

Modelling the galaxy–halo connection with semi-recurrent neural networks

Harry George Chittenden   and Rita Tojeiro 

School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, Scotland, UK

Accepted 2022 November 23. Received 2022 October 28; in original form 2022 March 24

ABSTRACT

We present an artificial neural network design in which past and present-day properties of dark matter haloes and their local environment are used to predict time-resolved star formation histories and stellar metallicity histories of central and satellite galaxies. Using data from the IllustrisTNG simulations, we train a TENSORFLOW-based neural network with two inputs: a standard layer with static properties of the dark matter halo, such as halo mass and starting time; and a recurrent layer with variables such as overdensity and halo mass accretion rate, evaluated at multiple time steps from $0 \leq z \lesssim 20$. The model successfully reproduces key features of the galaxy halo connection, such as the stellar-to-halo mass relation, downsizing, and colour bimodality, for both central and satellite galaxies. We identify mass accretion history as crucial in determining the geometry of the star formation history and trends with halo mass such as downsizing, while environmental variables are important indicators of chemical enrichment. We use these outputs to compute optical spectral energy distributions, and find that they are well matched to the equivalent results in IllustrisTNG, recovering observational statistics such as colour bimodality and mass–magnitude diagrams.

Key words: galaxies: evolution – galaxies: formation – galaxies: haloes – galaxies: star formation.

1 INTRODUCTION

The consensual basis for galaxy formation is that galaxies are formed from the contraction of the baryonic gas bound gravitationally to a dark matter halo (Wechsler & Tinker 2018). The aspects and future of a galaxy’s development are therefore determined by the properties of the halo and its surroundings. While the relationship between certain properties of the halo and galaxy may not necessarily be understood, some degree of correlation between the two is to be expected. Specifically, the rate of change in a host halo’s mass, the rate of merger events, and similar factors are expected to influence the halo’s likelihood of hosting galaxies (Bose et al. 2019), and the star formation rates, metallicities, and other intrinsic properties of the galaxies themselves (Wechsler & Tinker 2018). This relationship is commonly referred to as the galaxy–halo connection (GHC).

The GHC holds a number of important scientific questions. In terms of halo and stellar mass growth, the rate at which the halo assembles most of its mass, through accretion and through major mergers, will affect the time at which galaxies form the most stars, become quenched, and cluster together (Croton, Gao & White 2007; Hani et al. 2020; Cui et al. 2021; Montero-Dorta et al. 2021). In terms of local environment, galaxies in proximity to cosmic filaments are more prone to accreting metal-rich gas (Peng & Maiolino 2013; Donnan, Tojeiro & Kraljic 2022), and in dense regions of space are tidally quenched by more massive haloes, producing noticeably different luminosity and mass functions in their respective environments (Ayromlou et al. 2021; Hellwing et al. 2021; Lu et al. 2021). This relationship between galaxies and their haloes and environments is

nonetheless highly complex, and the causal interplay between haloes and galaxies, in different cosmic epochs and regimes of halo mass, local density, and interaction rates, remains poorly understood.

Several large volume cosmological simulations have, in recent years, synthesized a catalogue of haloes and galaxies (Schaye et al. 2014; Pillepich et al. 2017a; Davé et al. 2019) based on semi-analytic or hydrodynamical modelling of the astrophysical processes which regulate baryonic evolution in fine detail (for reviews of the simulation models see Somerville & Davé 2015; Vogelsberger et al. 2020). However, the complexity of the baryonic models make it computationally impractical to encapsulate the full extent of the GHC, and consequently the volumes of such simulations are limited to $(300 \text{ Mpc})^3$ (Nelson et al. 2019), while the larger simulations inevitably limit their mass resolution. The effect of this compromise is problematic in scientific applications, as the rarest of objects, such as high mass clusters, are significantly limited in number and resolution, while their unresolved substructures may contribute significantly to the evolution of causally connected galaxies.

A pure dark matter simulation relies solely on collisionless gravitational and cosmological dynamics, and so can be run on larger volumes without restricting the mass resolution significantly (Wechsler & Tinker 2018; Vogelsberger et al. 2020). The halo mass functions, correlation functions, and other statistics will appear similar to the equivalent baryonic simulation, given that the dark matter component constitutes the majority of the mass of a galaxy–halo system. A machine learning algorithm which can utilize the properties of a dark matter formation history to emulate their corresponding galaxies can therefore populate a large dark matter simulation with evolving galaxies across all times, in a fraction of the time taken to compute a hydrodynamical simulation of this level of complexity; while the connections of the GHC which are learned

* E-mail: hgc4@st-andrews.ac.uk

by this model can offer an explanation of the various galaxy formation mechanisms that take place in the history of the simulation. If applied to a high fidelity N -body simulation, it will result in a vast galaxy data set with which to test these connections in further detail.

Several studies have used machine learning methods to ascribe galaxy properties to dark matter haloes, including physical properties such as stellar and H I mass (Agarwal, Davé & Bassett 2018; Jo & Kim 2019; Lovell et al. 2021), observational properties such as band magnitudes and galaxy clustering (Wadekar et al. 2020; Xu et al. 2021; McGibbon & Khochfar 2022), and their dependence on time and cosmological models (Agarwal et al. 2018; Villaescusa-Navarro et al. 2021; Xu et al. 2021). This study aims to model the complete star formation history (SFH) and metallicity history (ZH) of central and satellite galaxies from the historical evolution of their dark matter haloes and local dark matter environment by means of a semi-recurrent neural network algorithm. From these properties, we then self-consistently predict observables such as optical spectra and broad-band colours. We design two such neural networks: one intended to simulate central galaxies, the other satellite galaxies.

Each network contains two input layers: one is a simple dense layer entailing time-independent halo properties, the second is a simple recurrent layer containing properties defined at multiple time steps. A recurrent layer contains an activation sequence which runs between successive data points, thereby establishing a one-way causal connection between them. For our objective of fashioning an evolutionary history of the dark matter halo and environment, whose temporal properties may play a role in governing the developmental aspects of galaxy formation, this framework proves valuable in enforcing causality and improving the precision of our results.

The motive for developing a semi-recurrent neural network which predicts the star formation rate and stellar mass weighted metallicity over cosmic time is to form our understanding of the dark matter properties which govern galaxy evolution, while determining how accurately a population of galaxies and their historical characteristics can be deduced exclusively from dark matter. Following the development of two neural networks for predicting central and satellite galaxy formation histories, these predictions are used to construct model spectral energy distributions using the flexible stellar population synthesis (FSPS) code (Conroy, Gunn & White 2009; Conroy & Gunn 2010). We recover observational characteristics, such as the bimodal relationship between stellar mass and colour (Baldry et al. 2006; Cui et al. 2021).

The neural networks are trained on data from the Illustris: The Next Generation (TNG) hydrodynamical simulation (Nelson et al. 2017, 2019; Pillepich et al. 2017a; Marinacci et al. 2018; Springel et al. 2017; Naiman et al. 2018), which consists of 100 ‘snapshots’ in time, over a redshift domain $0 \leq z \leq 20$, with a median time difference of 146 Myr. We access the collaboration’s public data repository,¹ namely the TNG100-1 and TNG300-1 simulations: the highest resolution simulations with volumes of respective cubic side length 100 and 300 Mpc. Combining these data sets provides a large, diverse sample of haloes of assorted mass and environment, gaining as impartial a training data set as possible. The TNG simulations are based on the Planck-2015 Λ CDM cosmological model ($\Omega_m = 0.3089$, $\Omega_\Lambda = 0.6911$, $\Omega_b = 0.0486$, $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$); assumed throughout this work. To ensure that halo and galaxy formation histories are well resolved at all redshifts, we impose a lower limit on the final stellar mass of $10^9 M_\odot$.

In this paper, we outline the properties, calculation, and justification of dark matter quantities in Section 2, and baryonic quantities in Section 3. Aspects of the neural network design and data pre-processing are discussed in Section 4. We evaluate aspects of the baryonic predictions and their derived observables in Section 5, and the importance of different features in the model in Section 6. The implications of the model are reviewed in Section 7 before summarizing our findings in Section 8.

2 NEURAL NETWORK FEATURES

In this section, we discuss the implementation of the various features of the neural network, including how they were calculated, normalized, and expected to benefit the predictability of results. A full summary of the quantities used in this network with details of their pre-processing, implementation, and testing is given in Table 1.

2.1 Time-dependent variables

In the TNG simulations, each halo is assigned a merger tree, which contains all progenitor subhaloes of the target halo at all prior snapshots of the simulation (Jiang & van den Bosch 2014; Nelson et al. 2019). A ‘branch’ is defined as a particular path taken by any given subhalo to the present halo, and is defined at all snapshots from the time of the subhalo’s formation to the target snapshot. For all objects, we utilize the main progenitor branch (MPB), which is defined as the branch describing the history of the subhalo of the highest mass (Nelson et al. 2019). All properties, such as stellar and dark mass components, metallicities, and angular momenta, whether applicable to the Friends-of-Friends (FoF) or SubFind object, are derived from the MPB.

Most variables in our model are normalized by Gaussian Quantile Transformation (see Section 4.1); however for temporal variables, there exist two normalization methods, which we term scalar and vector normalization (see Section 4.2). Scalar normalization consists of unique transformations for each time step, while vector normalization applies a single transformation to all data regardless of their time. As discussed in the pre-processing section, and the discussion of temporal quantities where relevant, each of these normalization methods add certain benefits to the treatment of different quantities.

2.1.1 Halo mass accretion history

The dark matter halo’s mass accretion rate is defined using the sum of masses of dark matter particles bound to the FoF group. We convert this to an accretion rate by finite differencing with respect to the time t_i at each snapshot:

$$\dot{M}_h(t_i) = \frac{M_h(t_i) - M_h(t_{i-1})}{t_i - t_{i-1}} \quad (1)$$

The subfind subhalo’s mass formation rate is defined equivalently:

$$\dot{m}_h(t_i) = \frac{m_h(t_i) - m_h(t_{i-1})}{t_i - t_{i-1}} \quad (2)$$

Where necessary, FoF (halo) and Subfind (subhalo) masses will be respectively denoted M_h and m_h to avoid discrepancy. The use of an accretion rate effectively directs the neural network to recognize its integral over any time interval; the full integral of course corresponding to the halo’s final mass (see Section 2.2.1). For the sake of treatment of accretion rate as a universal parameter to be integrated over intervals of time, accretion rates are vector normalized.

¹<http://www.tng-project.org/data/>

Table 1. A summary of the quantities used in both neural networks, grouped by layer and ordered by their placement in said layer. This entails the units of each quantity, and indicates which networks utilize them and how they are normalized. The section column indicates which section of this paper discusses this quantity. The shuffle group (final column) indicates which variables are simultaneously scrambled when testing for feature importance (see Section 6.1).

	Quantity	Notation	Network data			GQT	Logarithmic	Shuffle	
			Units	Section	Network				
Temporal features	Halo mass accretion rate	\dot{M}_h	$M_\odot \text{ Gyr}^{-1}$	2.1.1	Both	Vector	False	1	
	Subhalo mass accretion rate	\dot{m}_h	$M_\odot \text{ Gyr}^{-1}$	2.1.1	Satellite	Vector	False	1a	
	1Mpc Overdensity	δ_1	–	2.1.2	Both	Scalar	False	2	
	3Mpc Overdensity	δ_3	–	2.1.2	Central	Scalar	False	2	
	5Mpc Overdensity	δ_5	–	2.1.2	Central	Scalar	False	2	
	Circular velocity (proxy)	\bar{v}_{vir}	$\sqrt{(M_\odot/\text{Mpc})}$	2.1.4	Both	Vector	False	3	
	Dark matter half-mass radius	$R_{\frac{1}{2}}$	Mpc	2.1.4	Both	Vector	False	3	
	1Mpc Radial skew	μ_3	–	2.1.3	Satellite	Vector	False	4	
	3Mpc Radial skew	μ_3	–	2.1.3	Central	Vector	False	4	
	Distance to closest subhalo	d_{μ_3}	Mpc	2.1.3	Both	Vector	False	4	
Non-temporal features	Specific halo mass accretion gradient	β (c) β_{halo} (s)	$\log \text{ Gyr}^{-2}$	2.2.2	Both	None	False	1	
	Specific subhalo mass accretion gradient	β_{sub}	$\log \text{ Gyr}^{-2}$	2.2.2	Satellite	None	False	1a	
	Scaled infall time	a_{infall}	–	2.2.5	Satellite	None	False	1a, 2, 4	
	Scaled formation time	a_{max}	–	2.2.5	Satellite	None	False	1a	
	Infall mass ratio	μ	–	2.2.5	Satellite	None	True	1, 1a	
	Infall velocity	v_{rel}	km s^{-1}	2.2.5	Satellite	None	True	2	
	$z = 0$ Cosmic web distances	d_{CW}	kpc	2.2.3	Central	Scalar	True	2	
	Starting time	t_{start}	Gyr	2.2.4	Both	Scalar	False	All	
	$z = 0$ Halo mass	M_h	M_\odot	2.2.1	Both	Scalar	True	1	
	Maximum absolute halo accretion rate	$ \dot{M}_h $	$M_\odot \text{ Gyr}^{-1}$	2.2.1	Both	Scalar	True	1	
	$z = 0$ Subhalo mass	m_h	M_\odot	2.2.1	Satellite	Scalar	True	1a	
	Maximum absolute subhalo accretion rate	$ \dot{m}_h $	$M_\odot \text{ Gyr}^{-1}$	2.2.1	Satellite	Scalar	True	1a	
	Targets	Star formation history	\mathcal{S}	$M_\odot \text{ Gyr}^{-1}$	3.1	Both	Vector	False	N/A
		Metallicity history	\mathcal{Z}	Z_\odot	3.2	Both	Vector	False	N/A
		$z = 0$ Stellar metallicity	Z	Z_\odot	3.2	Both	Scalar	True	N/A
		$z = 0$ Stellar mass	M_s	M_\odot	3.1	Both	Scalar	True	N/A
		Mass weighted age	MWA	Gyr	3.1	Both	Scalar	False	N/A

2.1.2 Overdensity history

As a measure of the dark matter environment in proximity to the target halo, we compute and apply the mass-weighted density of haloes relative to the simulation mean: the halo overdensity. We compute this for each individual snapshot in order to quantify the environmental history of our halo data set.

For a given object, the local density is defined as the sum of masses of haloes within an arbitrary volume centred on the target’s centre of mass, divided by said volume (Agarwal et al. 2018; Bose et al. 2019), where TNG subhaloes contribute to the calculation if their centres of mass lie within this volume. The target’s local dark matter density, and thus the overdensity, is therefore a function of this volume.

We shall denote the overdensities calculated using a radius of x Mpc as δ_x , i.e. δ_1 for 1 Mpc. These overdensities are used as features in the recurrent input. For central subhaloes, we compute overdensities δ_1 , δ_3 , and δ_5 , as each of these will capture environmental structures on different scales. For satellite subhaloes, we are interested in smaller scale overdensities as a measure of the state of the halo environment. Agarwal et al. (2018) deem 200 kpc to be a useful overdensity radius for constraining zero-redshift baryonic properties, such as stellar mass and neutral hydrogen fraction. However, through investigating the history of multiple kpc-scale overdensities, while their physical values inevitably differ, as a function of time they are geometrically congruous. Smaller overdensities, however, are more susceptible to Poisson noise. Thus, the 1 Mpc parsec overdensity (δ_1) is used as the single overdensity measure for satellites.

We utilize the Grid Search In Python (GRISPY) (Chalela et al. 2021) package to compute overdensities. GRISPY is a regular grid processing and nearest neighbour searcher, specifically designed to

handle periodic boundary conditions, as is the case with the TNG simulations. The ‘bubble neighbours’ query returns the set of objects within a specified distance from the reference coordinate; in our case, haloes within x Mpc of the centre of mass of each targeted halo.

Unlike the halo mass accretion history, the overdensities are scalar normalized, in spite of being defined at the same time steps. The structure of the local environment is expected to vary to such an extent that the differences in overdensities at successive times are not meaningful, and significantly large that a common quantile transformation can fail to distinguish subsets of large and small value. In fact, vector normalization of the overdensity values have a strong adverse effect on the quality of predictions. Instead we prioritize the instantaneous environment over its vectorized history, as this is immune to temporal variation in cosmic structure and represents the local density field specific to each halo.

2.1.3 Radial dark matter skew

A third temporal parameter is the mass-weighted radial skew of the distribution of dark matter subhaloes, going radially outwards from the centre of the target subhalo. For a distribution of variables x_j with weights w_j , their statistical moments are given:

$$\mu_1 = \frac{\sum_{j=1}^N w_j x_j}{\sum_{j=1}^N w_j} \quad (3)$$

$$\mu_2 = \frac{\sum_{j=1}^N w_j (x_j - \mu_1)^2}{\sum_{j=1}^N w_j} \quad (4)$$

$$\mu_n = \frac{\sum_{j=1}^N w_j \left(\frac{x_j - \mu_1}{\sqrt{\mu_2}} \right)^n}{\sum_{j=1}^N w_j}, \quad \forall n \geq 3 \quad (5)$$

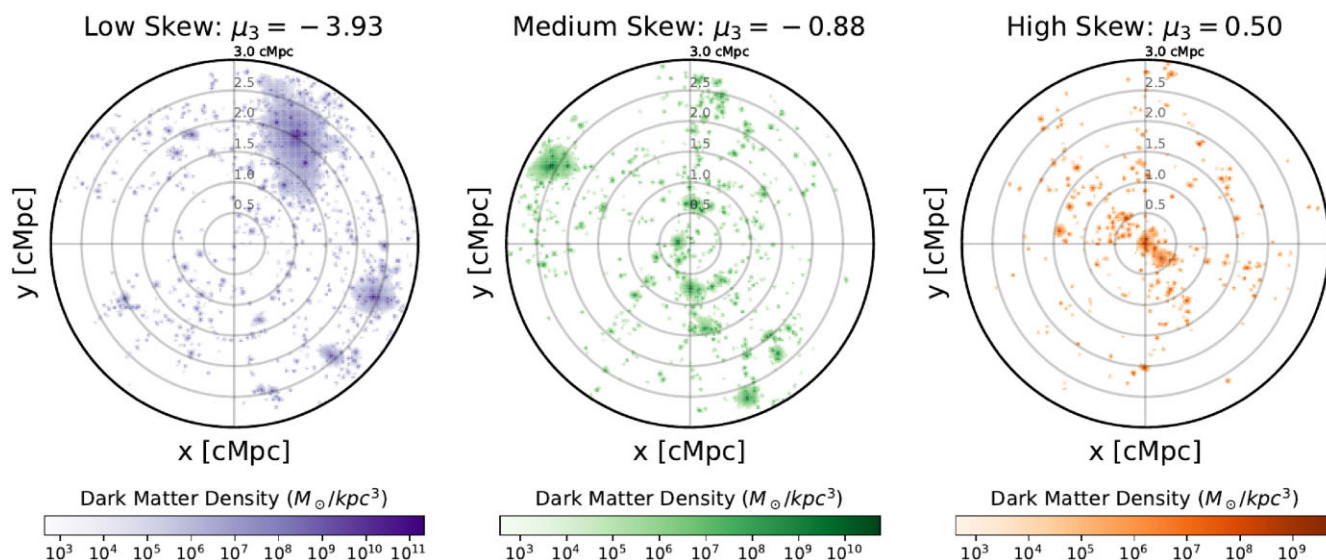


Figure 1. Obtained from the $z = 1$ TNG100-1 snapshot, this figure is a visualization of the x – y plane projection of the dark matter density distribution of subhaloes surrounding three central subhaloes of $\log M_h^{z=1}/M_\odot = 11.17$. The target subhaloes are not shown in these images, just as they play no role in the skew calculation. Dark matter cells are selected for this image provided that they lie within a sphere of radius 3 Mpc, centred on the target subhalo’s centre of mass, and are not gravitationally bound to the target subhalo. ‘Low’ and ‘high’ skewness refer to low and high quantiles of the skew data set, while ‘medium’ skewness are close to the median skew.

where the skew is the third moment by definition, and shall henceforth be denoted μ_3 . In our implementation, w_j are replaced by the mass of each subhalo, and x_j the distance to the target subhalo, in Mpc.

Merger events between dark matter haloes dramatically enhance star formation in the galaxies they host, by introducing cold gas and triggering large tidal disturbances in both merging galaxies. The scale of this is governed by the relative masses of haloes (i.e. the merger ratio) and the gas and star content of the galaxies in question (Wechsler & Tinker 2018; Bose et al. 2019). Desiring a temporal handle on significant merger events, a number of parameters have been trialled, such as the snapshot of the merger event and the mass ratio at this time. However, merger events take place on varying time scales and thus cannot be assigned a definitive redshift (Rodríguez-Gomez et al. 2015), and are prone to errors by nuances in the halo referencing system in Illustris (Poole et al. 2017). In fact, we find that a set of mean merger ratios occurring at sequential time steps fails to constrain the galaxy evolution significantly.

The skew as a function of time offers a parametrization of the merger history for each halo, as the most massive subhaloes will have the largest influence on the distribution of matter, and in the process of a merger, will skew the distribution more and more positively during infall. For satellites, we choose to evaluate the skew out to a 1 Mpc radius, aiming to measure both mergers in the central phase and collisions interior to the FoF halo in the satellite phase. For centrals, we evaluate the skew of the radial distribution up to 3 Mpc. This larger radius is necessary to contain data exterior to the largest of FoF haloes, whose accretion activity has a profound effect on the central galaxy. Satellite galaxies, on the contrary, are more dominated by the mass distribution within the FoF halo, justifying the use of a smaller scale skew measurement.

Typical environments for $z = 1$ subhaloes with low (low quantile), medium (near median), and high (high quantile) skew are exemplified in Fig. 1, where the target subhalo is omitted for clarity of the exterior mass distributions. An object with insignificant (medium) skew has an unbiased local distribution of matter, such as in the central

panel. For a low skew distribution, one or several subhaloes, large enough to dominate the local environment, will shift the distribution’s centre of mass far from that of the target subhalo. In highly skewed distributions the largest subhaloes are instead concentrated near the target subhalo, likely to merge with the target halo or at least invoke a significant tidal disturbance. The skew at a given time is therefore a measure of the local concentration of dark matter, whereas its variation with time describes the nature of flybys and collisions with the subhalo, potentially even its satellites.

As in Section 2.1.2, the GRISPY package is used to capture the subhaloes within a 3 Mpc radius of each target subhalo. The IDs and distances from the target to the objects are returned automatically, thus making calculation of the radial mass distribution easy. The skew is simply obtained from this using equation (5), with $n = 3$. This skew is computed for all target subhaloes, and at all snapshots of TNG, to gain a full data set of the skew histories.

In addition to μ_3 , the distance to the closest subhalo is also used for the input of the neural network. Denoted d_{μ_3} and measured in Mpc, this scales the distribution such that the skew correlates with the true location of the merging halo, and serves as a simple metric for the proximity of a merging halo itself. Both of these quantities are vector normalized.

2.1.4 Orbital velocity and half-mass radius

Lovell et al. (2021) use extremely randomized trees (ERTs) to predict zero-redshift baryonic properties from their dark matter haloes, and indicate that the maximum circular velocity of the subhalo’s rotation curve (v_{\max}) and the radius enclosing half of the subhalo’s dark matter mass ($R_{1/2}$) have the greatest predicting power in their data set when concerning stellar, black hole, and gas mass components, metallicity, and instantaneous star formation rate; most likely an indicator of the speed of collapse and thus starbursts and black hole accumulation (Davies et al. 2019; Lovell et al. 2021). The authors debate the potential use of historical properties in later work, as a

means of improved accuracy of their predicted stellar mass function and various mass relations.

A temporal measure of v_{\max} is not so simple to implement in our model, due to the presence of baryons influencing the rotation curve. Through inspecting merger trees in TNG, these velocities are larger in the hydrodynamical simulations when compared with the N -body equivalent, for objects in the satellite phase and for centrals at high redshift. We compute a term which is unaffected by the presence of baryons, and is loosely proportional to the temporal virial velocity of the subhalo, in terms of the dark matter half-mass radius $R_{1/2}$, and halo and subhalo mass, respectively, for central and satellite galaxies:

$$\tilde{v}_{\text{vir}}(t) = \sqrt{\frac{m_h(t)}{R_{1/2}(t)}}, \quad (6)$$

where we have ignored constant terms such as the Newtonian gravitational constant. This proxy assumes that TNG subhaloes possess a common radial density profile, and thus a simple scaling between its total mass and the mass enclosed within a given mass radius. As well as an estimate for virial circular velocity, it is similar to the proxy for NFW concentration used for TNG data by Bose et al. (2019).

We have implemented $R_{1/2}$ and \tilde{v}_{vir} as features at all redshifts in our data set. The required quantities are directly obtained from the TNG data catalogue, and the features are vector normalized in our pre-processing stage, thereby gaining a description of the growth of the subhaloes' physical sizes and rotation curves.

2.2 Time-independent variables

2.2.1 Final halo mass

The diversity of galaxy formation with respect to mass is an important one: the highest mass haloes, hosting the highest mass centrals and the greatest abundance of satellites, typically exhibit high mass elliptical galaxies, whose metal content is high and whose star formation has ended some considerable time ago. On the contrary, smaller haloes host younger, continually star-forming spiral galaxies, with large amounts of metal-poor gas (Wechsler & Tinker 2018). The halo mass is therefore an important quantity determining the properties of star formation and chemical enrichment history of our galaxies.

The final halo mass, while taken from the TNG data directly, equates to the integral of the halo rate over the full time of the simulation. As is the case with the accretion rates (see Section 2.1.1), the final subhalo dark matter mass is included in the satellite neural network alongside the halo mass, as an indicator of the present-day baryonic properties governed by differing gravitational regimes. Again, the subhalo mass is not considered for centrals due to their tight correlation from the central subhalo being the dominant mass.

As with all variables in Section 2.1, the zero-redshift halo mass is normalized using the Gaussian quantile transformation (GQT). The maximum absolute dark matter accretion rate across the halo's history is also used as an input variable, and is normalized the same way. This also serves as a measure of the magnitude of dark matter accretion.

2.2.2 Specific rate gradient

A pre-eminent parameter used by Montero-Dorta et al. (2021) is the gradient of the specific mass accretion rate, which they denote β .

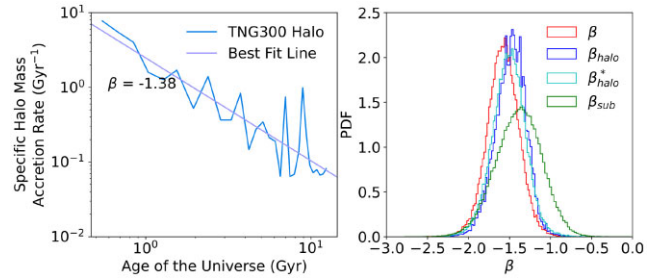


Figure 2. (Left-hand panel) An example of a straight line fit to the specific mass accretion rate of a randomly chosen halo, whose best fit gradient defines β . (Right-hand panel) The PDF distributions of our β data, including that of the unique subset of the satellite host values, denoted β_{halo}^* . These closely resemble Gaussian distributions and are well fit by the Gaussian function, with best fit parameters given in Table 2. β represents the gradients of the halo accretion histories in the central data, the rest are specific to satellite subhaloes (β_{sub}) and their host haloes (β_{halo} , β_{halo}^*).

On the left-hand panel of Fig. 2, β is calculated for one TNG300 halo by fitting a straight line between the logarithms of specific mass accretion rate and cosmic time.

β effectively identifies the fastest forming haloes at high redshift, whose galaxies maximize their star formation rate at similar times and go on to form high mass, quenched galaxies. Its inclusion in the neural network advocates a measure of specific accretion, which accounts for the halo's growth in proportion to its current mass, and the time scale of the halo evolving mass fractions. Like the halo mass, this too is an important factor in classifying galaxies of a certain evolutionary regime.

Additionally, we find that this parameter is useful for filtering out outliers, which can negatively impact the network's performance. For the lowest mass objects, in spite of the lower final stellar mass limit of $10^9 M_{\odot}$, certain accretion histories exhibit a noisy, stochastic appearance as one approaches the mass resolution of the simulation. More importantly, these under-resolved accretion histories are flat, and so their β values do not represent a typical specific mass accretion history. We therefore exploit the Gaussian nature of β by fitting to its distribution, and discard any samples with over a 5σ offset from the mean.

As β is Gaussian distributed, as can be seen in the right-hand panel on Fig. 2, its histogram is very similar in geometry to the GQT transformed features. Translated according to the best-fitting parameters, i.e. to a normal distribution of zero mean and unit width, the range of values is also the same as the GQT data. Therefore, we simply feed this translated β , hereafter α , into the neural network, with no need for quantile transformation.

For satellites, two distinct halo formation histories serve a role in the model: that of the FoF halo and that of the satellite subhalo. There are therefore two different β values each. Each β , like the β for centrals, is also Gaussian distributed, however for our satellite sample the means are larger and the standard deviations are similar (see Table 2), likely due to the most rapidly forming haloes hosting a greater abundance of high-mass satellites.

The α values for each formation history are also included in the satellite neural network, and satellites are selected under the circumstance that neither α has an absolute value above 5. To indicate the halo to which α or β refers, these will be denoted $\alpha_{\text{halo}}/\beta_{\text{halo}}$ for the main halo, and $\alpha_{\text{sub}}/\beta_{\text{sub}}$ for the satellite subhalo.

Note that our satellite data includes multiple subhaloes bound to the same halo; there are therefore duplicate values of β_{halo} in

Table 2. The best-fitting mean and standard deviation for all β distributions in our data. β_{halo} has two different fits due to the multiple occurrences of haloes which host multiple satellites in our data. β_{halo} denotes the data set used in training, while β_{halo}^* denotes the unique-valued subset of β_{halo} .

Network	β : Best-fitting Gaussian parameters		St. Dev.
	Quantity	Mean	
Central	β	−1.553	0.213
Satellite	β_{halo}	−1.489	0.192
	β_{halo}^*	−1.509	0.203
	β_{sub}	−1.379	0.274

the training set. Table 2 distinguishes the Gaussian fits for the training data β_{halo} and the unique subset β_{halo}^* , and shows that the true distribution is steeper and more similar to the central β distribution, which can be seen in Fig. 2. Their offset can be attributed to effectively sampling a different mass range; we impose a lower stellar mass cut of $10^9 M_{\odot}$ to all targets, and for satellites this implies a larger lower boundary of the host mass.

2.2.3 Cosmic web properties

The Discrete Persistent Structure Extractor (DISPERSE) code (Sousbie 2011) is a geometric algorithm which establishes the stationary points of a density field and quantifies its skeletal structure accordingly. We use the cosmic web catalogue data built on the TNG simulations using DISPERSE.² Specifically, we make use of the distances from the target halo to the nearest critical points and dark filaments, denoted collectively as d_{CW} :

- (i) d_{node} Distance to the nearest node (maximum) of the density field
- (ii) d_{minima} Distance to the nearest void (minimum) of the density field
- (iii) d_{saddle1} Distance to the nearest saddle point with one minimized dimension
- (iv) d_{saddle2} Distance to the nearest saddle point with two minimized dimensions
- (v) d_{skel} Distance to the midpoint of the nearest filament

This characterization of the dark matter environment gives a simple description of the level of anisotropy of the large scale environment of the halo, and can be interpreted as characterizing the tidal field that surrounds it. The cosmic web has been shown to modulate the accretion of matter onto haloes (Hahn et al. 2009; Borzyszkowski et al. 2017). Observationally, haloes of a similar mass will have distinct formation rates according to the surrounding density field (Tojeiro et al. 2017; Tinker et al. 2018), and tend to possess different morphologies and internal dynamics (Hellwing et al. 2021).

Properties such as halo mass, and scale-independent quantities such as NFW concentration, are also strongly correlated with their tidal environment (Hellwing et al. 2021; Ramakrishnan, Paranjape & Sheth 2021), therefore this will not be a unique indicator of the halo formation process. However, the cosmic web does influence the circumgalactic and intergalactic media directly, such as by transfer of metal-rich gas expelled by higher mass galaxies. Thus, they are in principle useful for identifying certain environmental aspects of the galaxy–halo connection.

These cosmic web distances, while useful for modelling the large scale environment influencing accretion of star-forming gas to central

haloes, is not considered so paramount to objects in the satellite phase. Simpson et al. (2018) show that the majority of satellite quenching stems from the ram pressure experienced upon infall, or the subsequent tidal effects of the host halo. While they briefly suggest that the cosmic web may quench some satellites, they suggest that this primarily affects low mass satellites which intersect the gas inflow from the filament to the host. Thus, we do not use DISPERSE quantities in the satellite neural network.

2.2.4 Starting time

Another parameter of the neural network is the time at which the halo first began to form, taken as the earliest snapshot at which the given merger tree is defined.

The structure of the recurrent layer requires identical time steps for all samples, yet haloes begin to form at different times. We interpolate the time-dependent properties and return their values at every third snapshot in TNG, meaning many will have no data at the earliest times.

While the recurrent layer enforces a causal relationship between time steps, the starting time is nevertheless used as a parameter to establish recently germinated haloes directly, and thus, the probable features of their galaxies.

2.2.5 Satellite infall

The properties of the satellite subhalo and its parent halo at the time of infall have been shown to be crucial measures of aptitude for star formation (Pasquali et al. 2010; Wetzel et al. 2013). Shi et al. (2020) study regimes of satellite evolution by categorizing according to a scaled formation time:

$$a_{\text{max}} \equiv \frac{1 + z_{\text{half}}}{1 + z_{\text{max}}}, \quad (7)$$

where z_{max} is the redshift of the TNG snapshot at which the subhalo’s mass is maximized, and z_{half} is the redshift of the snapshot at which half of this mass is attained, for the first time. The authors classify satellites as fast-accreting if this value is small, and slow-accreting otherwise; finding that fast-accreting satellites have greater star formation, gas abundance, and a systematically distinct stellar–halo mass relation (SHMR).

We compute this quantity using the redshifts of the snapshot at which these masses are attained, as Shi et al. (2020) did in TNG100-1. A second quantity is computed similarly:

$$a_{\text{infall}} \equiv \frac{1 + z_{\text{half}}}{1 + z_{\text{infall}}}, \quad (8)$$

where z_{infall} is the redshift at which the subhalo becomes bound to the central halo. a_{infall} effectively characterizes the stage in the subhalo’s growth history at the point of infall, or the time scale of its capture with respect to its growth, whereas a_{max} , however evidently indicative of satellite galaxy evolution, relates instead to its central-phase growth profile.

A difference in the properties of a_{infall} and a_{max} is that a_{max} is the ratio of two strictly consecutive times in the subhalo’s growth, therefore it has a lower bound of 1. For approximately one in 72 of our samples, $a_{\text{infall}} < 1$, indicating that infall occurs before the subhalo reaches half of its peak mass.

Additionally, two physical quantities are evaluated at the time of infall, each of which provide a measure of the properties and dynamics of the halo–subhalo system and how this will affect the future of the tidal environment. One is the absolute velocity of the

²https://github.com/Chris-Duckworth/disperse_TNG/

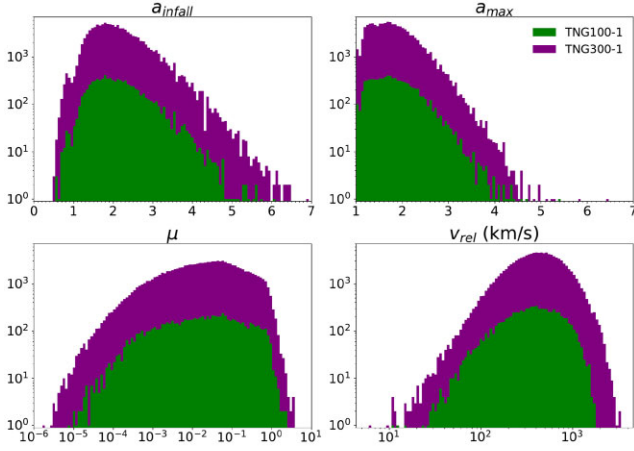


Figure 3. Stacked histograms of the four properties of infall satellites, discerned by simulation, used in the neural network for satellites. Clockwise from the top-left, these include the scaled infall time a_{infall} , the scaled formation time a_{max} from Shi et al., the velocity v_{rel} of the satellite relative to its host at the infall snapshot, and finally the mass ratio μ of the two haloes.

infalling subhalo relative to its host, calculated simply using the difference in peculiar velocity vectors from each object’s merger tree:

$$v_{\text{rel}}(z_{\text{infall}}) = \|\mathbf{v}_{\text{rel}}(z_{\text{infall}})\| = \|\mathbf{v}_{\text{pec}}^{\text{sub}}(z_{\text{infall}}) - \mathbf{v}_{\text{pec}}^{\text{halo}}(z_{\text{infall}})\| \quad (9)$$

In the satellite phase, the velocity of the satellite relative to the halo is related to the rate of the satellite’s mass-loss from ram pressure, and similarly its orbital velocity serves as a measure of its location in the halo’s potential and thus its likelihood of continued star formation (Behroozi et al. 2019; Slone et al. 2021). Including the velocity at the time of infall may prove a useful measure of the satellite’s trajectory.

The final infall parameter in this model is the ratio of the subhalo mass to its host halo’s mass at infall time:

$$\mu(z_{\text{infall}}) = \frac{m_{\text{sat}}(z_{\text{infall}})}{M_{\text{parent}}(z_{\text{infall}})}, \quad (10)$$

which introduces an important selection criterion which applies only to satellites.

It is assumed that the dark matter subhalo is the dominant mass of the galaxy–halo system. However there exist some low mass subhaloes in the TNG simulations where the dominant mass is gas or stars, which can be attributed to tidal dwarf galaxies in TNG (Haslbauer et al. 2019). While a satellite may initially be slightly larger than its future central if the latter is rapidly growing, a value of multiple positive orders of magnitude illustrates the collapse of this key assumption. We therefore only include satellite galaxies whose host halo obeys the following criterion: $m_s(z_{\text{infall}})/M_h(z_{\text{infall}}) < 0.1$, reducing the maximum μ by two orders of magnitude. The distributions of the infall parameters following this cut are shown in Fig. 3.

3 BARYONIC QUANTITIES

This section deals with the pre-processing and scientific interpretation of the galaxy properties which our neural network is designed to predict, and their complex role in governing observables.

Like the input variables, these properties are GQT normalized. The target data set consists of the zero-redshift stellar mass and mass-weighted metallicity of the galaxy, and the time-dependent

SFH and ZH, both vector normalized and defined for the same time steps as all time-dependent input features.

Defined only on the subhalo scale, the definitions of these historical baryonic quantities are identical for central and satellite galaxies, with the exception of resolution corrections in Section 3.3, in which the conditional means of these quantities must be evaluated separately for centrals and for satellites. Unlike dark matter properties, we consider only baryonic properties on the subhalo scale.

3.1 Star formation history

We define our star formation rate histories from the mass-weighted age distribution of star particles gravitationally bound to each subhalo at $z = 0$ – i.e. the stellar mass formed per unit time as a function of cosmic time, $\mathcal{S}(t)$. This definition differs from a stellar mass accretion history, as obtained directly from the merger tree, and analogous to the dark matter accretion rate in Section 2.1.1. We choose to focus on the first definition given our goal to produce spectral energy distribution for each galaxy. In principle, of course, both can easily be obtained from the simulation.

The total stellar mass ever formed is defined as the integral of the SFH \mathcal{S} over time. Due to recycling, this integrated stellar mass is larger than the $z = 0$ stellar mass in the merger tree, which we include as a time-independent feature. Unless stated explicitly, this paper refers to the integral of the SFH when referring to stellar mass.

We also define a mass-weighted age (MWA) for each galaxy as

$$\text{MWA} = \frac{\sum_{j=n}^N \mathcal{M}_n t_n^{\text{lookback}}}{\sum_{n=1}^N \mathcal{M}_n}, \quad (11)$$

where the sum is done over N snapshots, \mathcal{M}_n is the stellar mass formed in snapshot n , and t_n^{lookback} is the lookback time to snapshot n .

We compute the SHMR for the true and predicted galaxies. This effectively describes the mean stellar mass of a galaxy as a function of its halo mass, while the scatter at fixed halo mass encompasses the variance in SFHs associated with differing galaxy growth mechanisms, some of which are highly regulated by halo mass (Gu, Conroy & Behroozi 2016; Wechsler & Tinker 2018; Behroozi et al. 2019). Our model is considered adequately fit to the star formation histories provided that the numerical integrals of their SFHs accurately replicate the shape and scatter of the SHMR at $z = 0$.

3.2 Stellar metallicity history

The metallicity histories are computed as mass-weighted metallicities of all star particles associated with a subhalo at $z = 0$, in the same time bins as the star-formation history.

We also define a mass-weighted metallicity of the full galaxy as:

$$Z_s = \frac{\sum_{n=1}^N \mathcal{M}_n Z_n}{\sum_{n=1}^N \mathcal{M}_n}, \quad (12)$$

where the sum is done over N snapshots, and \mathcal{M}_n is the mass formed in snapshot n .

3.3 Resolution corrections

A prominent issue with using the two primary TNG simulation data sets is due to their difference in mass and spatial resolution. Pillepich et al. (2017a, b) explain that the stochastic star formation model in TNG is dependent on the density of gas mass which is

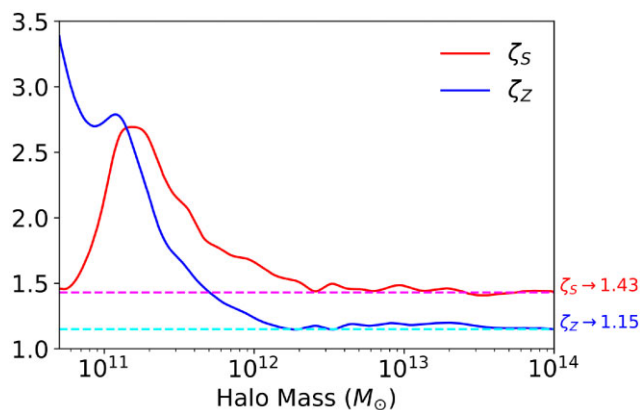


Figure 4. The ζ variables as a function of halo mass at redshift zero. The values of each ζ function for masses larger than that shown here (i.e. $>10^{14}M_{\odot}$) are taken as the average of the function over the $[10^{13}M_{\odot}, 10^{14}M_{\odot}]$ interval. The lower mass limit of this chart equates to the cutoff of TNG300 haloes used in our data.

identified, therefore simulations with lower resolution underestimate the star formation rate overall. Consequently, important summary statistics such as the SHMR and mass–metallicity relation (MZR) are underestimated in TNG300-1 in contrast with TNG100-1.

The same authors offer an adjustment to TNG300-1 by exploiting the fact that the resolution of TNG100-2 is identical to it. More generally, the n -th TNG300 simulation is chosen to recover the mean SHMR of the $n + 1$ -th TNG100 simulation. While they utilize this only to correct the SHMR, we will do the same to adjust our stellar metallicities. These corrections are based on the halo mass dependent ratio between the mean SHMRs and mean MZR of TNG100-1 and TNG100-2, and we denote them ζ_S and ζ_Z , respectively. The shapes of these functions of halo mass at redshift zero are depicted in Fig. 4.

At a fixed redshift, the ζ fractions are defined:

$$\zeta_S(M_h | z) = \bar{M}_{100-1}^*(M_h) / \bar{M}_{100-2}^*(M_h) \quad (13)$$

$$\zeta_Z(M_h | z) = \bar{Z}_{100-1}^*(M_h) / \bar{Z}_{100-2}^*(M_h), \quad (14)$$

where due to a lack of high mass samples, ζ_S and ζ_Z are assigned their mean value in the interval $10^{13}M_{\odot} \leq M_h \leq 10^{14}M_{\odot}$ if $M_h \geq 10^{14}M_{\odot}$.

The ζ corrections work adequately for adjusting the relevant relations within a single snapshot. However, the stellar age spectra used to compute the formation histories of the network are not defined with single-snapshot data, and so ζ is not a valid correction for these. Instead we apply a similar correction for the mean histories of objects in narrow bins of final halo mass, over which we apply a cubic spline interpolation, which can be neatly extrapolated outside the halo mass range of TNG100.

This returns a temporal variable, named ψ , whose geometry for star formation and metallicity as a function of both time and halo mass is realized in Fig. 5.

For a fixed halo mass, ψ is given mathematically as follows:

$$\psi_S(z | M_h^{z=0}) = \bar{S}_{100-1}(z) / \bar{S}_{100-2}(z) \quad (15)$$

$$\psi_Z(z | M_h^{z=0}) = \bar{Z}_{100-1}(z) / \bar{Z}_{100-2}(z) \quad (16)$$

To reiterate, the baryonic properties required for this study include the final stellar mass M_s and metallicity Z_s , whose TNG300 values are multiplied by ζ : a dimension-less function of halo mass at fixed redshift; and the star formation history \mathcal{S} and metallicity history

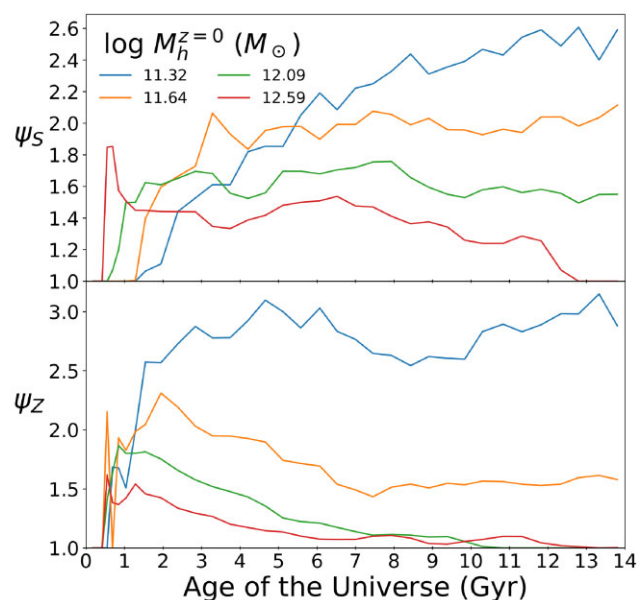


Figure 5. The ψ variables as a function of cosmic time, shown in four example bins of halo mass. The resolution correction for the star formation history ψ_S is shown in the upper panel, the metallicity history correction ψ_Z in the lower panel.

\mathcal{Z} , whose TNG300 values are multiplied by ψ : a dimension-less function of redshift at fixed final halo mass. ζ and ψ are computed independently for central and satellite galaxies, due to significant differences in their star formation histories.

Using these corrections, we find star formation and metallicity histories in TNG300 that are accurately matched to the TNG100 data, and therefore are suitable for the neural network. These matches are displayed in Figs 6 and 7.

3.4 Spectroscopy and photometry

We compute a set of spectral energy distributions from our original and predicted SFHs and ZHs using the PYTHON-FSPS module.³ This is based on a series of simple stellar population (SSP) spectra with initial mass function (IMF) in accordance with the Chabrier (2003) model. For each time step in our data, we emulate an SSP spectrum, parametrized by the current time and metallicity of the galaxy, and weight them according to the current star formation rate. Mathematically, the full spectrum \mathcal{F} is defined in terms of the SSP spectra f_j as follows:

$$\mathcal{F}(\lambda) = \sum_{j=1}^{N_{\text{snap}}} \mathcal{M}_j f_j(\lambda; \mathcal{Z}_j, z_j) \quad (17)$$

Photometric magnitudes are computed from these spectra using SDSS methodology. We mimic the flux passing through the five bandpass filters used in the survey (Fukugita et al. 1996) by integrating the spectrum over the response functions from SDSS, and compute absolute magnitudes and colours.

The distribution of colours from several combinations of bands is bimodal, where ‘blue’ galaxies have ongoing star formation, ‘red’ galaxies are quenched, and galaxies in the transition phase from blue to red are members of the ‘green valley’ (Nelson et al. 2017).

³<https://github.com/dfm/python-fsps>

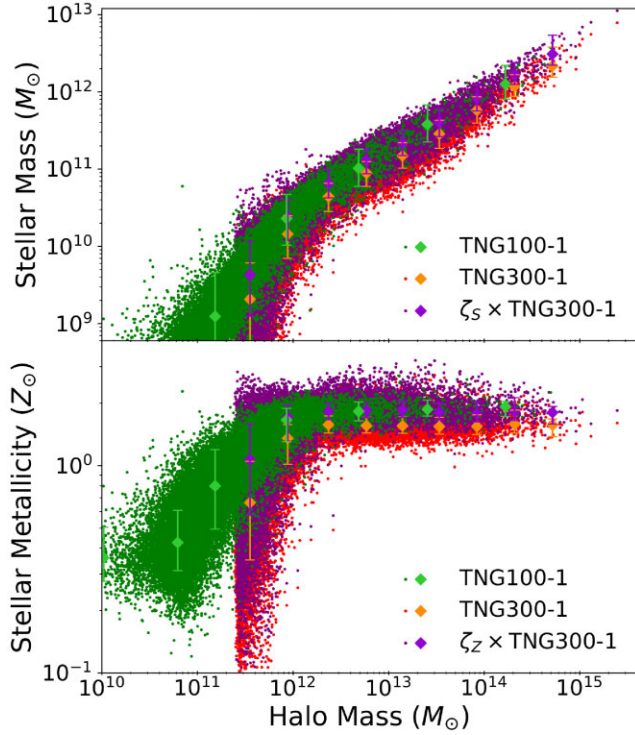


Figure 6. The SHMR and halo mass–metallicity relation at $z = 0$, showing the TNG100-1 data aligned with the TNG300-1 distributions, following amendment using the ζ functions. For each data set, similarly coloured errorbars indicate the median and range between 15-th and 85-th of stellar mass or metallicity in a given halo mass bin.

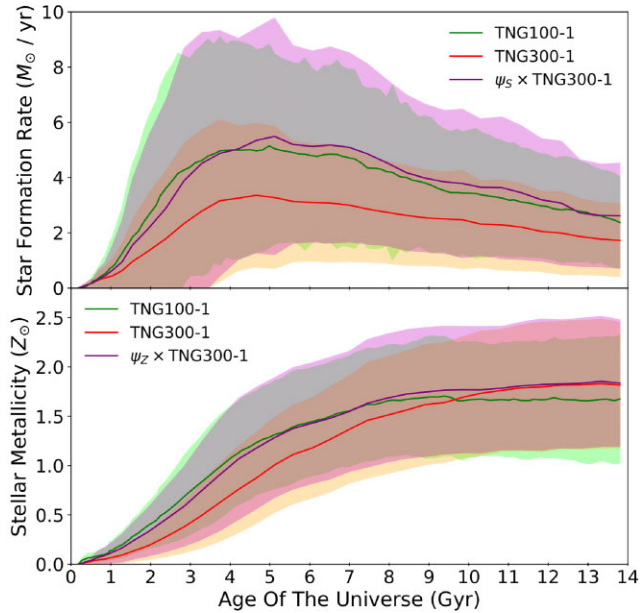


Figure 7. The mean star formation and metallicity histories for central galaxies of halo mass $10^{12} M_{\odot} \leq M_h^{z=0} \leq 10^{12.2} M_{\odot}$. The standard deviation of these data as a function of time are indicated by the size of the shaded regions. This shows the TNG100-1 data aligned with the TNG300-1 means, adjusted by the ψ parameters. To remove spurious features due to low numbers at early times, this figure only includes time steps with more than 100 nonzero values in all data. This applies only to times before a cosmic time of 1Gyr.

The colour–mass diagrams for centrals and satellites, inferred from network predictions are compared here with results from TNG.

We also compute the $H\alpha$ luminosities of our galaxies by taking the SFH-weighted sum of SSP values, as given by PYTHON-FSPS:

$$L_{\text{gal}}^{H\alpha} = \sum_{j=1}^{N_{\text{snap}}} \mathcal{M}_j L_j^{H\alpha} \quad (18)$$

4 NEURAL NETWORK DESIGN AND DATA PRE-PROCESSING

4.1 Quantile transformation

It is common practice, and indeed required, to normalize data sets in many applications of machine learning. Normalization is especially important for data sets with variables of very different and large ranges, such as ours.

For much of our data, we found that a simple scaling relation was inadequate as several quantities are not represented equally. For example, halo masses are heavily over-represented at low values, while large values are under-represented in our sample, leading to poor training. A partial solution to this issue was a GQT: an operation supported in the SCIKIT-LEARN (Pedregosa et al. 2011) Python Library.

Quantile transformations work as follows. Let x_i be a data point in the original data set. If the distribution of these data is normalized, one may interpolate through this and integrate it to establish a cumulative distribution function (CDF) Ψ , which increases monotonically from 0 to 1. Ψ therefore maps the data point x_i onto its corresponding quantile value. It is therefore very simple to map this onto a second probability distribution provided that its CDF and its inverse function are analytical. Specifically, if this second distribution has a CDF Φ , then the data x_i can be mapped onto this distribution like so:

$$y_i = \Phi^{-1}(\Psi(x_i)) \quad (19)$$

If we intend for the y_i to be Gaussian distributed, then equation (19) takes the specific form:

$$y_i = \sqrt{2} \operatorname{erf}^{-1}(2\Psi(x_i) - 1) \quad (20)$$

and can be inverted when returning predicted data to physical values:

$$x_i = \Psi^{-1}\left(\frac{1 + \operatorname{erf}\left(\frac{y_i}{\sqrt{2}}\right)}{2}\right) \quad (21)$$

A GQT was chosen specifically because the domain of our transformed data resides entirely between 5 and -5 , making it suitably normalized for the neural network. SCIKIT-LEARN also offer a transformation to a uniform distribution, however the Gaussian transformation proves much more suitable for our data. Fig. 8 shows how our TNG100-1 sample of logarithmic halo masses is mapped onto these distributions. In the uniform case, a number of data points are strongly skewed towards the edges of the distribution. This narrow margin containing such a broad range of values means that there can be a large error in the true data when making predictions based on a uniform distribution.

4.2 Vector and scalar normalization

In the context of time-dependent variables, the GQT can take two specific forms. In one form, we consider that the behaviour of a

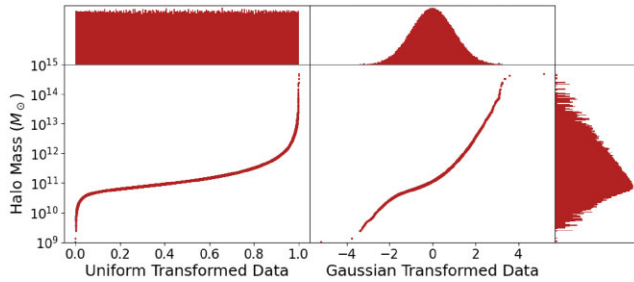


Figure 8. The mapping of the TNG100-1 set of halo masses (vertical axis) to a uniform distribution (left-hand panel) and Gaussian distribution (right-hand panel). Histograms of these data points are shown along every axis, including the halo mass distribution on the vertical axis. This graphic shows that a large range of data, namely halo masses above $10^{12.5}M_{\odot}$, corresponds to a very narrow range in the uniform distribution, which results in high sensitivity to small differences in the transformed data. This makes a uniform distribution unsuitable for making predictions from our data, and this is why we have chosen to transform to a Gaussian distribution instead.

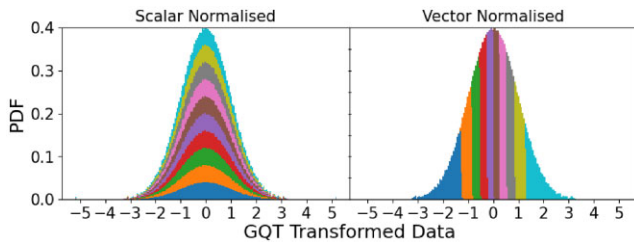


Figure 9. A depiction of the distribution of arbitrary, monotonic data after applying scalar (left-hand panel) and vector (right-hand panel) GQT normalization. Data points at successive time steps are shown in different colours. All time steps share the same normal distribution when transformed independently of other time steps, i.e. when scalar normalized. When vector normalized, the data points are transformed according to the full range of the quantity’s value over time, hence each time step’s distribution is relative to another. Stacking the full set of time step histograms results in the Gaussian distribution of the full data set, regardless of normalization method.

dark matter quantity as a function of time can influence the present-day state of the galaxy, and so we intend to preserve the geometry (or shape) of the property’s history. In the other, there may be physical or systematic differences between time steps which imply that there is no physical meaning to the gradient of the variable with time, and instead the absolute value at a given point in time is more significant. We introduce two forms of normalization: vector and scalar normalization, to incorporate time-dependent properties differently. The effect that each normalization has on temporal data is illustrated in Fig. 9.

With vector normalization, the GQT is fit to the variable at all time steps simultaneously. The transformed variable corresponds to the original variable’s value regardless of the time at which it is defined. Equation (20) refers to a GQT for a time-independent variable. If Ψ is the CDF of x_i : the 1D set of values of the temporal variable, the transformed variable as a function of time is defined by applying the transformation to the original, 2D data set:

$$y_i(t_j) = \sqrt{2} \operatorname{erf}^{-1} (2\Psi(x_i(t_j)) - 1), \quad (22)$$

where t_j is a specific time step. These are implemented into the recurrent input layer as a time-dependent vector, whose normalized value serves as an absolute indicator of the original data value, and therefore the temporal gradients of the variable are preserved.

Any variable for which the GQT is fit with no respect to time is considered scalar normalized. Here there exists a separate CDF for the variable at each time step, and thus a list of independent sets of transformed variables, separated in time:

$$y_i^k = \sqrt{2} \operatorname{erf}^{-1} (2\Psi_k(x_i^k) - 1) \quad (23)$$

4.3 Neural network architecture

We have developed two semi-recurrent neural networks in TENSORFLOW (Abadi et al. 2016), which have been trained separately on central and satellite galaxy data from TNG. In each network, the temporal component of the halo data constitutes the recurrent input data, while the remaining zero-redshift quantities such as halo mass are processed in the second, dense input layer.

While TNG contains 100 snapshots in time, for the sake of reducing the complexity of the neural network for better convergence, we evaluate all time-dependent halo and galaxy properties with a step size of three snapshots, thus computing a 33-element vector.

The architecture for the central neural network is depicted in Fig. 10. A sequence of number-labelled points indicate layers of different types and their dimensions. Each of our networks begin with a 2D input of features defined at 33 time steps, and a 1D set of scalar features. After multiple dense or recurrent layers, the 1D layers which succeed them are eventually concatenated, allowing both sets of data to predict baryonic properties. The recurrent and dense input quantities are discussed at length in Sections 2.1 and 2.2, respectively.

The satellite neural network is identical in design to the central network shown in Fig. 10, except that there exist seven temporal and eleven non-temporal input quantities, which again are discussed in Sections 2.1 and 2.2. The input layers therefore have dimensions of 11 and 33×7 nodes, combining in the same sequence to make a 44-node dense layer instead of 42. This is followed by a set of 45-layer nodes, and the rest of the satellite network remains true to the central design.

Two separate models exist due to the inclusion of variables which are not defined for central subhaloes, such as the time of infall into a larger host halo. Satellite galaxies also have different evolutionary properties such as decoupled growth of the subhalo and host halo, which are included in this model, and typically have larger stellar mass fractions and scatter in stellar mass at high halo mass when compared with centrals (Engler et al. 2020), and these differences may not have been distinguished in a composite model.

Due to the sparse nature of input quantities such as halo masses and overdensities, gradient saturation renders the network inoperable when using highly non-linear activation functions such as tanh or sigmoid. When trialling the rectified linear unit (ReLU) activation function, defined as follows:

$$\operatorname{ReLU}(x) \equiv \max [0, x] \quad (24)$$

we find no noticeable saturation effects. However, this network was subject to the Dying ReLU Problem (Lu et al. 2020), in which the gradient for any ReLU-activated node is automatically zero if its input is negative, thus all subsequent iterations from this node are zero, and will no longer contribute to training the model. To reduce the number of affected nodes, a common approach is to use the similar Leaky Rectified Linear Unit (L-ReLU) activation function:

$$\operatorname{L-ReLU}(x) \equiv \max [\alpha x, x]; \alpha \in (0, 1) \quad (25)$$

However, the discontinuity in the gradient of L-ReLU resulted in arbitrary discrepancies between similar samples in our data. Finally,

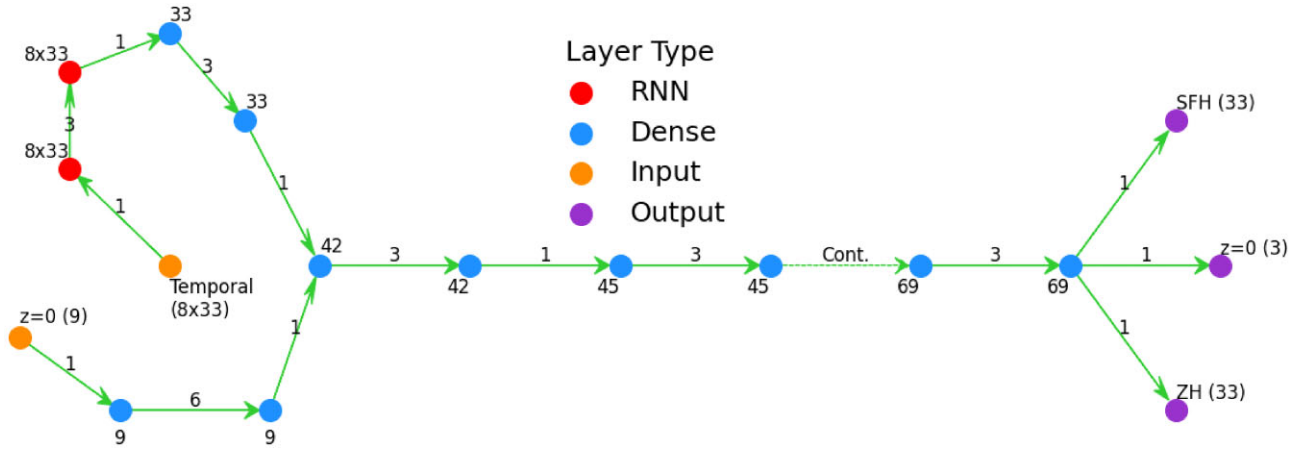


Figure 10. An illustration of the architecture of the neural network for central galaxies. Each point in this diagram represents a fully connected layer, whose dimensionality is indicated by a numerical label; with the exception of purple points, which represent a subset of the final, 1D output layer. An input layer exists for the time-dependent and time-independent halo properties each, whose outputs are conjoined at a 42-node dense layer. The temporal input layer and recurrent layers are 2D, comprising eight variables defined over 33 time steps. Each arrow indicates a connection between consecutive layers, while the arrow's label indicates the number of times that this connection repeats (i.e. '3' indicates that four consecutive hidden layers exist to accommodate three such connections). The dashed line arrow indicates that the process of adding three nodes to every fourth hidden layer repeats up to the point that there are 69 nodes per layer. The final set of connections returns the baryonic data output, consisting of star formation and metallicity histories, and zero-redshift galaxy properties.

the ELU activation function was tested:

$$\text{ELU}(x) \equiv \begin{cases} x & \text{if } x \geq 0 \\ \alpha(\exp(x) - 1) & \text{if } x < 0 \end{cases}, \quad (26)$$

which, with an α value of 1, offered a solution to both of these problems.

ELU activated networks are useful for scaling a layer's output to a near-zero mean and near-unity standard deviation, however this behaviour is only stable for sequential architectures of standardized inputs and initial kernel weights (Clevert, Unterthiner & Hochreiter 2015; Géron 2020). While ELU alone offers significant performance over a batch-normalized ReLU network (Shah et al. 2016), this instability along with the prospect of high gradients and saturation at highly negative inputs can limit the network's performance. The network is not fully deterministic, yet it converges adequately with ELU activations, unlike more direct ReLU-derived functions.

All layers in these networks are fully connected, i.e. there exist no dropout layers or other dilution techniques in either network; the reduction of connections this introduces is what the ELU activation function was chosen to mitigate, alongside the two-sided gradient problem of L-ReLU. We find that such dilation methods prove detrimental to the network's performance when using even ELU activation.

The TNG data are shuffled and split such that one quarter of the data is used in the testing phase, where the predicted galaxy properties are compared with those from the simulation. Of the remaining 75 percent used in the training phase, 20 percent is used for validation, computing a separate loss function alongside the one being optimized. This loss function is the mean-squared error between the true ($\{y_i\}$) and predicted ($\{f_i\} = f(\{x_i\})$) outputs:

$$\text{MSE}(\{y_i\}, \{f_i\}) = \frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2 \quad (27)$$

Finally, we find that a constant learning rate for the gradient descent process is inadequate; this initially has to be large to converge significantly, yet this leads to overshooting when approaching the loss function minimum. We use an exponentially decaying learning

rate, which reduces in value at each training epoch. Specifically, for training epoch number N , the learning rate Γ is defined:

$$\Gamma = \Gamma_0 \exp\left[-\frac{N}{N_0}\right], \quad (28)$$

where the most suitable values for N_0 and Γ_0 were found by trial and error to be 10 and 8×10^{-4} , respectively. As Γ will eventually fall to an insignificant value, the training phase is terminated once $\Gamma \leq 10^{-3} \Gamma_0$; therefore there exists a total of 70 training epochs.

4.4 Choice of architecture

The adoption of our semi-recurrent network design was the result of an amalgamation of tests of the network's ability to make basic predictions consistently. The use of a recurrent layer for temporal input quantities amended the lack of convergence of a basic neural network layout, in which all input variables were passed to a solitary dense layer. A 2D input layer also allows multiple temporal quantities to be processed in a time grid; a dense network containing all of ours would be impractical to train.

The number of hidden dense or recurrent layers was decided based on the minimum number of layers required to achieve convergence. Each input layer was followed by the optimal number of dense or recurrent layers for the network to recognize their equal contribution before the two were concatenated. The remaining hidden layers were again optimal in number, i.e. the minimum for which there was no noticeable difference in the accuracy or consistency of predictions, while gradually increasing in dimensionality to meet the number of output nodes.

To show improved convergence of a simple semi-recurrent network compared with a dense network, we incorporate two simplified designs of a basic (dense) network and its semi-recurrent equivalent. As with the main models, all layers in these networks are ELU activated and contain no dropout. Each network is trained with only star formation histories as output, and halo mass accretion history as the only temporal input variable; the dense network will not converge with the hundreds of input nodes introduced by implementing

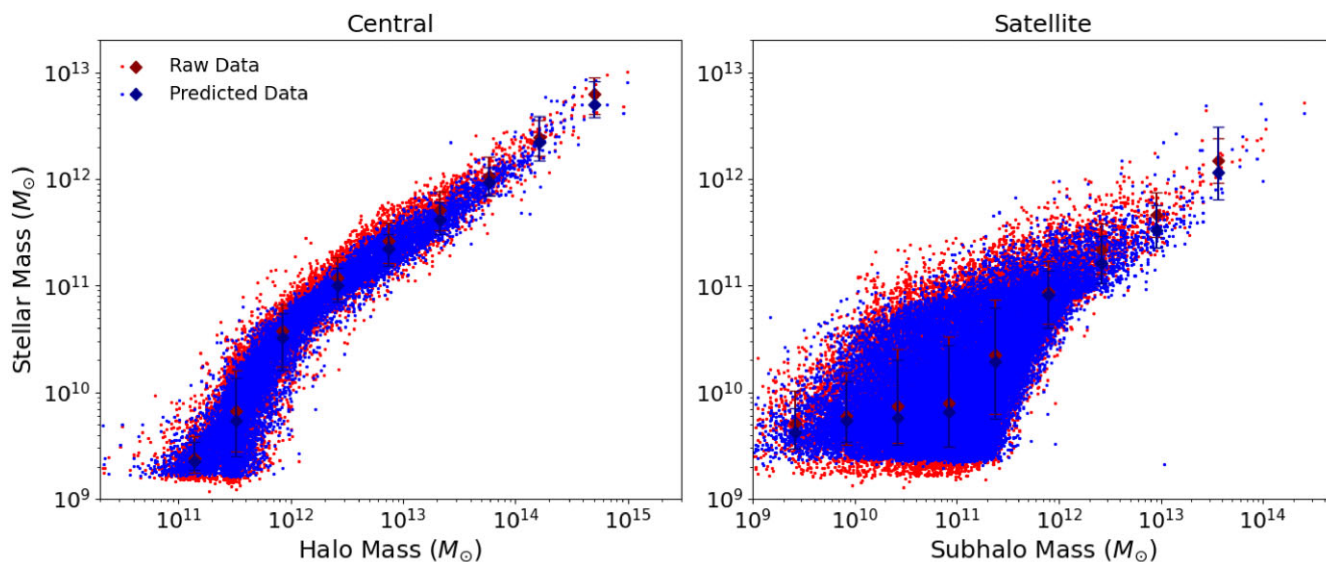


Figure 11. The numerical SHMR evaluated with the true and predicted star formation rates, for central galaxies (left-hand panel) and satellite galaxies (right-hand panel). This is shown as a function of FoF halo mass for the former, and subhalo mass for the latter. Data points from the original data set are shown in red, predictions in blue, while red and blue errorbars show the median and 15-th and 85-th percentiles of stellar mass in a given halo mass bin. The strong similarity of the shapes of these relations indicates a very good overall prediction of the star formation histories.

multiple temporal variables. The remaining input quantities include final halo mass, starting time and maximum absolute accretion rate.

The basic dense network therefore has 36 input nodes, 33 of which contain the accretion rate at different times; and 33 output nodes to accommodate the star formation history. There are 15 hidden layers from input to output, each with the same shape as the input layer.

The basic semi-recurrent network contains the accretion history in a recurrent input layer of shape (33, 1) and the remaining three quantities in a dense input layer. Following five hidden dense layers each, the two sequences merge to form a single 36-node layer, which is followed by 15 layers of the same shape before reaching the output layer.

The choice of number of layers in these models was decided as the optimum number of layers, as with the main model. Therefore this test of network design illustrates the difference in predictions of the simplest possible complete models. We use these two models to illustrate the advantage of the semi-recurrent design in Section 5.1.2.

5 PREDICTIONS

This section concerns the results of the neural network models for central and for satellite galaxies, comprising the properties of the simulated galaxy formation histories and a discussion of the physical consequences on their level of accuracy.

5.1 Galaxy properties

5.1.1 Stellar–halo mass relation

Numerical integration of the true and predicted SFH for each galaxy is used to plot the SHMR, shown in Fig. 11. Median absolute residuals in the logarithmic stellar mass are 0.079 dex for central galaxies, and 0.094 dex for satellites.

Despite the accuracy of this result, the network underpredicts the stellar mass of most galaxies. Another prevailing problem, discussed below, is the failure of the neural network to predict short bursts of star formation, which, despite their brevity, will have significant

influence on the spectroscopy of a galaxy. In general, this serves to reduce the scatter in the SHMR, and consequently this is also true for scatter in metallicity.

5.1.2 Star formation history

Fig. 12 shows the mean of the true and predicted SFHs from the simplified models described in Section 4.4, in six bins of halo mass. The general behaviour with halo mass is clearly predicted, with galaxies in higher mass haloes forming their stars earlier – i.e. these networks reproduce galaxy downsizing trends with recent star-formation shifting towards low-mass haloes.

The difference in the precision of the two networks is apparent: in each halo mass bin, the mean SFH from the semi-recurrent network is better constrained, and in most cases matches the true mean more consistently. Comparing predictions from ten independent runs, the variance in the predicted SFH at any snapshot was typically 1–2 orders of magnitude larger in a basic dense network than its semi-recurrent equivalent. The improved precision justifies our use of a recurrent treatment of historical dark matter properties.

For the main network, Fig. 13 shows the predicted and true SFH and ZH of a single galaxy in more detail. The broad shape of the SFH is well recovered, as is the stellar mass, but the NN is unable to predict the intrinsic variability of the order of 1 Gyr in the original data.⁴ The overall accuracy of the predicted geometries is visualized in Fig. 14, where we show the MWA of galaxies as a function of halo mass, for centrals and satellites separately. Here, it is easier to discern a tendency of the NN to overpredict MWAs, and to underpredict their scatter.

We find that these are general intrinsic shortcomings of the NN. The power in fluctuations on time scales lower than around 0.1 Gyr^{-1} are suppressed in the predicted data, which can explain the lower scatter in the MWAs at fixed halo mass. We will see later that this

⁴The true variability will of course have components at much shorter time scales, but is re-sampled here to 33 bins as described in Section 4.3.

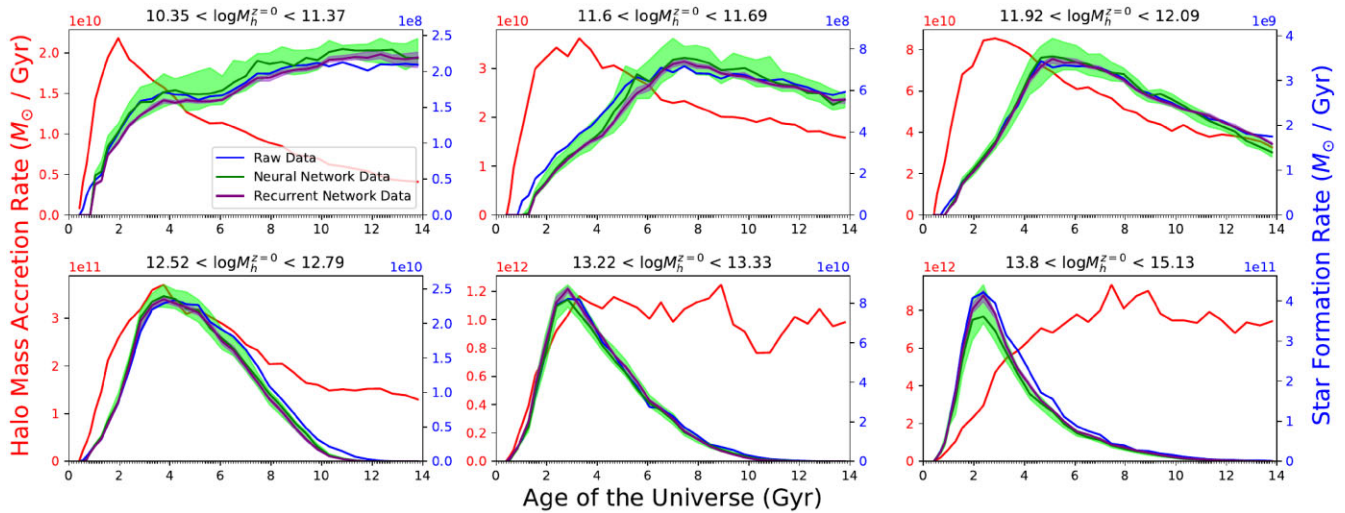


Figure 12. The shape of the mean star formation history in six bins of halo mass for the central galaxy data, where predictions from a basic dense neural network (green) are compared with those from our semi-recurrent architecture (purple). Their lines and shaded regions depict the median and interquartile ranges of the predicted mean from ten independent runs of each network. These are shown alongside the true mean star formation history (blue) and the mean mass accretion rate (red). The differences in interquartile ranges between predictions of the two networks shows a significant improvement in the converging power of the network when implementing the semi-recurrent design.

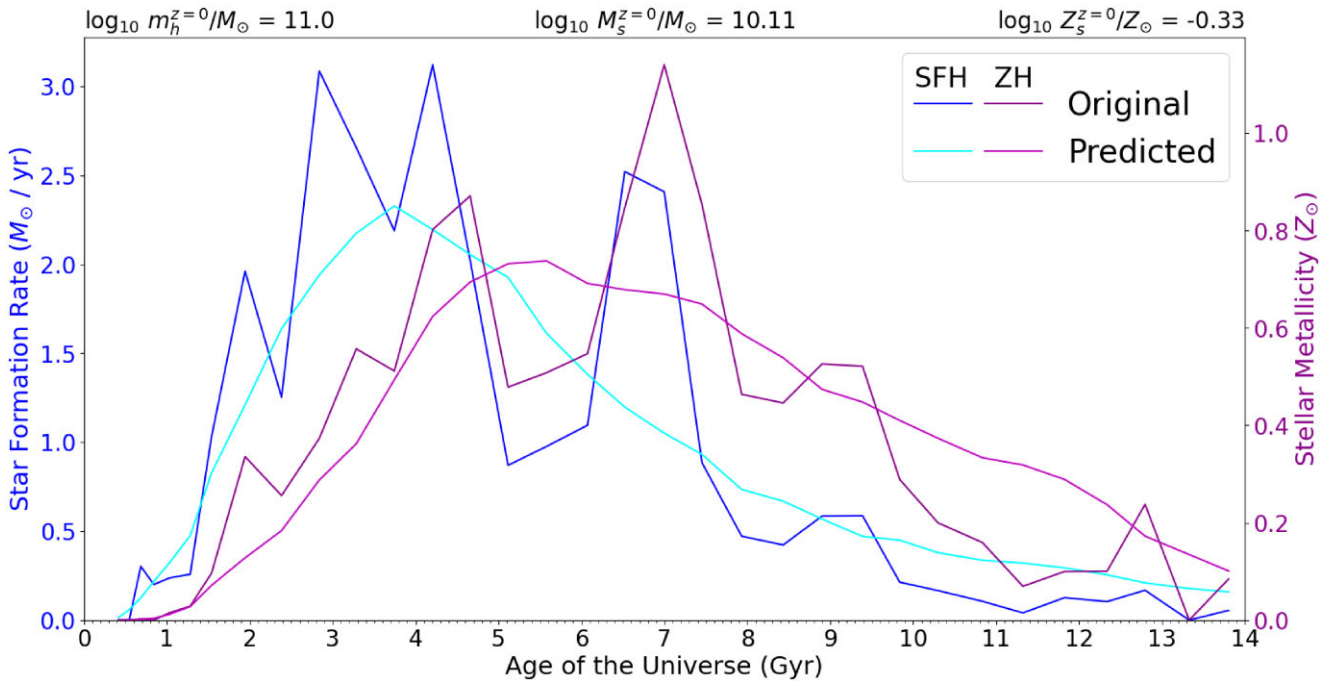


Figure 13. An example of formation history for a satellite galaxy of intermediate mass. The true star formation history (blue) and the corresponding prediction (cyan) are shown alongside the true stellar metallicity history (purple) and the predicted metallicity history (magenta). Numerical estimates of the subhalo and predicted stellar mass and metallicity are shown above the graph. This sample shows a modest fit to the shapes of the SFH and ZH, but not the variations on short time scales.

suppression is also correlated with discrepancies between predicted and true SEDs, and propose a possible way to overcome it.

In Fig. 15, we show how the time at which half the final (sub)halo mass is formed varies according to the SHMR for centrals and satellite galaxies, in both original and predicted data sets. This is an important aspect of the SHMR which illustrates the relation of speed of halo growth to stellar mass (Cui et al. 2021). While this quantity is not an explicit feature of the neural networks, its relationship with

the SHMR is preserved in the predictions of the neural network. This shows that the causal dependence of halo growth on star formation is captured by the neural network.

5.1.3 Chemical enrichment history

Fig. 16 depicts the relation between halo mass and mass-weighted metallicity of each galaxy. The scatter in metallicity is underesti-

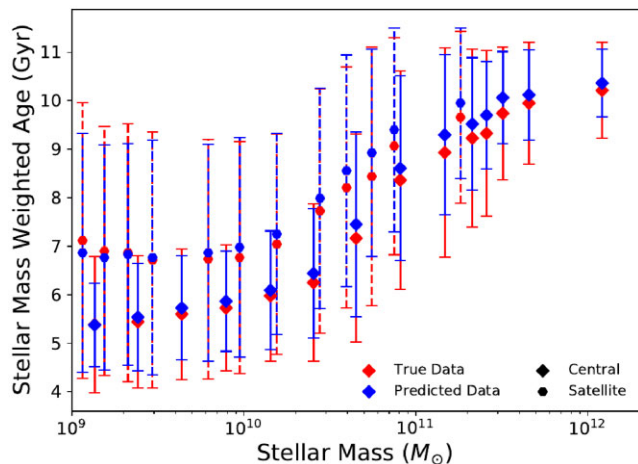


Figure 14. Stellar mass weighted age as a function of stellar mass, for central (solid, diamond) and satellite galaxies (dashed, hexagon). The plot points and error bars represent the median and interquartile range of mass weighted ages in each bin. The ages derived from the predicted stellar mass assembly history (blue) are overlaid upon the values computed from the original data, showing that the general trend of ages with respect to masses are well-matched, however the offsets and reduced ranges of certain error bars are indicative of the prevalence of samples with considerable differences in mass assembly geometry.

mated by the neural network, to a lesser extent for satellite galaxies. Metal-rich objects in particular have lower metallicities across stellar masses. We also see a lack of high frequency information in metallicity histories such as the example in Fig. 13.

We argue that this overall lack of high frequency information of star formation and metallicity histories is related to reduced scatter in both physical and observational summary statistics. This is discussed in Section 5.2.2.

5.2 Observational data

5.2.1 Spectra and line emission

Fig. 17 shows the mean of the stacked galaxy spectra, in bins of stellar mass, with the shaded regions indicating the standard deviation of the set. The shape and amplitude of the spectra in most bins are consistent in both the original and predicted galaxies. In high mass galaxies, the NN tends to underpredict the mean luminosity and scatter at short wavelengths, and at lower masses, it slightly overpredicts the luminosity at all wavelengths.

The spectral energy distributions for central galaxies exhibit clearly underpredicted mean and standard deviation compared with satellites. While the continuum and line emission features of these spectra tend to change in the correct way with halo mass, the mean amplitude of the predicted luminosities is often below of the true mean. This discrepancy may be attributed to stellar age and metallicity dependence on the mass-to-light ratio of an SSP (Gallazzi & Bell 2009). We will show how this is associated with the difficulty of the NN in predicting SFHs on short time scales. While this is also a problem for satellites, satellites are more likely to be quenched, and are therefore less susceptible to shortcomings in predicting SFHs on short time scales.

Although not shown, the numerical estimates of $H\alpha$ line luminosity behave similarly. The predicted distribution of luminosities is similar to the original data, yet we consistently underestimate the scatter in any given mass bin. This result suggests that line emission

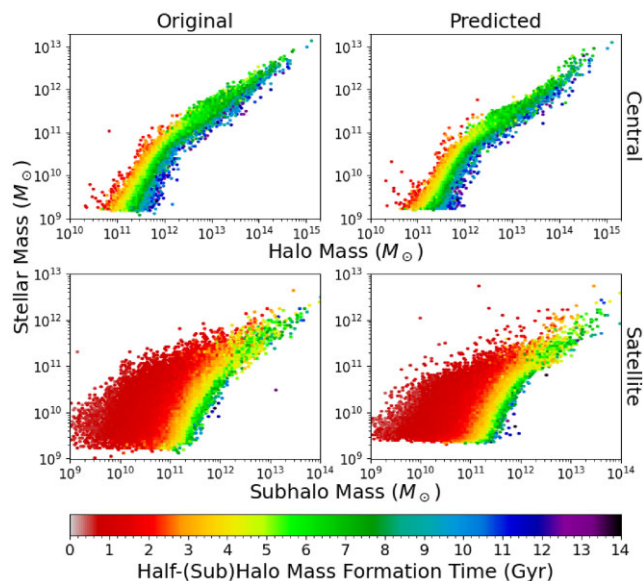


Figure 15. Stellar–halo mass relations for central galaxies (top panels) and satellites (bottom panels), in 2D bins coloured according to the mean cosmic time per bin at which half of the final mass of the halo (central) or subhalo (satellite) is formed. Each SHMR is shown separately for the original TNG data and the networks’ predictions by plotting according to respective stellar masses. The same trend of this property along the SHMR in both circumstances shows that this relationship is reflected in the network’s predictions. Note that the low occupancy of bins on the edges of the SHMRs are sensitive to minor differences in scatter, which results in an apparent distortion of the shape of the SHMR, which is misleading.

luminosities can be predicted, yet are sensitive to the subtle variability in the galaxy’s evolution, which serves to reduce their scatter at fixed stellar mass or star formation rate.

5.2.2 Residual luminosity and stochasticity

We show the dependence of the spectra of central and satellite galaxies on short time scale star formation events by calculating the mean of the fractional difference between the Fourier amplitudes of the true and predicted SFHs, from 0.3 Gyr^{-1} to the Nyquist frequency of approximately 1.2 Gyr^{-1} , and correlating these with the fractional difference in total luminosity in Fig. 18, and $H\alpha$ line luminosity in Fig. 19. These are each shown for galaxies of high star formation rate at $z = 0$. The frequency range of this data corresponds to variations on time scales from 0.8 to 3.3 Gyr.

In all of these plots, there is a clear correlation between the residuals in high-frequency Fourier modes of the SFHs and those in their derived luminosities. This indicates that there is a significant contribution to the spectra from short star formation variability, which the networks seldom predict. This is apparent in terms of the total luminosity, which is sensitive to the size of these high frequency attributes, and the $H\alpha$ luminosity, which is sensitive to recent star formation.

5.2.3 Photometry

The small variance in predicted luminosities translates into a narrower range in band magnitudes. Each magnitude is plotted against the galaxy’s stellar mass in Fig. 20, where despite the likeness of the distributions, the ‘scatter’ in magnitudes is smaller, particularly for

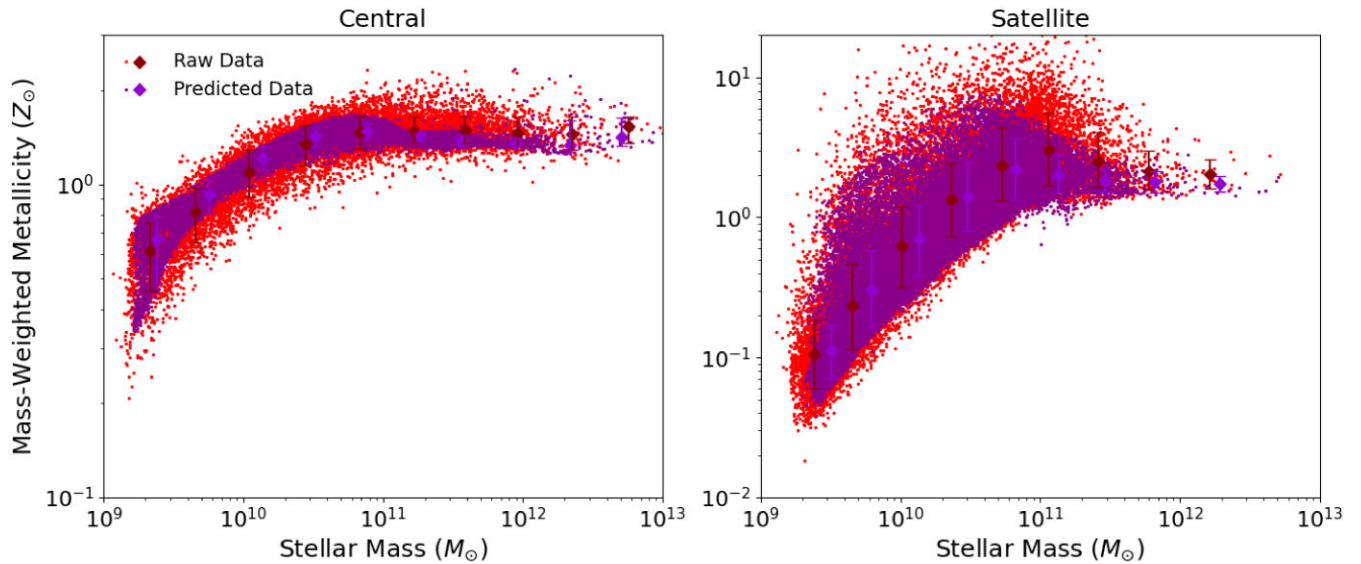


Figure 16. The numerical MZR evaluated with the true and predicted chemical enrichment histories, for central galaxies (left-hand panel) and satellite galaxies (right-hand panel). The true and predicted metallicities are, respectively, shown as a function of true and predicted numerical stellar mass. Data points from the original data set are shown in red, predictions in purple, while red and purple errorbars show the median and 15-th and 85-th percentiles of stellar metallicity in a given stellar mass bin. Like Fig. 11, the relation is well matched, however in this figure some extremities in scatter are unreached.

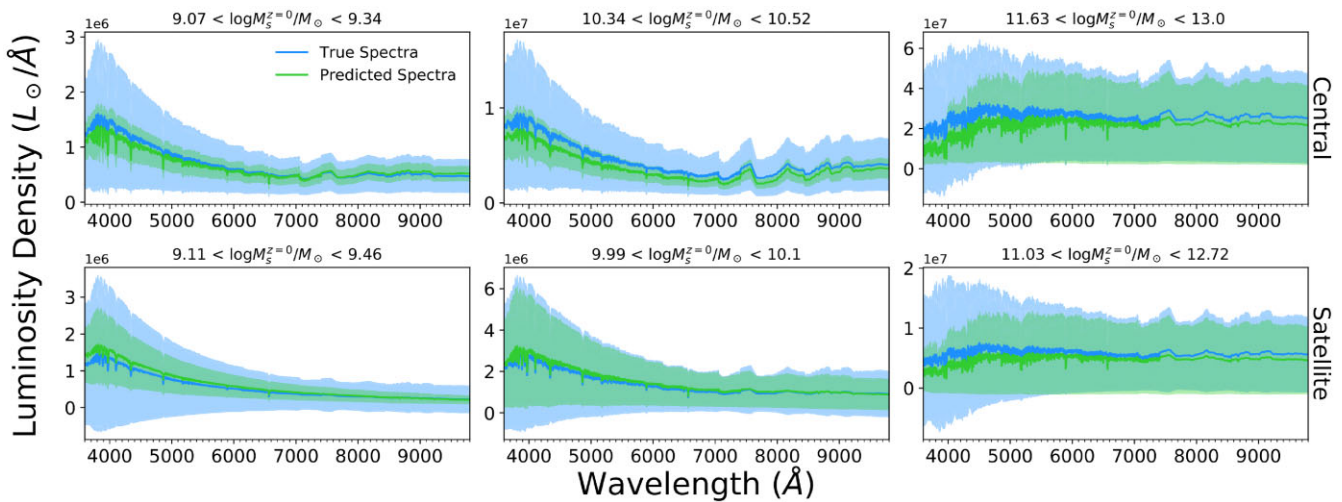


Figure 17. The mean and standard deviation for stacked central (top row) and satellite (bottom row) spectra in bins of stellar mass, shown for predicted star formation and metallicity histories in green, and TNG data in blue. Emission lines have been omitted from these plots for clarity. In the majority of samples, the continuum is generally well recovered, and is of similar amplitude. However, for high mass objects there is a reduced variance at short wavelengths, and lower mass galaxies have a smaller variance overall. This represents a poorer prediction of central galaxy spectra, with lower mean amplitudes and smaller variance than the spectra evaluated from TNG data.

central galaxies. Colour distributions are shown in Fig. 21, depicting the expected colour bimodality in multiple band differences, for both central and satellite galaxies. The network models therefore distinguish ‘blue’ star-forming galaxies and ‘red’ quiescent galaxies in their predicted galaxy formation histories.

The increased difficulty in predicting the luminosity at shorter wavelengths that is shown in Fig. 17 obviously translates to systematic offsets in the bluer photometric bands. We see this in colour distributions such as $u - g$ and $g - r$, where many red galaxies are shifted towards bluer colours. Despite offsets in colours evaluated at high mass, the general inclination of galaxy colour with regard to mass is such that high mass galaxies are redder.

We showed in Section 5.2.2 that the error in the total luminosity of a given galaxy can be likened to the absence of high frequency modes in the star formation rate. It has been shown that these short star formation events impact the accuracy of photometric colours, particularly when they occur at recent times (Chaves-Montero & Hearin 2021; Fraser, Tojeiro & Chittenden 2022). We show that the error in $g - r$ colour scales with residual Fourier modes in Fig. 22, and see that the behaviour of this correlation is similar to the effect on luminosities; but unlike the luminosity errors, this correlation is also visible for the Fourier modes in metallicity history. Noticing that some absolute colour residuals are similar in size to visible distortions of the $g - r$ distributions in Fig. 21, we see that the lack of short

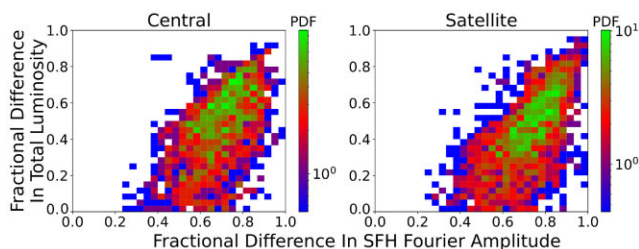


Figure 18. 2D histograms of the fractional difference between true and predicted total galaxy luminosity and the mean high-frequency Fourier amplitudes of their star formation histories. These are shown for galaxies between the 75-th and 95-th percentiles of $z = 0$ star formation rate, and shows data within a frequency range of $0.3\text{--}1.2\text{ Gyr}^{-1}$, i.e. a time scale range of $0.8\text{--}3.3\text{ Gyr}$. This correlation between the two residuals indicates the dependence of the calculated luminosity on high-frequency star formation events.

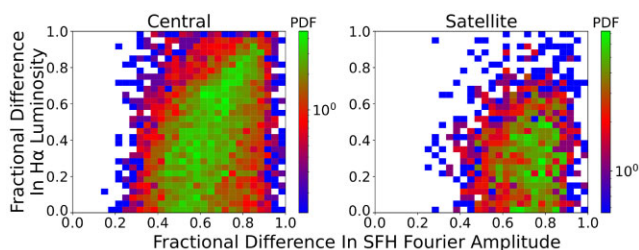


Figure 19. For the same galaxies as in Fig. 18, this figure shows the correlations between residuals of their high frequency SFH data and their total $H\alpha$ line luminosity. This indicates the importance of measuring short time scale star formation events as in Fig. 18, in particular at low stellar ages with the largest contribution of ionizing photons.

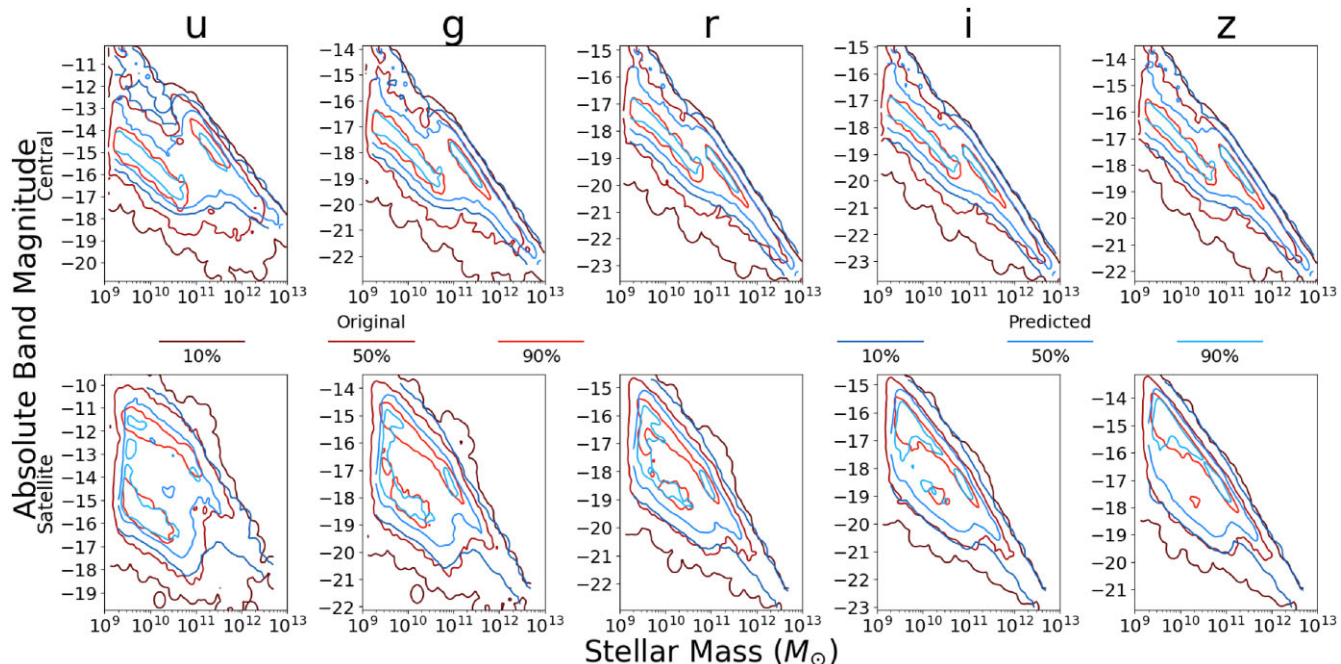


Figure 20. Estimates of the five SDSS band magnitudes from the true and predicted spectra of both central and satellite galaxies, shown as a function of stellar mass, with contour lines indicating percentiles of their 2D distribution. These show a reasonable similarity in all bands despite a slight reduction in the variance of magnitudes in the predicted data. In both central and satellite data, the bimodal distribution of magnitudes can be seen in relation to mass.

time scale events may explain distortions such as narrower peaks of the distribution. However, high-frequency ZH features are rare in the metal-rich galaxies whose metallicities are underpredicted, and the error in their colour may instead be due to inaccurate star formation histories.

6 IMPORTANT FEATURES

In this section, we discuss our methods of identifying the input features of the network which have the greatest predicting power over the star formation and metallicity histories.

A commonplace metric such as a random forest regressor is not useful for addressing the importance of historical features; the forward-feeding influence these variables make on the final result make it difficult to compute a metric as a single number. Instead, we perform a test in which the network is trained multiple times, while groups of similar features are scrambled, in an effort to eliminate their signal.

This method is similar to permutation importance, in that we compare the performance of the model after randomizing data subsets. However, it differs in the sense that the summary statistics of the disrupted model are derived from predictions, and compared with those calculated from the fiducial predictions. This allows us to measure physical properties of the data after scrambling, and identify the importance of the randomized quantities on these properties, whether or not they are explicitly given as model parameters.

6.1 Shuffle groups

We evaluate the properties of the SHMR, the halo mass–metallicity relation (HMZR), and metallicity history independently of star formation history, after randomizing one of these features by replacing the training and testing data with Gaussian random noise, and training the network with this data.

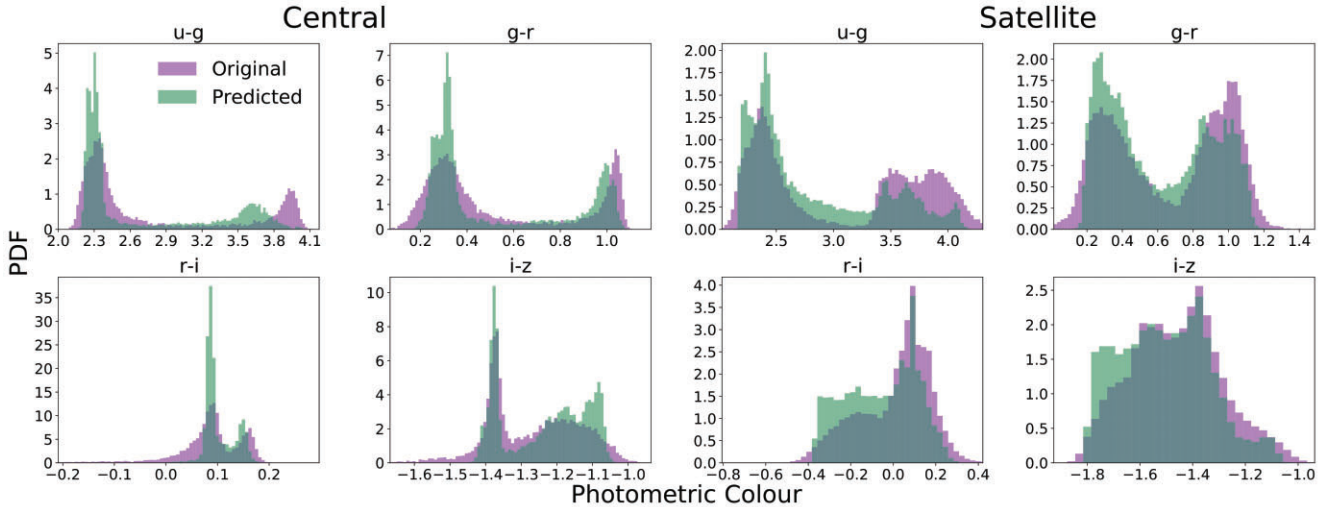


Figure 21. Photometric colour distributions across the five bands, showing the differences between two consecutive bands. The distributions, mostly bimodal, are shown in purple for the original data set, with those derived from predicted spectral energy distributions in green. We see rough agreement between data sets, however there are clear offsets in some of the data, such as bluer red galaxies in $g - r$, and significantly smaller predicted ranges.

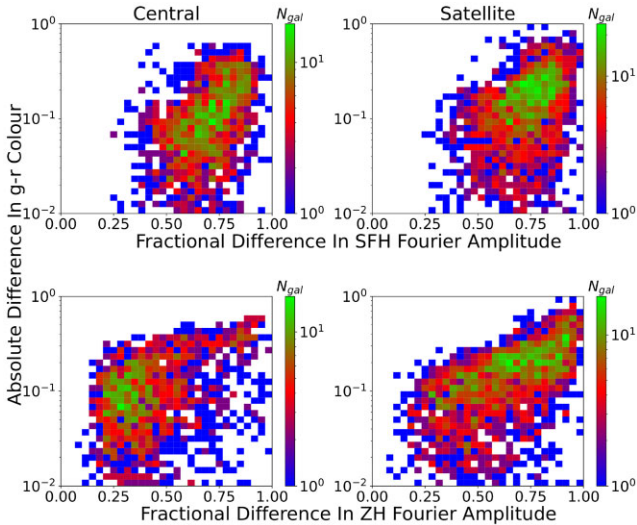


Figure 22. For the data shown in Figs 18 and 19, this figure shows 2D histograms of the absolute difference between true and predicted $g - r$ colours and the mean high-frequency Fourier amplitudes of their star formation histories (top row) and metallicity histories (bottom row). This clear correlation indicates the importance of measuring short time scale star formation and chemical enrichment events in the aim to calculate accurate colours.

There are two important aspects to this strategy. First of all, we choose these aforementioned relations in order to measure the effect on the scatter of a baryonic quantity at a fixed halo mass, thereby interpreting the effect that the shuffled quantities have on star formation and metallicity history, independently of halo mass. Secondly, we scramble data such that any directly related properties (e.g. δ_1 and δ_3) are also randomized, temporal or otherwise. Thus, each randomization is a test of a single group of interrelated features, which we term ‘shuffle groups’.

The shuffle groups to which each variable belongs are listed in Table 1. In summary, group 1 pertains to halo growth and contains accretion rate and final mass. Group 1a is unique to the satellite

network as it contains the properties related to the satellite subhalo, while group 1 refers to its host. Group 2 contains environmental features such as overdensity and cosmic web distances, group 3 relates to dynamical properties such as circular velocity, and group 4 relates to interactive variables such as the skew.

6.2 Scatter divergence

For ten independent runs of each randomized network, we compare the median predicted scatter to that of the un-randomized network, as a function of halo mass for central galaxies, and a function of subhalo mass for satellite galaxies. Instances where the median scatter is clearly offset from the fiducial prediction show where the scatter is supported by a property of this shuffle group.

The deviation in scatter is quantified by calculating the quantity Ξ :

$$\Xi \equiv \left\langle \log_{10} \left| 1 - \frac{\sigma_x}{\sigma_0} \right| \right\rangle, \quad (29)$$

where σ_x is the characteristic scatter of the given relation when shuffle group x is randomized, in dex, and σ_0 is that of the fiducial result. Ξ equates to $-\infty$ if σ_x and σ_0 are identical, otherwise a larger value of Ξ represents a stronger difference between scatters. The scatter is compared with the fiducial result and not the true data in order to discern the effect of scrambling features from the general inaccuracies of the model.

The angled brackets indicate that we have averaged this logarithmic ratio, over 80 loguniform bins of halo mass and weighting these bins according to their occupancy. This sample weighting is done to minimize biases from sample size limited areas, otherwise producing a misleading picture of the general size of the scatter. Additionally, Ξ values can be biased towards small values if the difference in scatter frequently changes sign; thus we discard samples where the local Ξ value is less than -2 .

In addition to the scatter in the SHMR and HMZR, we compute the scatter in the unweighted mean metallicity history, identifying dependencies on chemical enrichment independently of stellar mass, and thus for minority stellar populations such as late starbursts in early-type galaxies.

		Central				Satellite				
Galaxy-Halo Relation	SHMR (M_s)	-1.57 ± 0.15	-1.67 ± 0.18	-1.5 ± 0.11	-1.66 ± 0.21	-1.86 ± 0.1	-1.71 ± 0.12	-1.75 ± 0.07	-1.63 ± 0.21	-1.91 ± 0.25
	HMZR (Z)	-1.68 ± 0.12	-1.6 ± 0.12	-1.53 ± 0.28	-1.38 ± 0.1	-1.77 ± 0.22	-1.64 ± 0.22	-1.7 ± 0.03	-1.74 ± 0.24	-1.82 ± 0.14
	ZHMR (ZH)	-1.62 ± 0.28	-1.77 ± 0.1	-1.63 ± 0.14	-1.29 ± 0.08	-1.85 ± 0.23	-1.67 ± 0.13	-1.6 ± 0.05	-1.67 ± 0.29	-1.83 ± 0.17
		1	2	3	4	1	1a	2	3	4
		Shuffle Group								

Figure 23. Ξ values of the stellar–halo mass relation (top row), mass–metallicity relation (middle row) and mass–metallicity history relation (bottom row), for each shuffle group in the central model (left-hand panel) and satellite model (right-hand panel). The text in each cell shows the median and interquartile range of Ξ values obtained from ten independent runs of each network. To highlight the most significant shuffle groups, grid cells with smaller Ξ values are coloured in dark blue, transitioning into bright green for higher values. Higher Ξ values represent a greater importance of the given shuffle group as the model determines the given galaxy–halo relation.

As well as a Ξ quantity to measure differences in scatter, we apply the same method to evaluate the discrepancy in the median-filtered relations, thereby measuring the accuracy of the fit to the relations themselves. For this we replace the scatter in equation (29) with the logarithmic median of the given baryonic quantity. This Ξ analogue is not significantly affected by any particular shuffle group, signifying no quantity with distinctive effect on the amplitudes of the three relations.

6.3 Results

All values of Ξ for each network, and each baryonic quantity, are tabulated for centrals and satellites in Fig. 23, where we show the median and interquartile range of Ξ values obtained from ten runs of the neural network. Values with large median Ξ are considered significant, however a low interquartile range suggests that this deviation is a systematic result of the network, while higher interquartile ranges indicate poor convergence.

For central metallicity and metallicity histories, shuffle group 4, primarily involving skew, has a noticeable effect on the scatter. The skew therefore plays an important role in determining the dispersion of our metallicity histories, which may correspond to small-scale accretion events which lead to chemical enrichment. For satellites, groups 1a, 2, and 3, incorporating satellite history, overdensity, and circular velocity, are the groups which affect ZH significantly, however circular velocity is less precise. One can also see notable deviations in the HMZR scatter when scrambling groups 2 or 1a, suggesting that accretion and local environment are important to satellite enrichment.

For central galaxies, mass accretion history and circular velocity have notable offsets on the scatter of the SHMR, while circular velocity also has an effect on the satellite result. The robustness of the SHMR may be due to inherent correlations between some variables in different groups, such as the two aforementioned.

For satellites, the network performs similarly with either one of the halo and subhalo mass histories being randomized, with little change to the SHMR if only one is randomized. While the final masses and growth histories are important to the mean SHMR, their shuffle groups, containing infall parameters such as scaled infall time, have only a subtle influence on their scatter.

One can reconcile the halo growth rate and circular velocity with the SHMR scatter, as the halo’s early growth and internal dynamics will quickly determine the growth of its galaxy, which restricts

the feasible scope of future star formation rates. The future of its chemical enrichment can be more closely linked to environmental changes, which can be attributed to the abundance of star-forming or metal-rich gas in overdense regions, where the radial distribution would indicate the tendency for this to be accreted into the target galaxy. However, the expulsion of star forming gas owing to internal feedback mechanisms can impact the metallicity as well.

6.4 Median value divergence

We compare the distributions of mass, metallicity, and MWA in an intermediate halo mass bin for each network and each randomization in Fig. 24, shown next to the median of ten ordinary predictions, and the true data. As the network marginally underpredicts the scatter in stellar mass, and underpredicts metallicities, the true distributions are distinct from our predictions: histograms of stellar mass are marginally narrower in our predictions, and metallicity histograms are slightly offset. If the distribution from any randomization is closer to the true distribution, it suggests that a member of this shuffle group is misleading the network, while larger differences suggest the group contains necessary information.

We find no evidence of a shuffle group which improves the network’s performance if scrambled, as is the conclusion from evaluating a Ξ analogue for median data values; however there is a small difference in the shape of satellite MWA distributions when group 3 is randomized. While the scatter ratio shows the dependence of other shuffle groups in different halo mass regimes, this analysis of data in a narrow halo mass range shows that shuffle group 3 is additionally important for distinguishing similar samples. It is not explicit, however, whether this dissimilarity serves to deform the predictions, indicating that the shuffle group contains a necessary detail of the GHC; or whether these features correlate with the TNG distributions, which would imply that shuffle group 3 deceives the satellite network.

6.5 Subgroup randomizing

One can scramble individual quantities or subsets of the shuffle group, which may have an important influence on the galaxy’s formation history, in spite of physical correlations within the shuffle group. In Fig. 25 we show this result for two examples, one for each network, compared with the result from shuffling the remaining shuffle group members and the full shuffle group.

We test scaled infall time, formation time, and infall mass ratio in the satellite network. While these may be important quantities for satellite SFH, in Fig. 25 we see that the difference between predictions with these infall parameters and the rest of shuffle group 1a is small, suggesting that they have little effect on the performance of the satellite network. This may owe to the onset of the target’s satellite phase being inferred from the growth histories of the subhalo and its host. On the contrary, the deviation from scrambling infall parameters is larger, indicating that it is useful to utilize them explicitly. The metallicity deviation is also larger for infall only, however poorly constrained, which may indicate their importance.

Donnan et al. (2022) indicate that star formation histories and gas phase metallicities in TNG are modulated by the cosmic web. We randomize cosmic web distances in the central network, and compare the predicted scatter with the result when scrambling overdensities as the remaining data of shuffle group 2. Though not shown here, shuffling only the cosmic web distances has a noticeable effect on the scatter in the mass-weighted metallicity of a handful of high mass galaxies. Yet in our Ξ tables, showing the overall effect, the cosmic

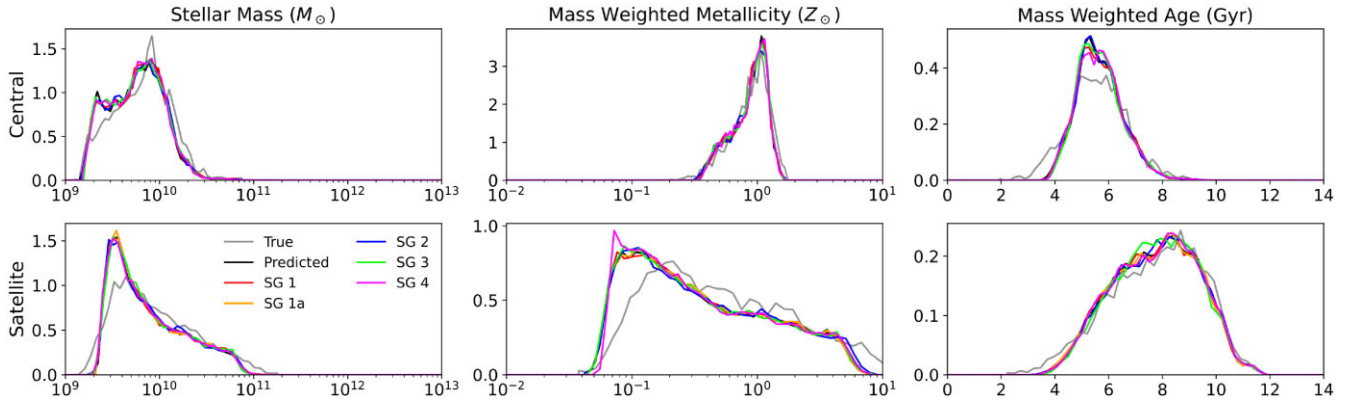


Figure 24. Smoothed PDFs of stellar mass, metallicity, and MWA derived from star formation and metallicity histories in an intermediate halo/subhalo mass bin. This shows the distributions obtained from the original TNG data in grey, the median from ten fiducial predictions in black, and the median result from each randomization in coloured lines. The distribution from each randomization is subtle and does not show a clear difference from the best predictions, nor do they bring the network closer to the true result. If the x-axis is logarithmic, the PDF is a function of the logarithmic value.

Galaxy-Halo Relation	Central			Satellite		
	SG 2	CW	5 Shuffled Data	SG 1a	Infall	Not Infall
SHMR (M_s)	-1.67 ± 0.18	-1.91 ± 0.22	-1.69 ± 0.17	-1.71 ± 0.12	-1.72 ± 0.2	-1.71 ± 0.23
HMZR (Z)	-1.6 ± 0.12	-1.76 ± 0.24	-1.58 ± 0.21	-1.64 ± 0.22	-1.65 ± 0.36	-1.7 ± 0.28
ZHMR (ZH)	-1.77 ± 0.1	-1.83 ± 0.24	-1.79 ± 0.24	-1.67 ± 0.13	-1.73 ± 0.4	-1.7 ± 0.23

Figure 25. Ξ values as in Fig. 23, instead showing the impact of shuffling certain subsets of a shuffle group. In each table, the left-hand column shows the result for the full shuffle group as in Fig. 23. The centre column shows the result for the data of interest, while the right-hand column shows the result for the remainder of the shuffle group. This shows overdensity to be the dominant component of the second shuffle group for central galaxies, while infall parameters have a noticeable effect in satellite galaxies.

web is insignificant, and the offset is larger and applies to most galaxies when randomizing overdensity. This, along with the larger interquartile ranges of Ξ for the cosmic web, would suggest that the network favours the overdensity components when effectively constraining the mass-weighted metallicity.

7 DISCUSSION

This NN model has been shown to infer the main trends of galaxies' star formation and metallicity histories with halo mass, and to broadly reproduce key observational results, such as downsizing and galaxy colour bimodality. In this section, we reflect on shortcomings, insights, and applications of the model.

7.1 Baryonic information

A possible factor limiting the predictability of the SFH is the lack of information of gas properties, such as mass, metallicity, and temperature, all of which are critical in the star formation process. Our model will predict any correlations with dark matter environment and stellar assembly for which gas properties are causally intermediate, however unconnected gas features which influence star formation may remain unrecognized. Studies show the importance of the gas content of merging haloes on star formation (Hani et al. 2020; Trevisan et al. 2021); something not modelled in our analogues for merger histories.

The metallicity of the gas depends on its location in the cosmic web, and is governed largely by gas fractions and specific inflow rates, independently of halo or stellar mass (Torrey et al. 2019; Van Loon, Mitchell & Schaye 2021; Donnan et al. 2022). However, our cosmic web distances carry little weight in predicting the SFH and ZH, likely due to the expectedly low sample size at fixed mass, or the fact that halo mass accretion histories and other quantities are better predictors of the shape of the SFH. Galárraga-Espinoza, Garaldi & Kauffmann (2022) argue that the effect of cosmic web filaments on star formation either enhance or quench galaxies depending on their scale, while the number of small filaments connected to the galaxy is a more robust measure of star formation enhancement. In this work, the larger deviation seen from scrambling overdensity suggest this is a stronger constraint than the cosmic web for modelling the metallicity history of central galaxies.

Additionally, the importance of the baryonic content of merging subhaloes would suggest that we may have captured star formation events more accurately if we had historical information of several of its most massive progenitors. However, designing such a network would prove difficult. Earlier attempts to characterize the merger history with a single variable in this work, such as median merger ratio per snapshot, were ineffective in improving the network's performance. A model may exist in future which predicts the baryonic growth of a halo's most massive progenitors, which by correlation with the MPB can predict the amplitude and time scale of merger-induced starbursts. Alternatively, one may use a temporal analogue to the specific merger rates evaluated for discrete time intervals in Dhoke & Paranjape (2021), or direct computation of a functional form for merger rates, relating to overdensity and the highest progenitor masses (Fakhouri & Ma 2009).

7.2 Stochastic star formation

With short time scale star formation events being difficult to model, it may be worthwhile implementing a stochastic description of the SFH, which has been achieved on Myr time scales (Tacchella, Forbes & Caplar 2020). This stochastic modelling is accomplished by computing the power spectrum of the SFH through implementing variable gas inflows and molecular cloud formation while conserving the system's mass. This power spectrum captures the time scales of these processes, allowing one to constrain the lifetime of molecular clouds and star-forming regions. In our data, the power spectra of

our predicted SFH and ZH are weaker at high frequencies when compared with TNG data. However, a network model in which we predict this spectrum directly may provide a statistical model for star formation on smaller time scales, potentially offering a stochastic amendment to observables.

We tested a network for central and for satellite galaxies, in which the same dark matter properties and network architecture were used to predict the amplitude of the Fourier transforms of the SFH and ZH, i.e. the square root of their power spectra. The shape of the spectrum and the scaling of its amplitude with halo mass are well recovered, with similar levels of accuracy as the SFH and ZH themselves. It is therefore possible to develop a model which could predict these power spectra and generate a stochastic signal to mitigate the errors in our predictions of SFH and ZH. The correlations between residual Fourier transforms and luminosities and colours seen in Section 5.2 suggest that this can help obtain more realistic galaxy SEDs.

7.3 Metallicity

We find that the complete networks roughly predict the mass-weighted stellar metallicity of all galaxies as a function of time, recovering the relations between stellar mass and metallicity, albeit with smaller scatter. High metallicity objects have their metallicity underestimated, which can be attributed to low sampling of the most metal-rich galaxies, or poor prediction of the star formation variability which weights it.

The omission of time-dependent overdensity, circular velocity, or radial skew undermines the ability to predict the scatter in the final metallicity and ZH variations. Overdensity and skew are environmental variables, providing measures of the concentration of the local mass distribution, and the density bias towards or away from the target subhalo, respectively. As functions of time, they measure the development of the most significant subhalo interaction events.

The influence of the skew on ZH scatter may be tracing the infall of subhaloes, whose relation to the target’s mass and environment adds to the likelihood of the halo progenitor hoarding star-forming or metal-rich gas. The importance of overdensity in determining metallicities would suggest that chemical enrichment owes to the abundance and properties of interacting subhaloes. For centrals, such a relationship may be a consequence of strong correlations of overdensities with the metal richness of gas inflows and the intergalactic medium (Peng & Maiolino 2013), while the tendency for overdensity to capture major interaction events which accelerate star formation (L’Huillier, Park & Kim 2015) may also contribute to the galaxy’s metallicity in future. However, overdensity appears more important for satellite metallicities. This significance may be attributed to environmental quenching, which is not a common effect on central galaxies, whose quenching is strongly correlated with the galaxy’s mass (Bluck et al. 2020). It is possible that the effect that overdensities have on central metallicities is captured by another shuffle group.

The skew, on the contrary, is weighted by the mass distribution and is independent of the scaling of the subhalo masses involved. Instead it can be considered a measure of the trajectory and frequency of subhalo interactions, most of which are minor. This could explain why overdensities are important to the galaxy’s mass-weighted metallicity, but the skew has notable secondary effects on the metallicity history.

The circular velocity also influences the metallicity scatter, albeit, with lower significance. This quantity strongly correlates with subhalo mass, yet it has a strong connection to the time scale of internal collapse, and thus the onset of AGN feedback which expels

star-forming gas (Davies et al. 2019), with higher values indicating rapid collapse and high halo concentration, thereby creating an environment which can alter the likelihood of metal production. This scaling with halo concentration, hence the star formation efficiency, is also dependent on environment and cosmic web distances (Hellwing et al. 2021), so this is far from an independent measure of these dynamics. Yet the omission of \tilde{v}_{vir} distorts the predicted HMZR scatter, suggesting the importance of dynamical measures of metal synthesis. Furthermore, shuffle group 3 contorts the distribution of metallicities on sample scales; a unique property of this shuffle group. While it is unclear whether this last distinction highlights a physical or a deceptive correlation, this could suggest that the dynamical nature of the galaxy–halo connection is the most independent of halo mass.

If the network is run with the baryon-biased maximum circular velocity in place of our virial approximation, the underprediction of high metallicity objects is reduced, and the MZR is fit more accurately. The full enrichment of these objects cannot be determined by our dark matter quantities and is additionally dependent on a purely baryonic phenomenon. Our model will undermine the metallicity scatter in a pure dark matter application. This result may prove more fruitful if our networks were trained on a semi-analytic model simulation, if this model returns suitable metallicities despite the dark matter determined rotation curve.

The dark matter half-mass radius, also a structural quantity, does not make a difference to predictions in metallicity, which has been shown to be the case at least at $z = 0$ (Lovell et al. 2021). The physical size of the halo is evidently insignificant next to its central density. This quantity does, however, appear important for predicting gas mass and instantaneous star formation rate (Lovell et al. 2021), which in principle constrains galaxy growth in the long term, however this is already dominated by mass accretion and other variables.

7.4 Stellar mass and star formation history

Building our networks with relatively few parameters, or randomizing several quantities, does not adversely influence the predictions of the SHMR or star formation histories. The differences in scatter or halo-binned stellar mass distributions are minimal between runs with differently randomized training data.

Final halo mass and circular velocity are powerful parameters to determine the stellar mass. Historically, stellar mass and SHMR scatter depend greatly on the origin time and the halo mass accretion rate. The difference with satellites is that while the halo mass is most crucial to the stellar mass in the time-independent framework, it is the subhalo’s mass accretion history which results in greater errors for the neural network as it is randomized. The properties and environment of the host can be attributed to the quantity of satellite galaxies and their evolution post-infall (Bose et al. 2019), therefore there may be a future study in which the assembly bias of the host halo is used to determine satellite formation history.

Quantities pertaining to the infall of satellite subhaloes are also important for their SFH. Though infall time is a significant factor in determining the SFH, as this will restrict the onset of satellite physics, we argue that the relationship between a_{max} and a_{infall} is a valuable asset to our model; the two share the half-mass formation time as a denominator, and therefore this relates the time of infall to the accretion history. Strictly, if $a_{\text{max}} > a_{\text{infall}}$, then the subhalo continues to grow after becoming bound to a larger halo, advocating a greater star formation rate in the satellite phase. On the contrary, the infall time along with t_{start} can constrain this relationship between central-phase accretion and post-infall star formation.

We have included starting time and scaled accretion times along the satellite’s MPB, yet the formation times affiliated with the main halo are known to constrain the population of satellite haloes and the properties of the galaxies they host (Artale et al. 2018; Bose et al. 2019). While these times concerning the host halo are likely to be captured by its mass formation rate, evaluating this explicitly in our model may indicate the typical properties of satellite galaxies within an early or late forming halo.

Inspecting star formation histories, it appears that the time at which star formation begins to decline and the subsequent gradient of the SFH are well matched for satellite galaxies. On the one hand, the mass ratio upon infall is likely to indicate the subsequent rate of stripping of the satellite galaxy, while the scaled infall time relates the time at which this ratio is achieved to the central-phase growth of the satellite galaxy. It would appear, however, that the deduction made by scaled time factors can be inferred by the combined growth histories of the subhalo and its host. The conclusions from Shi et al. (2020) indicating the clear distinction in the star formation rates linked to rapidly and slowly growing subhaloes could be derived from their temporal implementation, while the merger histories which also correlate with their scaled formation time can be inferred from overdensity and skew histories.

It may be worthwhile in future designs to include properties relating directly to quenching, particularly given that this is unconstrained for central galaxies. A halo’s internal velocity dispersion proves fruitful in classifying quenching central galaxies, being tightly related to black hole mass, while nearest neighbour overdensity is important for quenching satellites (Bluck et al. 2020). As these are valid measures of the harassment of a star-forming region, as a time-dependent input they may correlate with dips or drops in SFH. The difficulty is in obtaining a measure of velocity dispersion which is not affected by baryons.

7.5 Modelling observables

Systematic differences between predicted SFH and ZH may be mitigated in future by means of a stochastic measure of their fluctuations. We can train the same neural networks to predict the Fourier power spectra of these histories, correcting for the smaller amplitude of our predictions at high frequencies. We have shown that residual luminosities and line emission features correlate with this loss of amplitude, and therefore this correction could improve the accuracy of our spectra and photometry.

7.6 Mock surveys

Our model paves the way for large-volume high fidelity mocks and the construction of mock spectroscopic surveys with a consistent galaxy–halo connection. Such mocks can cover a large sample of galaxies and at a variety of redshifts. With respect to models that directly infer observational properties from dark matter properties, our model has the advantage that it links those observables with physical properties self-consistently. This expands the potential utility of these mocks in supporting the analysis of real surveys.

8 SUMMARY AND CONCLUSIONS

We have developed a neural network model capable of synthesizing the star formation rate and stellar metallicity of central and satellite galaxies as a function of cosmic time, using historical properties of the dark matter halo and environment as predictors. We have tested for the predicting power of certain dark matter quantities in

determining aspects of the galaxy’s evolutionary history, and computed observational properties from our predictions. The following statements summarize important aspects of our predictions:

(i) For each of our networks, designed separately for central and for satellite galaxies, the full diversity of star formation histories are accurately predicted. In Sections 5.1.1 and 5.1.2 we see that geometries of the predicted SFHs match those from the original TNG merger trees, from those of continually growing star-forming galaxies to quenched satellites and high-mass galaxies. The integrated stellar mass from these historical growths are well correlated with their values in the TNG data, and their mass from their merger trees. We show that our predicted star formation histories recover the complete SHMR in Fig. 11.

(ii) This work includes the introduction of novel parameters, such as the skew of the radial mass distribution surrounding the target object, and recently devised quantities such as the scaled formation time from Shi et al. (2020). These quantities have proven valuable in the prediction of our star formation and metallicity histories, and their relations with halo and galaxy masses have been outlined by the neural network.

(iii) The derived SHMR for either data set is accurate and resilient to the removal of predicting features, i.e. it can be maintained by relatively few parameters. However, we show in Section 6.3 that omission of certain properties can cause subtle differences in the scatter of the SHMR, as indicated by multiple shuffled runs. Removing quantities related to subhalo growth rate and circular velocity have such an impact.

(iv) In spite of these fruitful outcomes, this network model seldom predicts star formation events which occur on short time scales, such as star formation bursts or rapid quenching. We see the absence of such features in Fig. 13 and the general lack of high frequency information. Consequently, numerical stellar mass estimates where these sharp changes are prominent are systematically smaller than their true value, and the absence of such features can impact the accuracy of observational properties, which is often seen in Section 5.2. This is a greater issue for central galaxies, where these features are more commonplace. We can, however, train an identical network to fit the power spectra of the SFHs, shown in Section 7.2, potentially offering stochastic corrections in future. Untested properties relating to the host of a satellite galaxy may prove valuable, however more general properties such as velocity dispersion are difficult to characterize independently of baryons.

(v) In Section 5.1.3, we see that mass-weighted metallicity scatter is slightly underestimated, which in addition to the shortcomings of the star formation histories, may be attributed to the lack of baryonic influence on the rotation curve. We show in Section 6.3 that overdensity and skew, i.e. environmental measures, are paramount to the scatter in metallicity history; removing either of them affects their predictions significantly. Determining the variability in metallicity has been an issue in pre-existing studies (Agarwal et al. 2018; Lovell et al. 2021), however these include little detail around environmental history.

(vi) Despite some notable differences in metallicity, the predicted galaxy formation histories are observationally congruous to their hydrodynamical counterparts. While not predicting them directly, in Section 5.2.3 we show that the network recovers major observational statistics such as the mass–colour relations and colour bimodality of TNG. $H\alpha$ line fluxes are recovered less accurately as these are sensitive to variations in recent star formation, yet they obey similar trends to the original data.

(vii) Mock observables are shown to be well matched to those emulated from TNG data, despite inaccuracies such as smoother formation histories and lower metallicity scatter. Spectral energy distributions are computed more accurately for satellite galaxies than they are for central galaxies, as seen in Section 5.2.1, likely owing to greater quenching. We show in Section 5.2.2 that larger residuals in high-frequency modes of the SFH result in greater errors in luminosities, therefore a stochastic correction based on Fourier amplitudes may rectify this discrepancy. Other options include the introduction of dark matter quantities which may result in better predictions of formation history, or predicting star formation events on finer time scales.

(viii) In Section 6.5, while we motivate infall parameters as potentially important measures of satellite evolution, and properties of the cosmic web for central galaxies, we find that these have little isolated effect on our predictions, most likely owing to the inference of their effects from historical data. In high fidelity mocks, a higher volume of samples will be used to quantify the strength and concurrence of these correlations.

Ours is a practical model which can be used in high fidelity N -body simulations, statistically reproducing key relationships of the GHC, and producing mock surveys of unprecedented size and complexity. A future publication will describe the predictions made using data from an N -body simulation, and the spectroscopic mocks which we aim to make publicly available.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referee for useful suggestions which led to significant improvements in the clarity of the manuscript. HGC wishes to thank the UKRI Science and Technology Facilities Council for funding this research, under grant ID ST/T506448/1. We wish to thank the IllustrisTNG project for access to their data and their JUPYTERLAB workspace.

DATA AVAILABILITY

The data supporting the findings of this research is currently available in an online Zenodo repository. The code, and documentation of the code and data, are available in an online GitHub repository.

REFERENCES

Abadi M. et al., 2016, preprint (arXiv:1603.04467)
 Agarwal S., Davé R., Bassett B. A., 2018, *MNRAS*, 478, 3410
 Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, *MNRAS*, 480, 3978
 Ayramlou M., Kauffmann G., Yates R. M., Nelson D., White S. D. M., 2021, *MNRAS*, 505, 492
 Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *MNRAS*, 373, 469
 Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
 Bluck A. et al., 2020, *MNRAS*, 499, 230
 Borzyszkowski M., Porciani C., Romano-Díaz E., Garaldi E., 2017, *MNRAS*, 469, 594
 Bose S., Eisenstein D. J., Hernquist L., Pillepich A., Nelson D., Marinacci F., Springel V., Vogelsberger M., 2019, *MNRAS*, 490, 5693
 Chabrier G., 2003, *PASP*, 115, 763
 Chalela M., Sillero E., Pereyra L., Alejandro García M., Cabral J. B., Lares M., Merchán M., 2021, *Astron. Comput.*, 34, 100443
 Chaves-Montero J., Hearin A., 2021, *MNRAS*, 506, 2373
 Clevert D.-A., Unterthiner T., Hochreiter S., 2015, preprint (arXiv:1511.07289)

Conroy C., Gunn J. E., 2010, *ApJ*, 712, 833
 Conroy C., Gunn J., White M., 2009, *ApJ*, 699, 486
 Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303
 Cui W., Davé R., Peacock J. A., Anglés-Alcázar D., Yang X., 2021, *Nat. Astron.*, 5, 1069
 Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
 Davies J. J., Crain R. A., Oppenheimer B. D., Schaye J., 2019, *MNRAS*, 491, 4462
 Dhoke P., Paranjape A., 2021, *MNRAS*, 508, 852
 Donnan C. T., Tojeiro R., Kraljic K., 2022, *Nat. Astron.*, 6, 599
 Engler C. et al., 2020, *MNRAS*, 500, 3957
 Fakhouri O., Ma C.-P., 2009, *MNRAS*, 394, 1825
 Fraser T. S., Tojeiro R., Chittenden H., 2022, preprint (arXiv:2112.12516)
 Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
 Gallazzi A., Bell E. F., 2009, *ApJs*, 185, 253
 Galárraga-Espinosa D., Garaldi E., Kauffmann G., 2022, preprint (arXiv:2209.05495)
 Géron A., 2020, Hands-On Machine Learning With Scikit-Learn, Keras, And TensorFlow: Concepts, Tools, And Techniques To Build Intelligent Systems, 2nd edn., O'Reilly Media Incorporated, Sebastopol, Canada
 Gu M., Conroy C., Behroozi P., 2016, *ApJ*, 833, 2
 Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, *MNRAS*, 398, 1742
 Hani M. H., Gosain H., Ellison S. L., Patton D. R., Torrey P., 2020, *MNRAS*, 493, 3716
 Haslbauer M., Dabringhausen J., Kroupa P., Javanmardi B., Banik I., 2019, *A&A*, 626, A47
 Hellwing W. A., Cautun M., van de Weygaert R., Jones B. T., 2021, *Phys. Rev. D*, 103, 063517
 Jiang F., van den Bosch F. C., 2014, *MNRAS*, 440, 193
 Jo Y., Kim J.-H., 2019, *MNRAS*, 489, 3565
 L'Huillier B., Park C., Kim J., 2015, *MNRAS*, 451, 527
 Lovell C. C., Wilkins S. M., Thomas P. A., Schaller M., Baugh C. M., Fabbian G., Bahé Y., 2021, *MNRAS*, 509, 5046
 Lu L., Shin Y., Su Y., Karniadakis G., 2020, *Commun. Comput. Phys.*, 28, 1671
 Lu S. et al., 2021, *MNRAS*, 509, 2707
 Marinacci F. et al., 2018, *MNRAS*, 480, 5113
 McGibbon R., Khochfar S., 2022, *MNRAS*, 513, 5423
 Montero-Dorta A. D., Chaves-Montero J., Artale M. C., Favole G., 2021, *MNRAS*, 508, 940
 Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
 Nelson D. et al., 2017, *MNRAS*, 475, 624
 Nelson D. et al., 2019, *Journal of Computational Astrophysics and Cosmology*, 6, 2
 Pasquali A., Gallazzi A., Fontanot F., Van Den Bosch F. C., De Lucia G., Mo H. J., Yang X., 2010, *MNRAS*, 407, 937
 Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
 Peng Y.-J., Maiolino R., 2013, *MNRAS*, 438, 262
 Pillepich A. et al., 2017a, *MNRAS*, 473, 4077
 Pillepich A. et al., 2017b, *MNRAS*, 475, 648
 Poole G. B., Mutch S. J., Croton D. J., Wyithe S., 2017, *MNRAS*, 472, 3659
 Ramakrishnan S., Paranjape A., Sheth R. K., 2021, *MNRAS*, 503, 2053
 Rodríguez-Gomez V. et al., 2015, *MNRAS*, 449, 49
 Schaye J. et al., 2014, *MNRAS*, 446, 521
 Shah A., Kadam E., Shah H., Shinde S., Shingade S., 2016, in Proc. Third Int. Symp. Computer Vision and the Internet. VisionNet'16. Association for Computing Machinery, New York, NY, USA, p. 59
 Shi J. et al., 2020, *ApJ*, 893, 139
 Simpson C. M., Grand R., Gómez F., Marinacci F., Pakmor R., Springel V., Campbell D., Frenk C., 2018, *MNRAS*, 478, 548
 Slone O., Jiang F., Lisanti M., Kaplinghat M., 2021, preprint (arXiv:2108.03243)
 Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
 Sousbie T., 2011, *MNRAS*, 414, 350
 Springel V. et al., 2017, *MNRAS*, 475, 676

- Tacchella S., Forbes J. C., Caplar N., 2020, *MNRAS*, 497, 698
Tinker J. L., Hahn C., Mao Y.-Y., Wetzel A. R., 2018, *MNRAS*, 478, 4487
Tojeiro R. et al., 2017, *MNRAS*, 470, 3720
Torrey P. et al., 2019, *MNRAS*, 484, 5587
Trevisan M., Mamon G. A., Thuan T. X., Ferrari F., Pilyugin L. S., Ranjan A., 2021, *MNRAS*, 502, 4815
Van Loon M. L., Mitchell P. D., Schaye J., 2021, *MNRAS*, 504, 4817
Villaescusa-Navarro F. et al., 2021, *ApJ*, 915, 71
Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, *Nat. Rev. Phys.*, 2, 42
Wadkar D., Villaescusa-Navarro F., Ho S., Perreault-Levasseur L., 2020, preprint ([arXiv:2012.00111](https://arxiv.org/abs/2012.00111))
Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, *MNRAS*, 432, 336
Xu X., Kumar S., Zehavi I., Contreras S., 2021, *MNRAS*, 507, 4879

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