, 23
Stringer M. J., Shankar F., Novak G. S., Huertas-Company M., Combes F., Moster B. P., 2014, MNRAS, 441, 1570
Tacchella S. et al., 2019, MNRAS, 487, 5416
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tomczak A. R. et al., 2014, ApJ, 783, 85
Tonini C., Lapi A., Shankar F., Salucci P., 2006, ApJ, 638, L13
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van den Bosch F. C., 2017, MNRAS, 468, 885
van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029
van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padmanabhan N., 2014, MNRAS, 445, 1713
van den Bosch F. C., Jiang F., Campbell D., Behroozi P., 2016, MNRAS, 455, 158
Wang L., Jing Y. P., 2010, MNRAS, 402, 1796
Weinberger R. et al., 2017, MNRAS, 465, 3291
Whiley I. M. et al., 2008, MNRAS, 387, 1253
Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., Chu Y., 2004, MNRAS, 350, 1153
Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153
Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41
Zanisi L. et al., 2020, MNRAS, 492, 1671
Zanisi L., Shankar F., Bernardi M., Mei S., Huertas-Company M., 2021a, MNRAS, 505, L84
Zanisi L. et al., 2021b, MNRAS, 505, 4555
Zavala J., Avila-Reese V., Firmani C., Boylan-Kolchin M., 2012, MNRAS, 427, 1503
Zhang Y. et al., 2017, preprint (arXiv:1710.05908)

APPENDIX A: TESTING THE SELF-CONSISTENCY OF DECODE

We develop a variant of our model, which we refer to as *weighted* method, to check the self-consistency of our approach in DECODE (referred to as *discrete* method) in reproducing the average properties of dark matter haloes and galaxy evolution. In the main body of the paper, we make use of the discrete method only. We describe in this section the details of the weighted method. We reiterate here that the weighted method is more difficult to generalize to all model variants, for example when multiple galaxy properties are included in the evolution, which makes the discrete method more flexible.

First of all, we focus our attention on the mergers history of DM haloes. To compute the latter, in the weighted method we employ the recipe from section 3.1.3 of G19, where we interpret the difference of the SHMF between redshifts z and $z + \mathrm{d}z$ as the weight (or probability) of the infalling subhaloes in each mass bin. In other words the weight, at given redshift step and mass bin, is the fractional average number of subhaloes of that mass which cross the virial radius of the parent halo at that redshift. The comparison between the weighted and discrete methods for one parent halo mass bin is shown in the upper panel of Fig. A1, where the total mass assembly history of van den Bosch et al. (2014) is also shown for completeness. Despite the fact that in the discrete method we perform our analysis on the assumption of identical mean accretion for all haloes competing to the same bin of host halo mass at $z = 0$, the resulting mean contribution from subhalo mergers in the discrete method appears to be in very good agreement with the one computed from the weighted method, further supporting the validity of our discrete approach.

Similarly to the host dark matter haloes, the bottom panel of Fig. A1 compares the merger contributions to the central galaxy

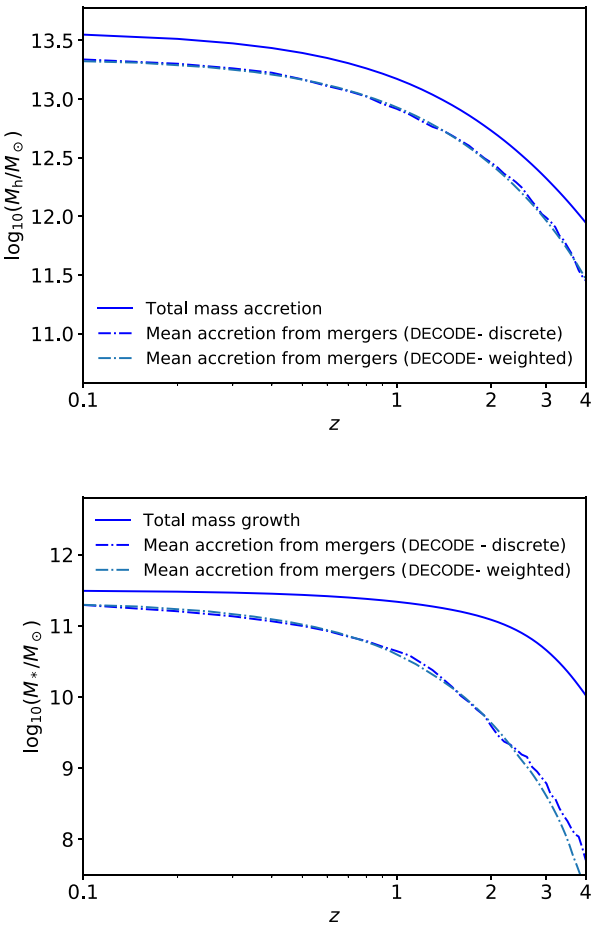

Figure A1. Upper panel: total halo mass accretion history (solid line) along with the comparison between the mergers history of from the discrete and weighted methods (dashed lines). Lower panel: same as upper panel but for galaxy stellar mass.

growth computed via the discrete and weighted methods, as labelled, showing again very good agreement between the two methods. We note that this agreement is, as expected, independent of the choice of the input SMHM relation or dynamical friction time-scales, as long as the same parameters are adopted in both methods. We specify once again that, as already noted by G20, each merger tree generated in the *discrete* DECODE can show sometimes a merger history that goes beyond the total stellar mass growth of the galaxy, which might seem not physical. This is a direct consequence of the fact that each merger tree in DECODE is a stochastic realization of the mass functions and probability distributions used as input. However, we test that in the SMHM models that we adopt in this work, the average merger history always lives below the total mass growth and is also fully consistent with what the *weighted* method predicts, as already shown in Fig. A1. We show in the left-hand panel of Fig. A2 the merger history for our fiducial Model 2 for a galaxy mass bin of $\sim 10^{11.5}\,\mathrm{M}_\odot$, where we see that all the single merger histories lie below the total mean mass assembly of the galaxy. On the other hand, in the right-hand panel we show the same results but for Model 2a, where we can clearly see the impact of a lower high-mass end in the SMF leading to a much higher number of mergers, and in many cases not physical as it goes beyond the total growth. The effect of generating an unphysical merger history that, on average, is larger than the total mean stellar mass growth, could be, at least in part,

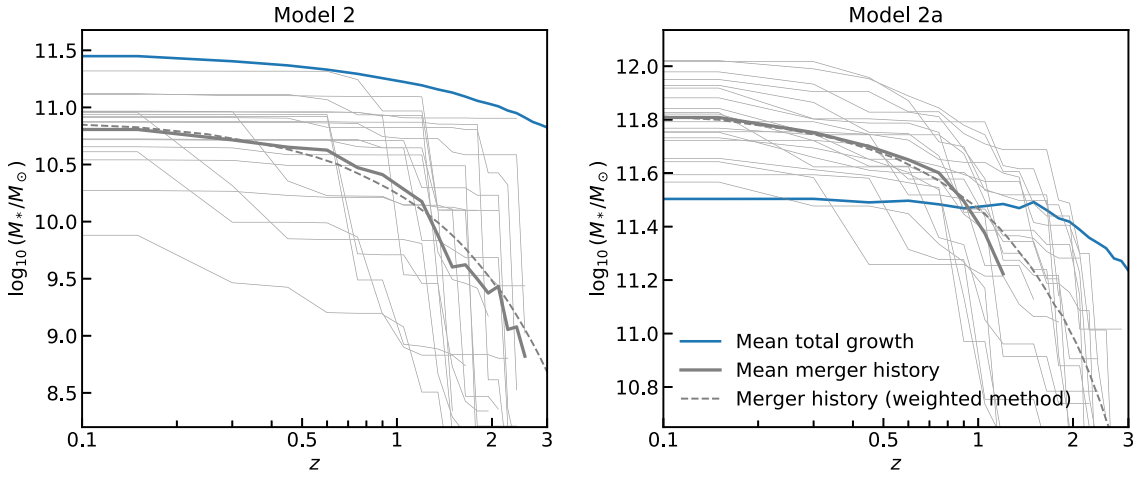

Figure A2. Merger histories of a sample of galaxies with stellar mass at $z = 0$ of $\sim 10^{11.5}\,\mathrm{M}_{\odot}$, for Models 2 and 2a, as labelled by the panels. The solid blue and grey lines how the average total stellar mass growth and merger history, respectively. The dashed grey lines show the average merger history from the weighted model.

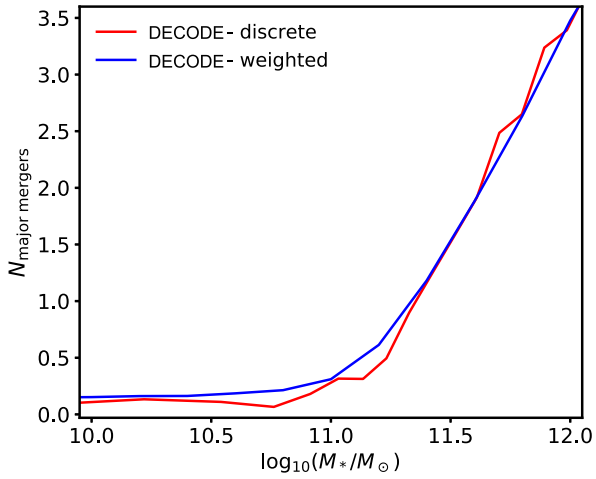

Figure A3. Number of major mergers predicted by the discrete version of DECODE compared to that computed with the weighted method.

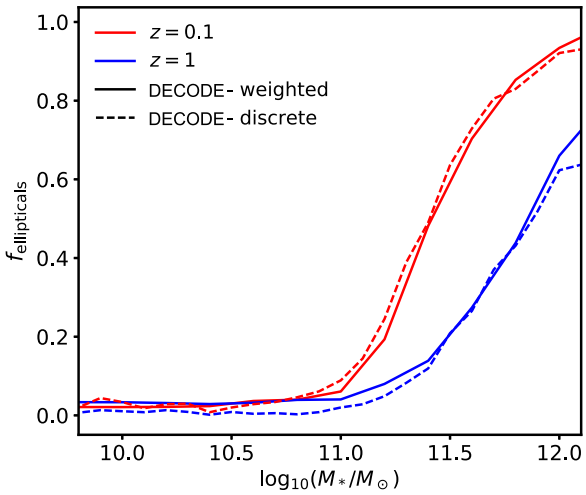

Figure A4. Fraction of ellipticals predicted by the discrete version of DECODE compared to those computed with the weighted method, for redshifts 0.1 and 1.

be alleviated by including strong stellar stripping (see e.g. Cattaneo et al. 2011; Smith et al. 2016) and/or stellar mass loss in mergers (e.g. Moster et al. 2018). We will further explore these interesting variants to the model in future work, in combination with the amount of associated star formation our stellar mass accretion tracks predict.

We also provide the number of major mergers, implied fraction of ellipticals, and mean B/T ratios as predicted from the two methods in Figs A3, A4, and A5, respectively. The selection of the major mergers and the ellipticals in the discrete method is already described in Section 4.4. In the weighted method, the number of major mergers is computed by directly integrating the merging satellites SMF at each redshift over the range of mass $M_{*,\mathrm{sat}}/M_{*,\mathrm{cen}} > \mu$, with μ being the major mergers mass ratio threshold. Concerning the fraction of ellipticals, we also label the galaxies that had at least one major merger as ellipticals, similarly as we do in the discrete method. To this purpose, we initialize the fraction of ellipticals at redshift $z_{\mathrm{ini}} = 4$ to 0, assuming that all galaxies are disc-like at that time. From that epoch, we proceed forward in time and at each redshift we analytically compute the probability of a galaxy to have had at least one major merger, which we interpret as the fraction of ellipticals itself, according to the following formula

$$\mathcal{P}_{1\mathrm{MM}} = 1 - \mathcal{P}_{\mathrm{MM}}^{W}, \tag{A1}$$

where $\mathcal{P}_{1\mathrm{MM}}$ is the probability of having at least one major merger, $\mathcal{P}_{\mathrm{MM}}$ is the probability of a generic merger to be a major one, and the exponent W is the weight integrated over the major mergers stellar mass range. At each time-step, we update the fraction of spirals and ellipticals according to equation (A1). Finally, we also provide the comparison of the B/T ratio for Model BT2 (Section 4.5), which we compute in the weighted model as the cumulative sum of the probabilities of having at least one major merger in each time bin. It is clear from the results reported in Figs A3–A5 that the both methods provide extremely consistent predictions on the aforementioned quantities, further validating the use of the discrete method to predict mean galactic properties.

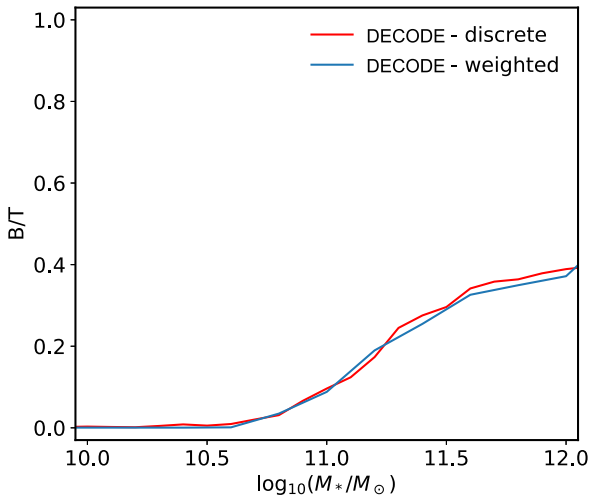

Figure A5. Bulge-to-total ratio predicted by the discrete version of DECODE compared to that computed with the weighted method for Model BT2.

APPENDIX B: CORRECTION TO THE HALO MASS FUNCTION

The abundance matching procedure of equation (8) between the galaxy SMF and the dark matter HMF also includes galactic satellites in the former, and dark matter subhaloes in the latter. However, the HMF is usually given as a fit to the number densities of only the parent haloes existing in any given simulation snapshot, and thus it must be corrected by the abundances of the unstripped subhaloes surviving at any given epoch. To determine this correction, we first compute the number densities of unstripped and surviving subhaloes in DECODE at any redshift of interest, and then fit it following, for convenience and for ease of comparison, the same analytical expression adopted by Behroozi et al. (2013)

$$\frac{\phi_{\rm satellites}(M_{\rm h})}{\phi_{\rm centrals}(M_{\rm h})} \sim C(a) \log\left(\frac{M_{\rm cutoff}(a)}{M_{\rm h}}\right), \tag{B1}$$

where $a = 1/(1+z)$ is the scale factor. Our DECODE Monte Carlo simulations provide the following fitting formulae for the two free parameters in equation (B1)

$$\log(C(a)) = -2.42 + 11.68a - 28.88a^2 + 29.33a^3 - 10.56a^4, \tag{B2}$$

$$\log(M_{\rm cutoff}(a)) = 10.94 + 8.34a - 0.36a^2 - 5.08a^3 + 0.75a^4, \tag{B3}$$

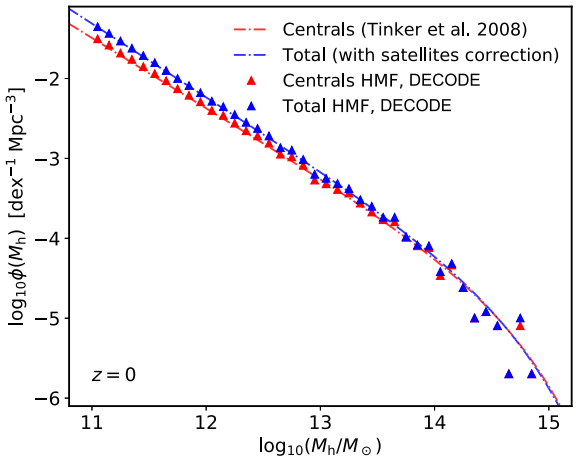

Figure B1. Halo mass function for parent dark matter haloes of Tinker et al. (2008) (red dashed–dotted line) and the total HMF obtained by applying the correction with satellites (blue dashed–dotted line), compared to the mass functions calculated from DECODE (triangles).

We show in Fig. B1 the Tinker et al. (2008) HMF for centrals along with the aforementioned correction for including unstripped, surviving (i.e. unmerged) satellites, and we compare them with the numerical mass functions calculated from DECODE.

APPENDIX C: HALO AND STELLAR MASS GROWTHS

As already described in Section 3.1, we assign the parent halo with a mean mass assembly history. For the details of how we numerically calculate the latter, we readdress the interested reader to the work of van den Bosch et al. (2014). We show the mean DM accretion for four different values of stellar mass bin (as labelled in the legends) in the left-hand panels of Figs C1 and C2, where we compare with the data from the TNG simulation and EMERGE, respectively. This is a cross-check that the predictions of the TNG simulation and EMERGE are consistent with the mean assembly history that we are employing.

Similar considerations are valid for the total mean galaxy stellar mass growths. We show in the right-hand panels of Figs C1 and C2 the results predicted by our model DECODE using as input the SMHM computed from the two simulations, and we compare with the data from the simulations themselves. The stellar mass growth histories for the four masses shown comes out to be consistent within ~ 1 dex with TNG and EMERGE.

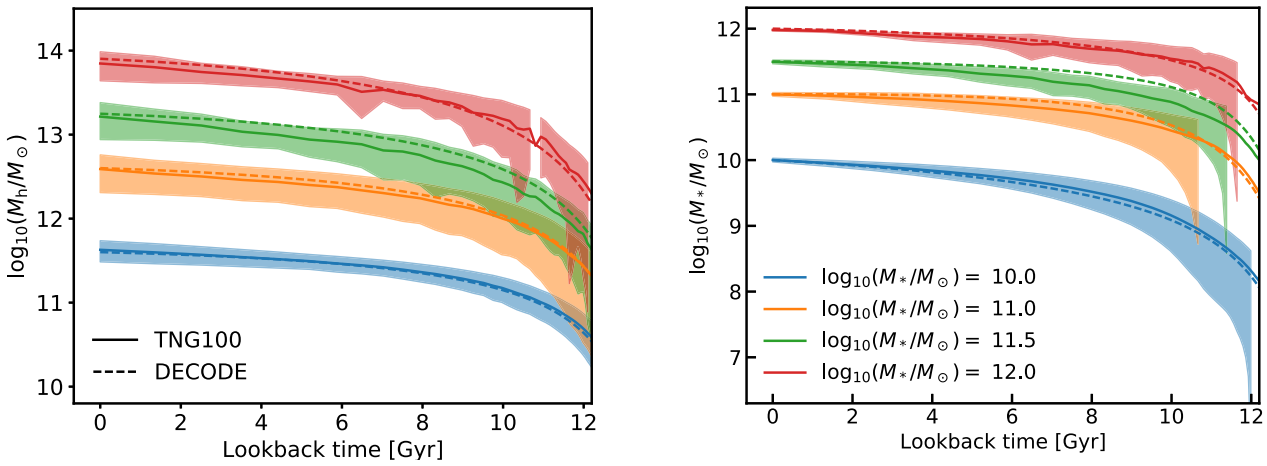

Figure C1. Left-hand panel: halo mass assembly history from the TNG100-1 simulation, for four stellar mass bins at $z = 0$. Solid lines and shaded areas show the mean and 1σ error from the simulation, and the dashed lines show the accretion history from (van den Bosch et al. 2014) that we adopt in this work. Right-hand panel: total stellar mass growth for the same four mass bins. The dashed lines show the mean galaxy growth computed using the SMHM relation of TNG as input in DECODE, while the solid lines show the results extracted directly from the TNG data base.

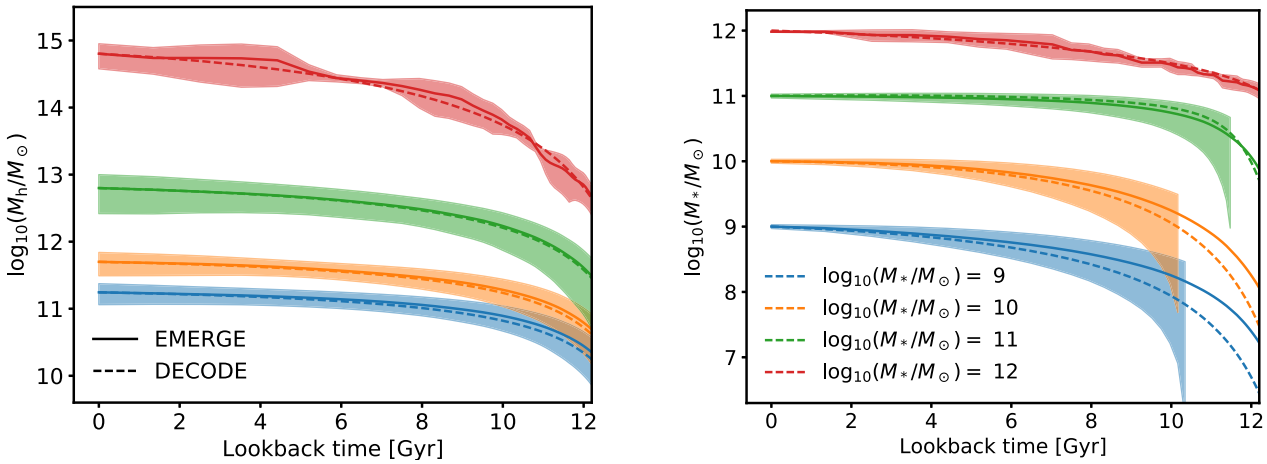

Figure C2. Same as Fig. C1 but for EMERGE.

APPENDIX D: BULGE-TO-TOTAL RATIOS MODELLING

We show in Fig. D1 a comparison of the B/T ratios from different observational data sets. In particular, we compare the MaNGA data (black error bars), described in Section 2.4 and used as a reference for the models in this work, with the *SDSS* data from Mendel et al. (2014) who selected a subsample of the Simard et al. (2011) catalogue. We also show the B/T ratio of *SDSS* that we have computed directly from the Simard et al. (2011) catalogue (grey error bars), as well as the predictions of Models 2 from this work (green-dotted and blue-dashed lines). Interestingly, our results for *SDSS* are not consistent with those from Mendel et al. (2014). Nevertheless, our results discussed in the main text are still valid, irrespective of the exact data set chosen for computing mean B/T ratios. All the three observational B/T ratios show in fact that models based only on mergers, such as our BT1 and BT2 described in Section 4.5, are not sufficient to reproduce the measure B/T ratios, at least at low masses. On the other hand, models that include also disc instabilities perform much better in reproducing the observational data. In summary, all observational B/T data suggest that at low masses some level of disc instabilities is still expected in addition to mergers in order to well describe the evolution of galactic bulges.

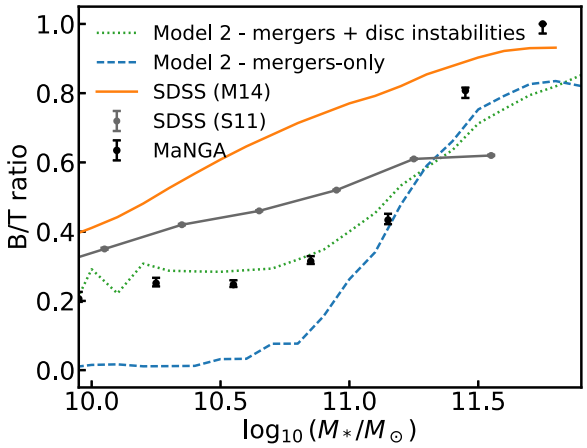

Figure D1. Bulge-to-total ratios as a function of stellar mass from this work compared to different observational samples. The green dotted and blue dashed lines show the Model 2 predictions for the mergers + disc instabilities and mergers-only toy models, respectively. The orange line shows the mean B/T of *SDSS* using the sample of Mendel et al. (2014), as shown in Devergne et al. (2020). The grey error bars show the *SDSS* B/T computed using the Simard et al. (2011) catalogue and the black error bars the MaNGA data.

APPENDIX E: SUBHALO ABUNDANCE MATCHING

For completeness, we compare our SMHM relations computed from direct abundance matching between the SMF and the HMF, with the SMHM relation derived from the stellar mass-peak velocity (SMPV) relation (e.g. Guo & White 2014; Chaves-Montero et al. 2016; Contreras et al. 2021; Favole et al. 2022). To this purpose, we first compute the SMPV relation using equation (8), where we input the SMFs of Models 1 and 2, as described in Section 4.1, and we replace the HMF with the peak velocity function. We extract the peak velocity function from the MultiDark simulation (Klypin et al. 2016) at different redshift snapshots. Once the SMPV relation is computed, we calculate the implied SMHM relation and its dispersion at fixed halo mass by using the halo masses competing to each $V_{\rm peak}$ in the simulation.

We show the resulting SMHM relation in Figs E1 and E2 at different redshifts (solid blue lines and blue regions), compared with the SMHM relations from DECODE's Models 1 and 2, respectively. We see that the relations obtained via the SMPV relation match very well with the SMHM relations from Models 1 and 2. This agreement, on one hand, provides a further validation of our unstripped, surviving subhalo abundance matching technique presented in Section 3.5 and, on the other hand, further highlights that the systematics in the shape and/or redshift evolution of the observed SMF have a strong impact on the resulting mapping between stellar mass and halo mass.

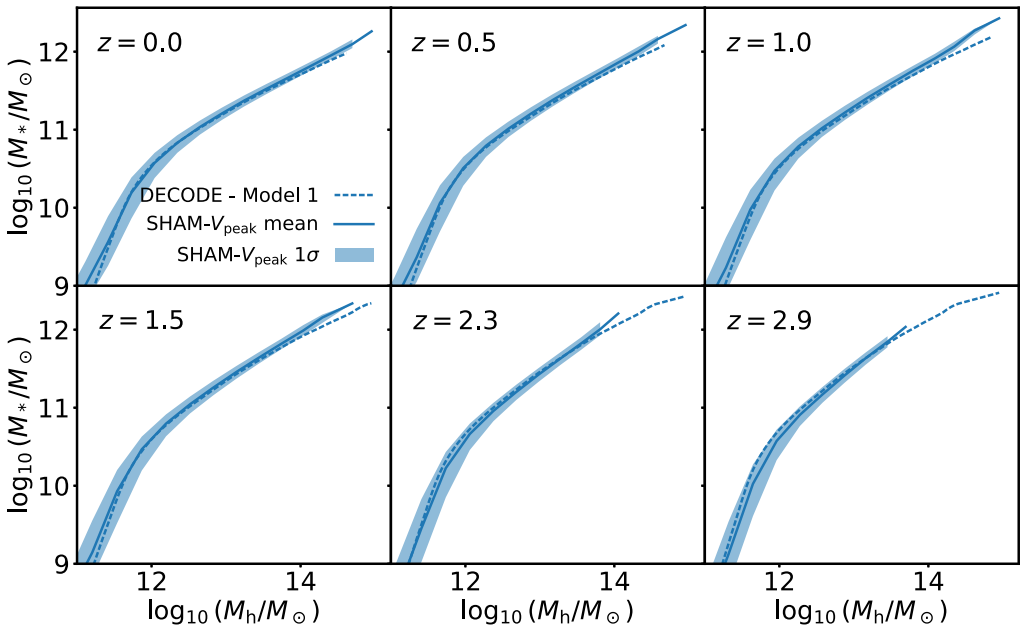

Figure E1. SMHM relation computed via the SHAM technique using the DM $V_{\rm peak}$ data from the MultiDark compared to DECODE at different redshifts. The dashed lines show Model 1 SMHM relation from DECODE, at different redshifts, as described in Section 4.1. The solid lines and shaded areas show the SMHM relation computed via SHAM with the velocity function using the Model 1 SMF as input and the 1σ dispersion.

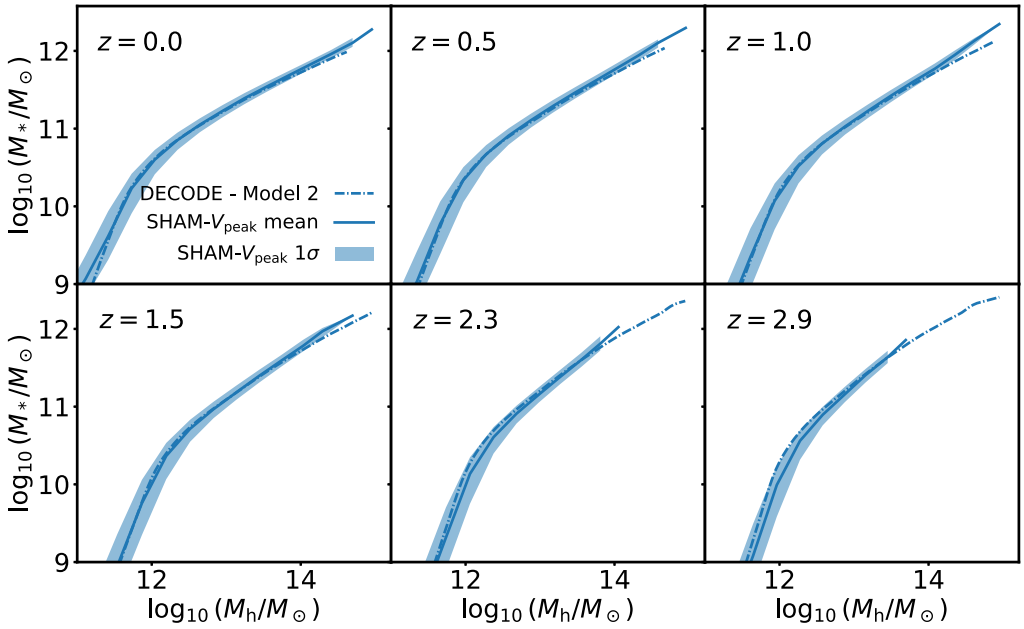

Figure E2. Same as Fig. E1 but for Model 2.

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