



IllustrisTNG and S2COSMOS: possible conflicts in the evolution of neutral gas and dust

Jenifer S. Millard¹,¹★ Benedikt Diemer²,² Stephen A. Eales,¹ Haley L. Gomez,¹ Rosemary Beeston¹ and Matthew W. L. Smith¹

¹*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK*

²*Department of Astronomy, University of Maryland, College Park, MD 20742, USA*

Accepted 2020 October 8. Received 2020 September 28; in original form 2020 August 12

ABSTRACT

We investigate the evolution in galactic dust mass over cosmic time through (i) empirically derived dust masses using stacked submillimetre fluxes at 850 μm in the COSMOS field and (ii) dust masses derived using a robust post-processing method on the results from the cosmological hydrodynamical simulation IllustrisTNG. We effectively perform a ‘self-calibration’ of the dust mass absorption coefficient by forcing the model and observations to agree at low redshift and then compare the evolution shown by the observations with that predicted by the model. We create dust mass functions (DMFs) based on the IllustrisTNG simulations from $0 < z < 0.5$ and compare these with previously observed DMFs. We find a lack of evolution in the DMFs derived from the simulations, in conflict with the rapid evolution seen in empirically derived estimates of the low-redshift DMF. Furthermore, we observe a strong evolution in the observed mean ratio of dust mass to stellar mass of galaxies over the redshift range $0 < z < 5$, whereas the corresponding dust masses from IllustrisTNG show relatively little evolution, even after splitting the sample into satellites and centrals. The large discrepancy between the strong observed evolution and the weak evolution predicted by IllustrisTNG plus post-processing may be explained by either strong cosmic evolution in the properties of the dust grains or limitations in the model. In the latter case, the limitation may be connected to previous claims that the neutral gas content of galaxies does not evolve fast enough in IllustrisTNG.

Key words: galaxies: evolution – galaxies: ISM – galaxies: luminosity function, mass function – galaxies: statistics – submillimetre: ISM.

1 INTRODUCTION

Despite constituting around only 0.1 per cent of the baryonic content of a galaxy by mass (e.g. Vlahakis, Dunne & Eales 2005; Dunne et al. 2011; Clemens et al. 2013; Beeston et al. 2018; Driver et al. 2018), cosmic dust has a profound effect on our ability to study and understand the evolution of galaxies. Dust obscures and absorbs the optical and ultraviolet (UV) light from stars, and re-emits this energy at far-infrared (FIR) and submillimetre (sub-mm) wavelengths. Dust is thought to have absorbed around half of the starlight ever emitted in the Universe (e.g. Puget et al. 1996; Fixsen et al. 1998; Dole et al. 2006).

Cosmic dust is believed to originate from three main sources: the stellar winds of evolved low-mass stars, such as asymptotic giant branch (AGB) stars (e.g. Salpeter 1974; Valiante et al. 2009), the violent deaths of massive stars (supernova explosions, e.g. Dunne et al. 2003b; Morgan & Edmunds 2003; Matsuura et al. 2011; Gomez et al. 2012; De Looze et al. 2017, 2019; Chawner et al. 2019, 2020; Cigan et al. 2019), and grain-growth in the ISM (e.g. Dwek, Galliano & Jones 2007; Draine 2009; Dunne et al. 2011; Asano et al. 2013; Zhukovska et al. 2016; De Vis et al. 2017a,b). Since dust is a by-product of star formation, studies examining cosmic

dust allow us to probe the past star formation history of a galaxy. Furthermore, since cosmic dust is an important constituent of the ISM and pervasive throughout it, dust has been found to be useful as a tracer to estimate the molecular gas content of galaxies, the fuel for future star formation. The optically thin dust continuum emission detected at a single sub-mm wavelength can be used as a tracer of molecular gas (Eales et al. 2012; Scoville et al. 2014, 2016, 2017; Tacconi et al. 2018; Millard et al. 2020).

To enable studies of the gas and dust content of galaxies over much of cosmic history, astronomers have conducted sub-mm surveys of the sky. For example, the *Herschel Space Observatory* (hereafter *Herschel*, Pilbratt et al. 2010) was used in large-scale FIR surveys covering well-studied areas of sky (e.g. H-ATLAS, Eales et al. 2010), and the Submillimetre Common-User Bolometer Array 2 (SCUBA-2) on the James Clerk Maxwell Telescope (JCMT; Holland et al. 2013) has recently been used to study the COSMOS field (Scoville et al. 2007) at unprecedented depths at sub-mm wavelengths (Simpson et al. 2019). The Atacama Large Millimetre Array (ALMA) has also been used to observe the dust continuum of a significant number of high-mass galaxies (e.g. Scoville et al. 2017; Tacconi et al. 2018, and references therein).

Despite several studies illustrating the usefulness of sub-mm dust emission as a tracer for gas mass (e.g. Béthermin et al. 2015; Genzel et al. 2015; Scoville et al. 2016, 2017; Tacconi et al. 2018; Millard et al. 2020), such calculations involve significant assumptions. For

* E-mail: jenifer.millard@astro.cf.ac.uk

example, the studies implicitly assume that metallicity does not evolve with redshift and that the fraction of metals incorporated in dust grains is a constant, or equivalently that the dust-to-gas ratio is constant with time. A more direct measurement to make with sub-mm data is to measure dust masses, in which the only assumption is that the properties of dust are constant in space and time.

One of the fundamental measurements that can be made to describe the dust content of galaxies, particularly at different cosmic epochs, is the dust mass function (DMF) which is defined as the space density of galaxies as a function of dust mass (Beeston et al. 2018). The first measurements of the local DMF were made using sub-mm observations at 450 and 850 μm , using an indirect technique and were hampered by small number statistics (Dunne et al. 2000; Dunne & Eales 2001; Vlahakis et al. 2005). Early efforts to probe beyond the local Universe ($z = 1$, Eales et al. 2009; $z = 2.5$, Dunne, Eales & Edmunds 2003a) also suffered from poor statistics and a variety of other experimental limitations.

The advent of *Herschel*, with its better sensitivity and faster mapping speed, led to the first estimates of the DMF with good statistics. Using sub-mm data at 250 μm on 1867 sources out to $z = 0.5$, Dunne et al. (2011) found evidence for dramatic evolution in the dust content of galaxies, with five times more dust observed at $z = 0.5$ than $z = 0$, i.e. over the past 5 billion years of cosmic history. More recently, Beeston et al. (2018) presented an estimate of the local DMF with the best statistics to date, using 15 750 galaxies in the equatorial GAMA fields (Driver et al. 2011) out to a redshift of 0.1. Driver et al. (2018) later constructed an optically selected sample to probe the dust density out to redshifts of 5, finding little evolution from $0 < z < 0.5$, in contrast to Dunne et al. (2011). Driver et al. (2018) found an increasing dust mass content at higher redshifts, with dust mass density peaking at $z = 1$.

Although more detailed observations of the Universe over the past two decades have led to a deeper understanding of the content of galaxies, we still lack knowledge of the underlying physics governing galaxy evolution. This information is difficult to garner from observations alone. One solution to this problem is the development of suites of complex cosmological hydrodynamical simulations. Modern efforts include IllustrisTNG, which evolves a mock universe from shortly after the big bang through to the present day, including many of the physical processes that drive galaxy evolution, such as gas radiative mechanisms, star formation in the dense ISM, stellar population evolution, chemical enrichment, and feedback and outflows (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018a; Pillepich et al. 2018a; Springel et al. 2018).

However, cosmological hydrodynamical simulations such as IllustrisTNG are tuned to reproduce a selected set of observational constraints, such as the stellar mass content of galaxies and cosmic star formation rate density at $z = 0$ (Pillepich et al. 2018a). Therefore, it is vital to test these simulations by comparing their predictions to properties that they have not been tuned to reproduce.

Whilst current simulations, such as IllustrisTNG, do not model dust (although such models are in preparation, e.g. McKinnon et al. 2019), dust properties can be examined using post-processing analysis (e.g. Schulz et al. 2020; Vogelsberger et al. 2020). A recent example of this is presented in Baes et al. (2020), who examined the infrared luminosity functions and dust mass functions for galaxies from the EAGLE cosmological simulation (Crain et al. 2015; Schaye et al. 2015) for $z < 1$. Synthetic multiwavelength observations for EAGLE galaxies were generated using the radiative transfer code SKIRT (Baes et al. 2011; Camps & Baes 2015; Camps et al. 2016, 2018). Dust masses were then estimated by fitting a simple modified blackbody model to the synthetic luminosities generated

using SKIRT at wavelengths of 160, 250, 350, and 500 μm . They found that EAGLE predicted only mild evolution in the DMF out to $z = 1$, in contrast with the observations of the DMF of Dunne et al. (2011) but consistent with the milder evolution in the dust density found by Driver et al. (2018).

In our first study, (Millard et al. 2020, Paper I), we used the sub-mm emission from dust as a tracer of extra-galactic gas content over much of the history of the Universe. We estimated the average molecular gas mass fractions of galaxies in the COSMOS field using stacked 850 μm fluxes and the gas scaling relations derived in Scoville et al. (2016, 2017). Stacking combines the fluxes from similar galaxies whose individual data signals are otherwise buried beneath noise to allow estimations of average galactic properties. Information on individual galaxies is lost, but is gained on the population as a whole. We calculated inverse-variance weighted average sub-mm fluxes for $\sim 63\,000$ binned sources with stellar masses $9.5 \leq \log_{10}(M_*/M_\odot) \leq 12$, out to $z < 5$, extending to higher redshifts than previous studies, and including more ‘normal’ star-forming galaxies on the Main Sequence (MS; e.g. Daddi et al. 2007; Karim et al. 2011; Whitaker et al. 2012; Madau & Dickinson 2014; Lee et al. 2015).

In this study (Paper II), we instead use the average stacked sub-mm fluxes from Paper I to examine the evolution of dust mass in galaxies over cosmic time out to $z < 5$, which requires only the assumption that the properties of dust are a constant. We compare our observational results to the evolution of dust mass in galaxies predicted by IllustrisTNG using a post-processing analysis, to examine how well modern simulations trace dust. We also calculate the dust mass functions for IllustrisTNG galaxies, and compare these to the observations of Dunne et al. (2011) and Beeston et al. (2018).

We use the cosmological parameters from *Planck* (Planck Collaboration 2016) and make use of `astropy.cosmology` (Astropy Collaboration 2013; Price-Whelan et al. 2018) assuming Λ CDM and $H_0 = 67.7 \text{ km Mpc}^{-1} \text{ s}^{-1}$, $\Omega_{\text{m},0} = 0.307$, and $\Omega_{\text{b},0} = 0.0486$.

2 DATA

Here, we will briefly summarize the observational data used to derive average dust properties of physical galaxies over cosmic time, and also describe the simulations from IllustrisTNG used in this study. The observational data sets and catalogues used to compare with the simulations are described in more detail in Millard et al. (2020).

2.1 S2COSMOS map

The JCMT is equipped with a 10 000 pixel bolometer camera which operates in the sub-mm regime, specifically at wavelengths of 450 and 850 μm . This camera, SCUBA-2 (Holland et al. 2013), was recently used to produce the deepest and most sensitive 850 μm map of the COSMOS field (Scoville et al. 2007), as part of the SCUBA-2 COSMOS (S2COSMOS; Simpson et al. 2019) survey. The map combines archival data from the SCUBA-2 Cosmology Legacy Survey (S2CLS; Geach et al. 2017; Michałowski et al. 2017), and new data from S2COSMOS, resulting in a complete and homogeneous map of the COSMOS field at 850 μm .

Instrumental noise dominates over confusion noise in the S2COSMOS survey; the median instrumental noise in the central 1.6 deg^2 of the 850 μm map is $\sigma_{850\mu\text{m}} = 1.2 \text{ mJy beam}^{-1}$. Additional coverage of 1 deg^2 has a median instrumental noise level of $\sigma_{850\mu\text{m}} = 1.7 \text{ mJy beam}^{-1}$ (Simpson et al. 2019). In this work, we use the match-filtered map, which is more sensitive to point-source emission.

For further details of the maps produced in the S2COSMOS survey, we refer the reader to Simpson et al. (2019).

We derive the dust masses of galaxies in the S2COSMOS field by stacking on the 850 μm map (Millard et al. 2020). As such, we require a source catalogue to provide the locations of COSMOS galaxies, which we briefly discuss in the next section.

2.2 Astrophysical source catalogues

The COSMOS field is one of the most extensively studied areas of sky, with observations spanning almost the entire electromagnetic spectrum. As such, comprehensive photometric catalogues of the region have been created, and several spectral energy distribution (SED) fitting routines have been used with such data sets to constrain the physical properties of galaxies. In this work, we make use of the source catalogue presented in Driver et al. (2018), wherein physical parameters of galaxies were estimated using the SED fitting routine MAGPHYS (da Cunha, Charlot & Elbaz 2008). We also use the COSMOS2015 catalogue from Laigle et al. (2016).

2.2.1 MAGPHYS catalogue

MAGPHYS (da Cunha et al. 2008) produces probabilistic estimates of the physical properties of galaxies by using a χ^2 minimization technique to fit pre-determined libraries of physically motivated model SEDs to panchromatic photometry data. Examples of the possible physical parameters returned include stellar mass estimates and star formation histories. The model SEDs of MAGPHYS cover galactic emission from the UV through to the FIR. Stellar emission is modelled using synthetic spectra from Bruzual & Charlot (2003), assuming a Chabrier (2003) stellar initial mass function (IMF). This starlight can then be attenuated by dust present within the galaxy; either in spherically symmetric birth clouds, wherein lies a ‘warm’ dust component with temperature 30–60 K, or in the ambient interstellar medium (ISM), where the ‘cold’ dust component is assumed to have a temperature of 15–25 K (Charlot & Fall 2000). MAGPHYS employs an energy balance between the UV-to-near-Infrared (UV-NIR) and the mid-Infrared-to-far-Infrared (MIR-FIR) model components – the stellar emission attenuated by dust must be re-emitted in the FIR.

In this work, we use the MAGPHYS catalogue (MagPhysG10v05) of Driver et al. (2018), which is based on the 22-band panchromatic photometry catalogue of Andrews et al. (2017) and contains 173 399 sources. The photometric catalogue used with MAGPHYS (G10CosmosLAMBDARcatv06) is i -band < 25 mag limited (Driver et al. 2018). As stated in Andrews et al. (2017), the source catalogue is complete for objects with i -band < 24.5 mag and partially complete to $i < 25$ mag.

Redshifts for the MAGPHYS catalogue are sourced from Davies et al. (2015), who performed an independent extraction of spectroscopic redshifts from the zCOSMOS-Bright sample (Lilly et al. 2007) and combined these redshifts with archival spectroscopic redshifts from the PRIMUS, VVDS, and SDSS (Cool et al. 2013; Le Fèvre et al. 2013; Ahn et al. 2014) surveys. If spectroscopic redshifts were not available, sources were assigned photometric redshifts from COSMOS2015 (Laigle et al. 2016).

2.2.2 COSMOS2015 catalogue

The COSMOS2015 catalogue from Laigle et al. (2016) contains half a million NIR selected objects with photometry from the X-ray range, through to the radio. In the deepest K_s -band regions, for

stellar masses $> 10^{10} M_\odot$, the catalogue is 90 per cent complete out to $z = 4$. Where possible, sources in the catalogue have spectroscopic redshifts; otherwise photometric redshifts ($\sigma_{\Delta z/(1+z)} = 0.007$ for $z < 3$, and $\sigma_{\Delta z/(1+z)} = 0.021$ for $3 < z < 6$) are assumed.

2.2.3 Final astrophysical source catalogue

Andrews et al. (2017) advise that sources in their catalogue without matches in COSMOS2015 (Laigle et al. 2016) should be treated with caution. We cross-match the MagPhysG10v05 catalogue (Driver et al. 2018) to the publicly available COSMOS2015 catalogue (Laigle et al. 2016) on RA and Dec, to ensure catalogue completeness, and to extend the wavelengths for which photometric measurements are available. The cross-matched MAGPHYS-COSMOS2015 catalogue covers the central 1 deg^2 of the S2COSMOS map, where instrumental noise is lowest.

Emission from active galactic nuclei (AGNs) can contaminate measurements of galactic emission, particularly in the infrared, and subsequently systematically impact estimates of physical parameters like stellar masses and star formation rates (Ciesla et al. 2015). Since disentangling galactic emission from that originating from AGN is non-trivial, we choose to remove AGN from the cross-matched MAGPHYS-COSMOS2015 catalogue. We identify AGN using IR, radio and X-ray data, and also remove sources with MAGPHYS stellar masses greater than $10^{12} M_\odot$. A full description of this process is available in Millard et al. (2020).

In this work, we limit ourselves to considering galaxies in the MAGPHYS-COSMOS2015 catalogue with $\log_{10}(M_*/M_\odot) \geq 9.5$. In total, our final MAGPHYS-COSMOS2015 astrophysical source catalogue, without AGNs, consists of 63 658 galaxies, of which 13 955 have reliable spectroscopic redshifts.¹ We note that the MAGPHYS-COSMOS catalogue provides the RA, Dec, M_* , and z for our observed galaxies. We do not use the MAGPHYS dust masses in this work.

2.3 IllustrisTNG

IllustrisTNG is a suite of the state-of-the-art hydrodynamical cosmological simulations designed to illuminate the underlying physical processes that drive galaxy formation (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018a; Pillepich et al. 2018a; Springel et al. 2018). The simulations model both baryons and dark matter, and encode many physical processes, such as stellar evolution, galactic winds, and chemical enrichment schemes (Weinberger et al. 2017; Pillepich et al. 2018a,b), to try and create realistic representations of the universe from the big bang until the present day. IllustrisTNG utilizes the moving mesh code AREPO (Springel 2010) to model the evolution of galaxies over cosmic time. Galaxies and haloes are identified using the SUBFIND algorithm (Davis et al. 1985; Springel et al. 2001; Dolag et al. 2009). The updated galaxy formation model used in IllustrisTNG is based on the one originally presented in the Illustris simulation (Vogelsberger et al. 2013, 2014a,b; Genel et al. 2014; Torrey et al. 2014; Sijacki et al. 2015). The Illustris TNG simulations follow the cosmology of the Planck Collaboration (2016).

The simulations are run under three different resolutions – in this work, we focus on TNG100, with a box side length of 106.5 Mpc.

¹For consistency the redshifts used in this study are from Driver et al. (2018), since these are the redshifts used in producing the MAGPHYS estimates of the stellar masses.

Table 1. Number of TNG100 galaxies considered at each redshift, after removal of unphysical galaxies and resolution mass cuts to $200m_b$ (N_{filt}), and after additional mass cuts to allow for comparison with our observations obtained from stacking galaxies in the S2COSMOS field ($N_{\text{filt,obs}}$; see Sections 2.2 and 2.3).

Snapshot	z	$2r_{0.5}$	N_{filt}	$N_{\text{filt,obs}}$
17	5		3185	300
21	4		7394	996
25	3		14 338	2631
33	2		22 902	5531
40	1.5		26 270	7227
50	1		27 456	8680
59	0.7		27 442	9350
72	0.4		26 734	9563
78	0.3		26 393	9513
84	0.2		26 168	9462
91	0.1		25 660	9298
99	0		25 254	9252

We consider snapshots from the present day, out to $z = 5$, such that we explore a similar redshift range compared to our observational data. Of particular interest to this study are the predictions for the evolution of the neutral gas and metallicity (e.g. Diemer et al. 2018, 2019; Nelson et al. 2018b; Pillepich et al. 2018a).

2.3.1 Simulated galaxies from TNG100

In TNG100, there are a few objects that may be artefacts (Nelson et al. 2019). From our sample, we therefore remove galaxies with zero neutral gas and no metals (these are mostly satellite galaxies that have presumably undergone excessive stripping through galaxy–galaxy interactions; see Diemer et al. 2019).

When comparing to data from the literature, we impose a lower stellar mass limit of 200 times the baryonic mass resolution of the simulation ($200m_b$; $m_b = 1.4 \times 10^6 M_\odot$) to avoid galaxies that may be poorly resolved in the simulations and therefore not reliably modelled (Shen et al. 2020). In other words, we perform a stellar mass selection on IllustrisTNG galaxies, without considering other galactic baryonic mass components e.g. gas mass. When comparing to only our stacking data, we consider only TNG100 galaxies with stellar masses within the bounds imposed by the astrophysical data i.e. $9.5 \leq \log_{10}(M_*/M_\odot) \leq 12$.

In this study, we use parameter values for a galaxy that are calculated for particles and cells within two stellar half-mass radii ($2r_{0.5}$). However, to ensure the robustness of our conclusions and to check that our choice of TNG100 data set does not influence our conclusions, we also perform the same calculations (re-running all results) but for parameter values calculated using all particles gravitationally bound to the galaxy. We do not discuss these results in the paper, other than to say we find that using parameter values calculated for all particles and cells gravitationally bound to a source does not change our conclusions. Table 1 illustrates the number of TNG100 galaxies considered at each redshift, for the $2r_{0.5}$ data set.

2.4 Stellar mass distributions

Before we compare dust masses for the observational and simulated data, we perform a quick check and examine the normalized stellar mass distributions of galaxies in the two data sets. We bin the galaxies

in the observational MAGPHYS-COSMOS catalogue into the same redshift bins as the IllustrisTNG snapshots (Fig. A1 and Table 1). We see that broadly, the stellar mass distributions are similar. We can be confident that the observed and simulated galaxies represent similar galactic populations.

3 DUST MASSES

3.1 Empirical dust masses from stacking on S2COSMOS 850 μm map

Typically, the 850 μm emission from individual galaxies is below the noise level of the S2COSMOS map (Simpson et al. 2019). Therefore, in Paper I, we studied the 850 μm emission of galaxies in the COSMOS field in a statistical manner – we binned galaxies by redshift and stellar mass to create subpopulations of these sources, and used well-established inverse-variance weighted (IVW) stacking methodologies to estimate mean sub-mm fluxes for sources in each bin. Although this does mean that we lose information on individual galaxies, we gain information on each subpopulation as a whole. We estimated 850 μm stacked flux errors using Monte Carlo (MC) simulations on the location of galaxies within the S2COSMOS map. See Millard et al. (2020) for full details.

For sub-mm emission measured on the Rayleigh-Jeans tail of dust emission, dust masses, M_d , can be estimated using

$$M_d = \frac{\langle S_{\nu_0} \rangle D_L^2}{\kappa_{\lambda_e} B_{\nu_e}(T_d)(1+z)} \quad (1)$$

where $\langle S_{\nu_0} \rangle$ is the average stacked flux for galaxies in a given ($M_* - z$) bin, measured at the wavelength of observation (in this case, 850 μm), D_L is the luminosity distance of the source,² κ_{λ_e} is the assumed dust mass absorption coefficient scaled to the rest-frame emission wavelength, $B_{\nu_e}(T_d)$ is the Planck function calculated at the rest-frame emission frequency for an assumed dust temperature, T_d , and z is the middle of the redshift bin under consideration (e.g. Dunne et al. 2003a).

The value κ_{λ_e} is notoriously uncertain. Indeed, a recent study by Clark et al. (2016) showed that estimates of κ_{500} in the literature span over 3.5 orders of magnitude. We have used the freedom produced by this uncertainty to choose a value of κ_{500} that gives as good agreement as possible between the observed low-redshift DMFs (from Dunne et al. 2011 and Beeston et al. 2018) and the one predicted by IllustrisTNG (Section 4.1). Rather than a formal fit, given the uncertainties and small differences in the shape of the predicted and observed $z = 0$ DMFs, we determined κ_{500} by performing a rough fit by eye to the high-mass end of the $z = 0$ DMFs. In particular, we ensured a good fit with the $z < 0.1$ DMF from Dunne et al. (2011). By performing this ‘self-calibration’ of κ_{500} at low redshift, we are still able to test whether IllustrisTNG predicts the correct shape of the low-redshift DMF and whether it predicts the cosmic evolution seen in the observations. We assume $\kappa_{500} = 0.14 \text{ m}^2 \text{ kg}^{-1}$, which is close to that found by James et al. (2002) and roughly around the middle of the range of values given by Clark et al. (2016). Note that we scale any dust masses taken from the literature to our chosen κ_{500} values. We scale the dust mass

²In this work, we make use of the Python package `astropy.cosmology`, assuming Planck15 cosmology, to calculate the luminosity distance using the centre of the redshift bin for the subpopulation in question.

emissivity coefficient to the rest frame emission wavelength using:

$$\kappa_{\lambda_e} = \kappa_{500} \left(\frac{\lambda_{500}}{\lambda_e} \right)^\beta \quad (2)$$

where β is the dust emissivity spectral index. We assume $\beta = 1.8$ (Planck Collaboration 2016), consistent with our previous stacking work and other studies upon which this work was based (see Millard et al. 2020 for details; see also Scoville et al. 2016; Scoville et al. 2017). Once again, for our stacked data from the S2COSMOS field, the rest frame emission wavelength is estimated by using the middle of the redshift bin currently being investigated. By scaling the dust mass emissivity coefficient to the wavelength of emission in the rest frame, we implicitly assume that the properties of dust grains themselves are constant throughout the Universe. We assume a mass-weighted dust temperature of 25 K, as in Millard et al. (2020) (see also Scoville et al. 2016, 2017). We note that our method for calculating dust masses is simple, only depends on a few parameters, and does not have the black-box complexity of spectral energy distribution-fitting methods such as MAGPHYS. We discuss the impact of our assumptions on the values for β , κ_{500} and mass-weighted dust temperature in Section 5.1.

3.2 Dust masses from IllustrisTNG

Dust masses are not one of the physical parameters explicitly calculated for galaxies in IllustrisTNG. Therefore, we have to calculate them in post-processing. We calculate dust masses, $M_{d,TNG}$, on a cell-by-cell basis, summing over all cells within two stellar half-mass radii of a given galaxy:

$$M_{d,TNG} = \epsilon_d \left(\sum_{i=1}^{N_{2r}} \frac{f_{H,neutral,i}}{f_{H,i}} M_{gas,i} Z_i \right) \quad (3)$$

where $M_{gas,i}$ is the mass of gas in a given cell, Z_i is the metallicity of the gas in the cell, $f_{H,neutral,i}$ is the fraction of gas in the cell that is neutral hydrogen (atomic or molecular hydrogen; HI or H₂), $f_{H,i}$ is the fraction of all the gas in a cell that is hydrogen, and N_{2r} is the number of cells within two stellar half-mass radii for a given source. ϵ_d is the fraction of metals in a galaxy's ISM that are locked up in dust, a free parameter. We justify our choice of ϵ_d at the end of this Section.

In IllustrisTNG, the neutral gas fraction is calculated using two different prescriptions, depending on whether cells are classified as star-forming or not. For cells below the star formation threshold density, the neutral gas fraction is calculated self-consistently within the simulations (Vogelsberger et al. 2013). The combination of the cooling rate (from primordial hydrogen and helium cooling, metal-line cooling and inverse Compton cooling off CMB photons), the photoionization rate (from the ultraviolet background; UVB) and gas self-shielding provide an overall UVB photoionization rate, which is used with CLOUDY look-up tables (Ferland et al. 1998) to calculate the neutral gas fraction. However, for star-forming cells where the gas density is higher, these approximations break down; the neutral gas fraction is not calculated self-consistently within the simulations (Springel & Hernquist 2003). For these cells, the model splits gas into a hot and cold phase, where the cold phase is assumed to be entirely neutral gas. The fraction of neutral hydrogen in star-forming cells is then estimated based on the density fraction of cold gas clouds, with a typical fraction between 0.9 and 1 (Springel & Hernquist 2003).

3.2.1 A constant dust-to-metal ratio?

The most direct estimates for the dust-to-metal ratio, ϵ_d , come from using UV spectroscopic observations of stars to measure the metal abundances within stars and in the ISM, the difference between the two being attributed to the metals depleted from the interstellar gas and locked up in dust grains (James et al. 2002). Clark et al. (2016) list 12 different estimates of ϵ_d , which have a mean value of 0.5 and a standard deviation of 0.1. Most of these measurements, however, are ultimately based on depletion measurements in the Milky Way or rely on theoretical models of dust grains. The only other galaxies for which depletion measurements have been made using UV spectroscopy of stars are the Magellanic Clouds (Roman-Duval et al. 2019), which have much lower stellar masses than the galaxies in our COSMOS sample. The depletion measurements imply that ϵ_d in these galaxies is lower than in the Milky Way, by a factor of 1.5 for the LMC and a factor of 4 for the SMC. In an alternative approach to this question, Rémy-Ruyer et al. (2014) and De Vis et al. (2017b, 2019), in studies of large samples of nearby galaxies, find a linear relationship between dust-to-gas ratio and metallicity for galaxies with $12 + \log(O/H) > 8.0$. The former work implies a constant value of ϵ_d of ~ 0.3 and the latter implies ~ 0.2 for evolved galaxies,³ suggesting that only the lowest mass, lowest metallicity galaxies would likely deviate from this constant value locally. Both studies observed a broken power law where the dust-to-metals ratio starts to vary with metallicity below a ‘transition’ value of $12 + \log(O/H) \sim 8$. An illustrative comparison of the dust-to-metals ratio observed in the local Universe is shown in Fig. 1, including predictions from the chemical evolution models of De Vis et al. (2017b).

It is not immediately clear what values of ϵ_d are appropriate at high redshifts, where high stellar mass galaxies are at an earlier stage of their evolution and therefore may have lower metallicities. However, De Vis et al. (2019) found galaxies in the local DustPedia survey with high dust-to-metal ratios (0.2–0.5) even at early evolutionary stages (defined as galaxies with high gas fractions). For the highest gas fractions in their sample ($f_{gas} > 0.8$), they began to see dust-to-metal ratios as low as 0.04. Observations at high redshifts include gamma-ray bursts and damped Lyman alpha systems (De Cia et al. 2013; Zafar & Watson 2013) where the observed values of ϵ_d for high-metallicity galaxies is similar to that for the Milky Way (see also Yajima et al. 2014). De Cia et al. (2013) suggest that ϵ_d could be lower for very low-metallicity galaxies. Similarly, no obvious trend with redshift was seen by Wiseman et al. (2017), but extreme changes in ϵ_d are postulated for galaxies at $z > 6$ (see Vogelsberger et al. 2020 for more discussion). Li, Narayanan & Davé (2019) however see little evolution in the dust-to-metal ratio from $z = 0$ to $z = 6$, instead showing it is most sensitive to the gas phase metallicity Z (though decreasing values are seen at lower stellar masses and high gas fractions; De Vis et al. 2017b).

For the redshift range used in this work, both the hydro-simulations from McKinnon, Torrey & Vogelsberger (2016) and semi-analytical models of Popping, Somerville & Galametz (2017) imply that ϵ_d is different at $z = 0$ and $z = 2$. More recently, using the IllustrisTNG simulations we use in this work, Vogelsberger et al. (2020) required a

³We note that these quoted values for ϵ_d were calculated assuming a different dust mass emissivity coefficient compared to this study. Scaling to the same κ_λ as used in this work results in a dust-to-metal ratio ~ 20 and ~ 30 per cent higher than quoted in the original studies, thus reducing any perceived difference between the absolute values of quoted ϵ_d here and in previous works.

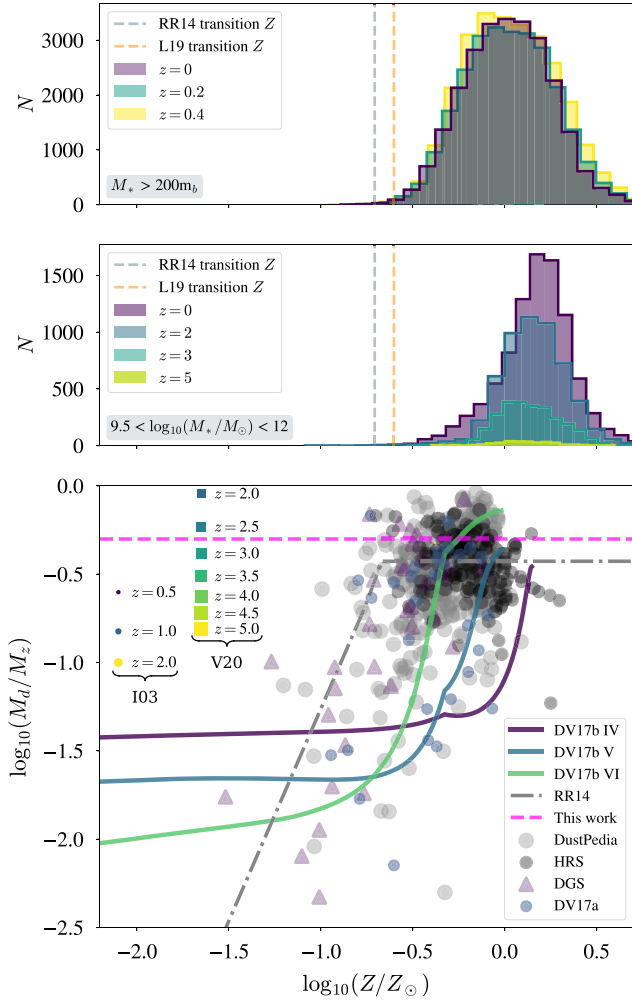


Figure 1. The dust-to-metal ratio for the local Universe and our samples. In the top two panels, we show the distribution of the metallicities of IllustrisTNG galaxies for the two subsamples we use in this work. The vertical lines indicate the transition metallicity from Rémy-Ruyer et al. (2014) (grey) and from Lagos et al. (2019) (orange). In the bottom panel, we show local galaxy samples DustPedia (De Vis et al. 2019), the gas-selected HSS sample (De Vis et al. 2017a,b), the revised parameters for the Dwarf Galaxy Sample (DGS; Madden et al. 2013) from De Vis et al. (2017b) and the *Herschel* Reference Survey (HRS; Boselli et al. 2010). Following De Vis et al. (2019), the metal mass M_Z is defined as $M_Z = (Z \times M_{\text{gas}}) + M_{\text{dust}}$ where the fraction of metals in the ISM by mass Z is defined as $Z = 27.67 \times 10^{12 + \log(\text{O}/\text{H}) - 12}$ where the solar value of $12 + \log(\text{O}/\text{H})$ is 8.69. Following IllustrisTNG, we use a solar value of $Z_{\odot} = 0.0127$ (Wiersma et al. 2009). A broken power-law relationship with Z from Rémy-Ruyer et al. (2014) (their table 1, using the Milky-Way CO – H₂ conversion factor) is shown as the dot-dashed line. Predictions of the evolution of dust-to-metal ratios are illustrated by the models from De Vis et al. (2017b) (solid lines). Coloured circles show the range of the dust-to-metals predicted as a function of z from Inoue (2003, 103), with markers increasing in size and colour-coded by redshift. Square coloured boxes show the values proposed to match observed UV luminosity functions from $2 < z < 9$ (Vogelsberger et al. 2020, V20). These are included to show potential variations in ϵ_d with redshift.

dust-to-metal ratio (their f_Z parameter) to vary with redshift over the range $2 < z < 8$ in order to match observed UV luminosity functions. Their relationship ($\epsilon_d = 0.9 \times (z/2)^{-1.92}$) predicts ϵ_d values of 0.9–0.1 for the redshift range which overlaps with this work ($2 < z < 5$),

see Fig. 1. We note that this $z = 2$ value is higher than that predicted by Inoue (2003; Fig. 1) and McKinnon et al. (2016).

Fig. 1 compares the distribution of metallicities in the two subsamples of the IllustrisTNG simulations used in this work: the $200m_b$ cut used for comparing with local galaxies (redshifts $0 < z < 0.4$), and the COSMOS stellar-mass cut ($9.5 \leq \log_{10}(M_*/M_{\odot}) \leq 12$) we use for comparison with our stacking data out to redshift $z = 5$. To investigate the impact of metallicity on the dust estimates used in this work, we compare with the broken power-law of Rémy-Ruyer et al. (2014) which suggest departures from a constant ϵ_d only occurs at $Z/Z_{\odot} < 0.2$ ($Z/Z_{\odot} < 0.25$ for Lagos et al. 2019). For the $200m_b$ low z sample, some galaxies have metallicities equal to, or below, the Rémy-Ruyer et al. (2014) and Lagos et al. (2019) transition metallicity, but these are small numbers – $N \sim 12$ ($N \sim 60$) for Rémy-Ruyer et al. (2014) (Lagos et al. 2019). The majority have metallicities where a constant dust-to-metal ratio is appropriate. For the COSMOS sample, very few of our sources are at low enough metallicities, even at high redshifts. In summary, we show there is observational evidence that both our local samples of IllustrisTNG galaxies and high-metallicity systems have a fairly constant value of ϵ_d over the redshift range appropriate in this work, with a value similar to that of the Milky Way. Fig. 1 illustrates that a constant value of $\epsilon_d = 0.5$, the average value from the meta-analysis of Clark et al. (2016), is appropriate for the TNG100 sources used in this study. We will return to this issue in Section 5.

Finally, we note that a similar method was used to derive dust masses for IllustrisTNG galaxies by Hayward et al. (2020). They define ISM gas using a temperature-density cut from Torrey et al. (2012) (summing over all particles within 25 kpc of the subhalo centre) and use this to compute an ISM metal mass, which they convert to a dust mass using a constant dust-to-metal ratio of 0.4.

4 RESULTS

4.1 Local DMF: IllustrisTNG

Before examining the evolution of dust mass in IllustrisTNG galaxies over much of cosmic time, we first investigate the dust content of IllustrisTNG galaxies in the local universe by calculating the IllustrisTNG DMF for $z < 0.5$. We compare these low-redshift DMFs to ones from the literature – specifically, from Dunne et al. (2011) and Beeston et al. (2018). Dunne et al. (2011) used a sample of 1867 sources selected at $250 \mu\text{m}$ from the Science Demonstration Phase of H-ATLAS (Eales et al. 2010) with reliable Sloan Digital Sky Survey [SDSS (DR7); Abazajian et al. 2009] counterparts. Dust masses are estimated by using both a single and multiple temperature modified blackbody models for the SEDs, assuming a κ_d value consistent with that used in MAGPHYS. Their sample covers a redshift range of $0 < z < 0.5$. Beeston et al. (2018) uses MAGPHYS to estimate dust and stellar masses for optically selected galaxies from the three equatorial fields of the Galaxy and Mass Assembly Catalogue (GAMA; Driver et al. 2011), which are the fields that also have FIR/sub-mm data from *Herschel*-ATLAS (Eales et al. 2010). In total, they consider 15750 sources covering around 145 deg^2 of the sky, at redshifts $z < 0.1$.

We calculate the IllustrisTNG DMFs over a similar dust mass range to those in the literature,⁴ and bin our galaxies in stellar mass bins of

⁴Specifically, our lower dust mass limit in Fig. 2 is dictated by the average dust mass of galaxies in the lowest stellar mass bin of Fig. 3 to avoid using dust mass bins which are affected by the sharp stellar mass cut-off of $200m_b$ introduced in Section 2.3.1.

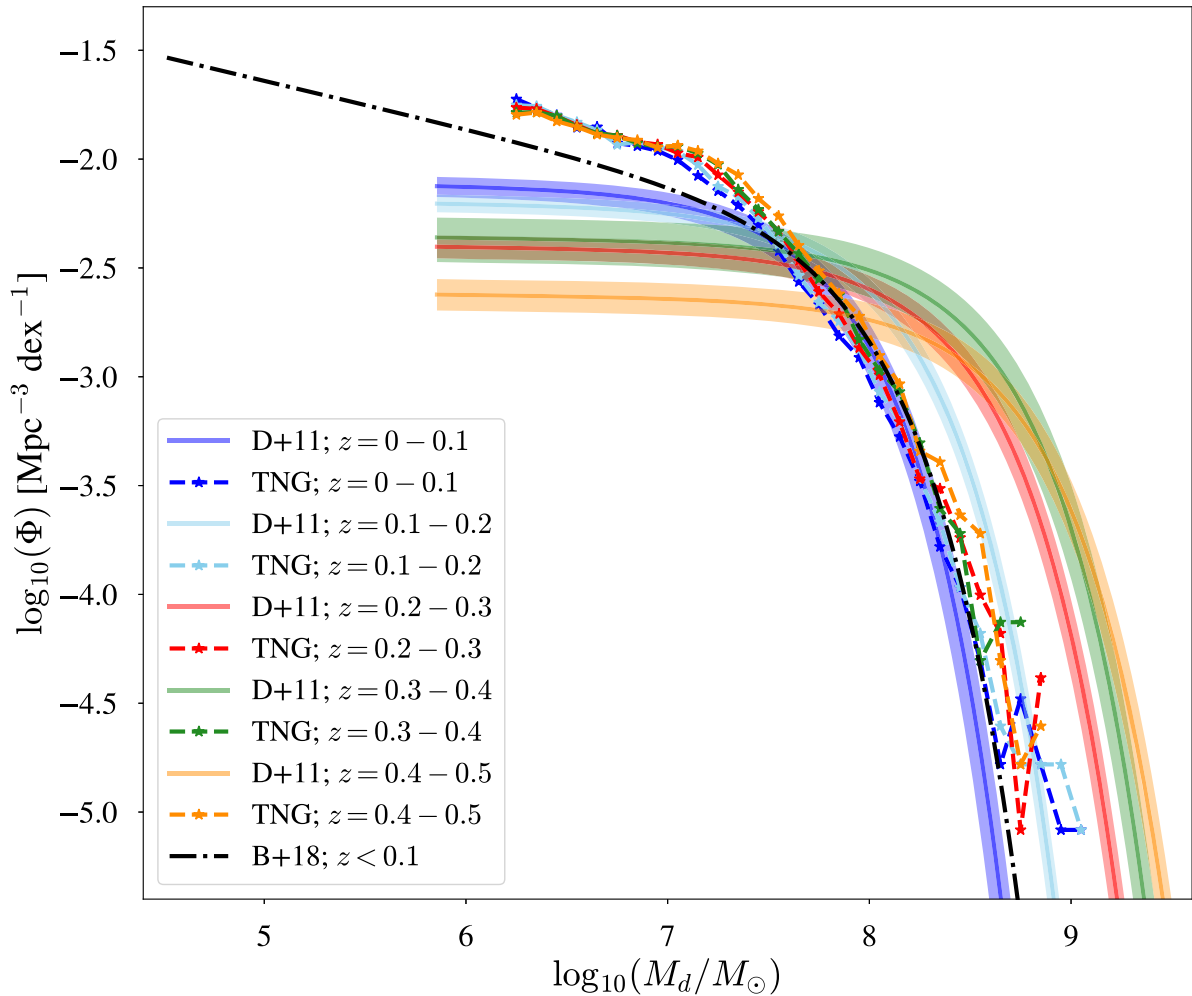


Figure 2. Low-redshift DMFs for IllustrisTNG galaxies (stars and dashed lines) and local DMFs from Beeston et al. (2018) (black dash-dot line) and Dunne et al. (2011) (solid colour lines). The DMF from Beeston et al. (2018) is based on an optically selected sample of 15 750 galaxies within the redshift range $0.002 \leq z \leq 0.1$. The DMFs from Dunne et al. (2011) are based on a $250 \mu\text{m}$ selected sample of 1867 galaxies with redshifts $z < 0.5$. The width of the dot-dashed line represents the error for the Beeston et al. (2018) DMF; the coloured shaded regions show the errors for the Dunne et al. (2011) DMFs. The colours of the IllustrisTNG DMFs and DMFs from Dunne et al. (2011) are chosen to indicate the same redshift bins. Local observed DMFs are scaled to IllustrisTNG cosmology with the dust masses scaled to κ_{500} .

0.1 dex. We scale the DMFs from Dunne et al. (2011) and Beeston et al. (2018) to the same cosmology as assumed in the IllustrisTNG simulations, and to κ_{500} (Fig. 2). Scaling to the κ_{500} used in this study increases the dust masses quoted in the literature studies by a factor of 1.43.

In Fig. 2, we see that at high dust masses, beyond the knee of the DMFs, the $z = 0$ TNG DMF agrees well with the DMF from Beeston et al. (2018), and with the $z < 0.1$ DMF from Dunne et al. (2011). We note that some of this agreement is by construction since we calibrated κ_{500} by insisting on good agreement between the observed and theoretical DMF at high dust masses, but the agreement between the shapes of the DMFs (beyond the knee) is still significant. At lower dust masses, we find a higher number density of galaxies for TNG, compared to Beeston et al. (2018). We attribute the excess number of sources at low dust masses to galaxies classified as ‘satellites’ in the simulation, which have likely undergone excessive stripping (Diemer et al. 2019; Stevens et al. 2019), therefore reducing their dust masses as calculated using our post-processing technique and causing an excess of sources with low dust masses as compared to observations. We return to this issue and explore the differences

between ‘satellite’ and ‘central’ TNG100 galaxies in Section 5.2. However, the slope of the DMF at low dust masses seems similar to that found by Beeston et al. (2018), even if there is an offset in the absolute number density.

The most striking difference between the observational and theoretical DMFs (Fig. 2) is that we do not see the evolution in the DMF over the redshift range $0 < z < 0.5$ that is seen in the observational data. The TNG DMFs show little evolution up to $z = 0.5$, which is in stark contrast to the dust mass evolution shown in the observational data from Dunne et al. (2011). However, it is worth noting that Dunne et al. (2011) caution that their DMFs at $z > 0.3$ are estimated from galaxies that mostly only have photometric redshifts. Even so, the strongest evolution in the DMFs from Dunne et al. (2011) is at redshifts below this threshold, and this evolution is clearly not apparent in the TNG100 DMFs.

We now examine the $z = 0$ ($M_* - M_d$) distribution of TNG100 galaxies and compare this distribution to local observational studies, as a sanity check that the dust mass estimates for TNG100 galaxies are consistent with those observed in the local galaxy population (Section 4.2).

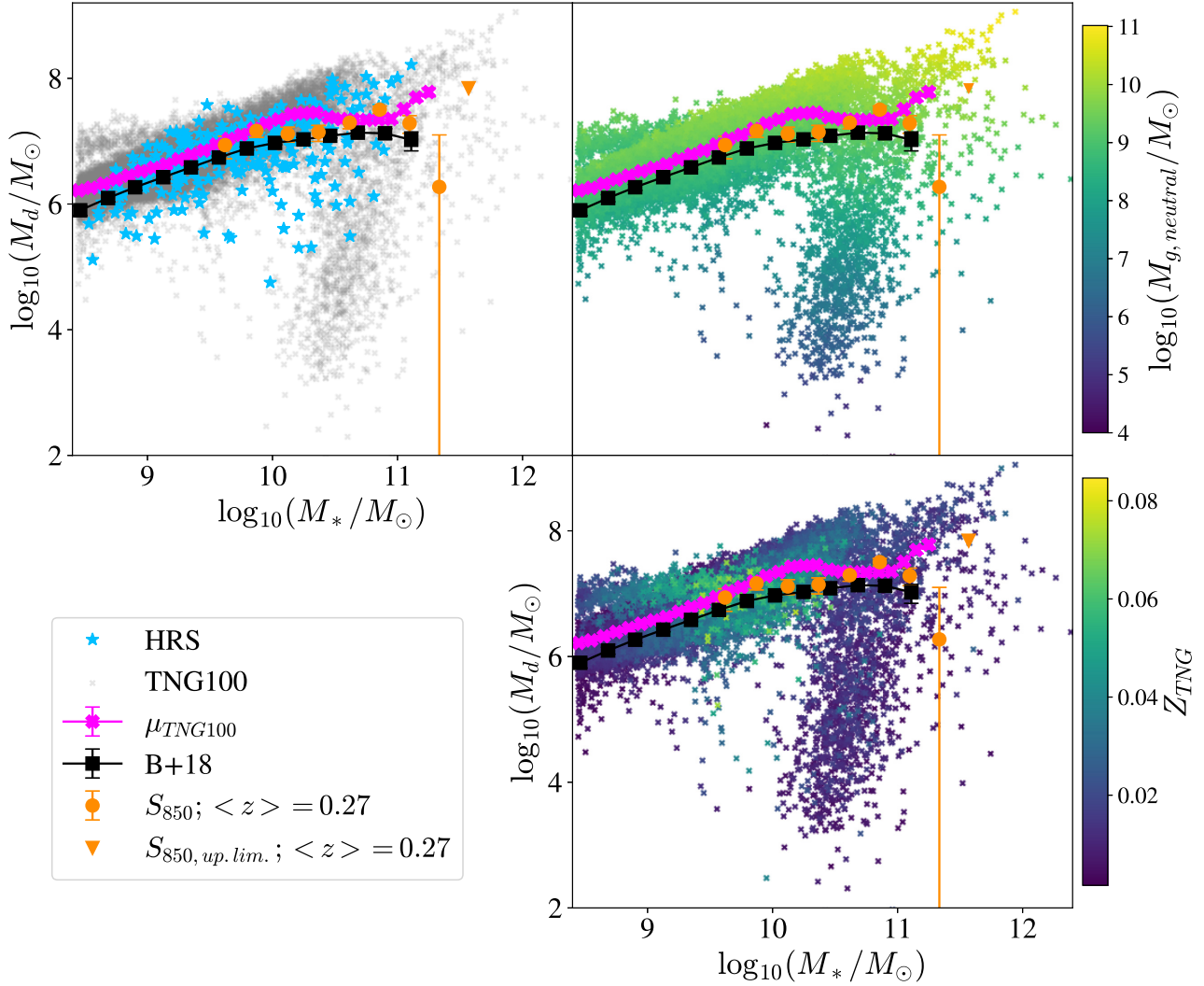


Figure 3. Dust mass versus stellar mass for the TNG100 galaxies at $z = 0$ and for various low-redshift galaxy samples. The left-hand panel shows the TNG100 galaxies as grey points; the right-hand top panel shows them as crosses with a colour given by each galaxy’s neutral gas mass (see colour bar); and the right-hand bottom panel shows them as crosses with a colour given by each galaxy’s metallicity (see colour bar). The mean and standard error on the mean for the TNG100 galaxies in a given stellar mass bin are shown by fuchsia crosses. The galaxies in the Herschel Reference Survey are shown as blue stars (De Vis et al. 2017a), the mean dust masses from the sample of Beeston et al. (2018) as black squares, and the mean dust masses for the galaxies in the low-redshift bins from our S2COSMOS study as orange circles, with the orange triangle showing an upper limit.

4.2 Dust mass estimates at $z = 0$

We compare the dust masses calculated using the $z = 0$ TNG100 data to dust masses calculated using our stacked S2COSMOS fluxes, and to dust masses from other studies of the local Universe from De Vis et al. (2017a) and Beeston et al. (2018) (Fig. 3). De Vis et al. (2017a) use MAGPHYS to estimate dust and stellar masses for 323 galaxies that comprise the Herschel Reference Survey (HRS; Boselli et al. 2010). The HRS is a stellar-mass selected volume-limited sample of galaxies in the nearby Universe. Sources have distances between 15 and 25 Mpc, and are K -band selected, to minimize dust selection effects caused by extinction. The sample contains both late- and early-type galaxies. A brief description of the data used in Beeston et al. (2018) can be found in Section 4.1. The dust masses from De Vis et al. (2017a) and Beeston et al. (2018) are scaled to κ_{500} . In Fig. 3, we bin the TNG100 sources by stellar mass and calculate the mean and standard error on the mean for galaxies in a given bin, as long as there are at least 50 sources in the bin.

We see that, broadly, the dust mass estimates from local studies agree with the $z = 0$ galaxies from TNG100 (Fig. 3). However, it is worth noting that dust masses for the K -band-selected HRS, and the average dust mass in each stellar mass bin for the TNG100 simulations (which agree well with each other) are a little higher compared to the larger, but optically selected, H-ATLAS sample from Beeston et al. (2018). The offset between our mean TNG100 values and the mean values for Beeston et al. (2018) is ~ 0.2 dex. The $850\mu\text{m}$ stacked sample from S2COSMOS lies somewhere between the HRS and Beeston et al. (2018). It may be possible that the small absolute offsets in average dust masses seen here are enhanced by the different selection effects and sampling issues of the observational studies. In particular, the optically selected sample of Beeston et al. (2018) have average dust masses sitting slightly below the FIR-selected samples of H-ATLAS sources (see their fig. 10). Conversely, the TNG100 galaxy sample is not subject to such observational selection effects. The deviation could also be attributed to the methods used to calculate stellar and dust masses in the

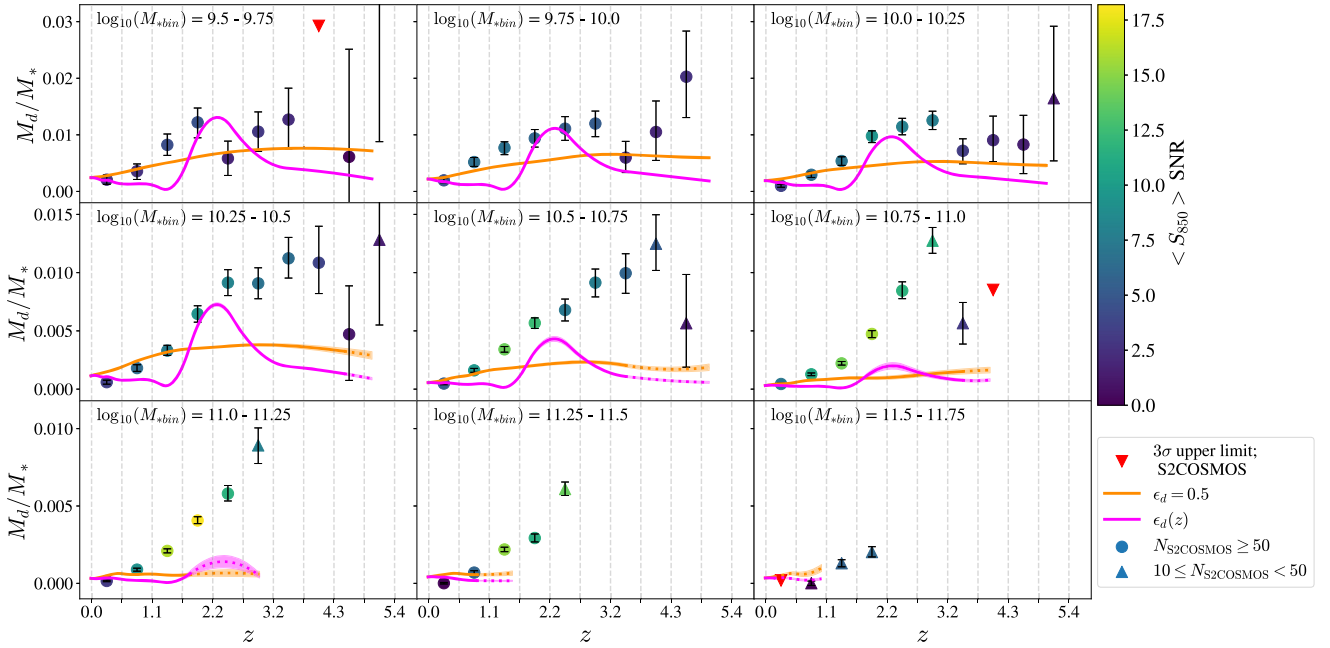


Figure 4. Dust-to-stellar-mass ratios obtained from the stacked S2COSMOS fluxes (symbols, Millard et al. 2020) and TNG100 galaxies (orange and magenta lines). Red triangles: observed dust-to-stellar-mass ratios estimated using flux 3σ upper limits. Circles: observed dust-to-stellar-mass ratios, where colour of point represents SNR of flux used to estimate dust masses, and there are at least 50 MAGPHYS-COSMOS sources in $(M_* - z)$ bin. Triangles: number of MAGPHYS-COSMOS sources in $(M_* - z)$ bin, N_{S2COSMOS} , between 10 and 50. Orange line: dust-to-stellar-mass ratio for TNG100 galaxies assuming a constant $\epsilon_d = 0.5$ (quadratic interpolation, see Fig. B1). The magenta line shows instead the results when a redshift-dependent relation for $\epsilon_d(z)$ is assumed (Inoue 2003 for $z < 2$ and Vogelsberger et al. 2020 for $z \geq 2$). Note the sharp transition at $z \sim 1.7$. Solid line: at least 50 TNG100 sources in $(M_* - z)$ bin. Dashed line: number of TNG100 sources in $(M_* - z)$ bin, N_{TNG} , between 10 and 50.

observational studies and the simulation data, which are not identical. However, qualitatively, the results for the data sets considered in Fig. 3 are reasonably consistent, despite the different techniques and samples used to estimate physical parameters. We again remind the reader that the agreement shown in Fig. 3 is, in part, by construction due to the choice of κ_{500} (see Section 4.1).

At $\log_{10}(M_*/M_\odot) \sim 10.5$ in Fig. 3, there is a distinct population of TNG100 galaxies with a lack of dust (grey crosses). Since these are massive galaxies, low resolution is not likely to be the cause of such low dust masses. As can be seen in the top-right panel of Fig. 3, these galaxies are devoid of neutral gas. These TNG100 galaxies are ‘quenched’ galaxies that have had their gas supply disrupted due to AGN feedback mechanisms. In IllustrisTNG, this stellar mass regime is the threshold at which kinetic mode feedback from AGN (in the form of black hole driven winds) becomes a dominant feedback effect (Weinberger et al. 2017, 2018; Terrazas et al. 2020; Zinger et al. 2020).

In Fig. 3, at lower stellar masses, there is a spur of galaxies with high dust masses, which is not seen in the observed samples. These galaxies also have high metallicities (Fig. 3, bottom-right panel).

Despite some small differences, the galaxies from TNG100 broadly follow the same $(M_* - M_d)$ distribution that we see in observations at $z = 0$, showing that with the exception of the low-mass end of the IllustrisTNG DMF, our post-processing method for estimating the dust mass does a good job of predicting the dust properties of the local galaxy population. Having performed this low-redshift check, we now investigate whether the discrepancy between the observed and predicted evolution over the redshift range $0 < z < 0.5$ continues to higher redshifts.

4.3 High-redshift dust mass evolution

After calculating the dust masses for each galaxy in each TNG100 snapshot (see Section 3.2), we bin the galaxies into the same stellar mass bins used in our stacking analysis. For a given $(M_* - z)$ bin that has at least ten sources, we calculate the mean of the dust-to-stellar-mass ratio (M_d/M_*). The errors on the calculated dust-to-stellar-mass ratios are taken to be the standard error on the mean. Since the TNG100 snapshots are at different redshifts to the redshift bins used in our stacking analysis, we perform quadratic interpolations to the resulting TNG100 dust-to-stellar-mass ratios in each stellar mass bin, to allow a comparison with our stacking data. Here, we focus on comparisons between the observations and simulations when considering a constant $\epsilon_d = 0.5$. In Section 5.1, we return to this comparison and consider the impact of an evolving $\epsilon_d(z)$ with redshift.

Fig. 4 shows a comparison of the model predictions and observations. The model predicts dramatically less evolution than the observations (Fig. B1 shows just the TNG100 predictions with a scale chosen to highlight details of the predicted evolution more clearly). At low stellar masses both the simulation and observations agree that the peak value of (M_d/M_*) is at $z \sim 3$, but TNG100 predicts much less evolution than is seen (Fig. 4). It is interesting to note that the peak in the dust-to-stellar-mass ratio of galaxies is a little before the peak of the star formation history in the Universe. At low redshifts, the dust-to-stellar-mass ratios estimated for the TNG100 galaxies agrees well with the dust-to-stellar-mass ratios calculated in the stacking analysis for the observed galaxies, in all stellar mass bins, but we remind the reader that the agreement at low redshifts is by construction (see Section 4.1).

The simulation and the observations agree that at low redshift the galaxies with high stellar masses generally contain less dust than those with lower stellar masses. However, the difference between the predicted and observed evolution for the high-mass galaxies ($> \log_{10}(M_{*,\text{bin}}) \sim 10.5$) is dramatic (Fig. 4). The observed evolution is very large and continues out to the highest redshifts probed by our sample. In comparison, the simulated galaxies still have very small amounts of dust at high redshifts. Fig. 4 shows that the difference in dust-to-stellar-mass ratio evolution between the simulated and observed data is highest at the highest stellar masses – except for the highest stellar mass bin, where a lack of galaxies in both data sets prohibits meaningful conclusions on the evolution of dust-to-stellar-mass ratio to be drawn.

In the next section, we explore reasons to explain the conflicting dust mass evolution shown between observations and simulations.

5 DISCUSSION

5.1 Caveats and assumptions

First, we consider the effect of the fundamental assumptions we have made in this work. Two big assumptions we have made are that the following parameters are time invariant: (i) κ_{500} and (ii) T_d .

There is no obvious reason why we would expect κ_{500} to evolve with redshift, and we would need κ_{500} to have increased by a factor of ~ 5 by $z = 3.5$ for galaxies with stellar masses $\log_{10}(M_{*,\text{bin}}) = 10.5$ – 10.75 to bring the observations and the simulations into agreement. We are not aware of any predictions or observations showing such a large evolutionary effect in the properties of the dust itself, but we cannot rule this out.

In equation (1), a higher dust temperature would lead to a lower calculated dust mass. In this work, it is important to remember that the dust temperature assumed in equation (1) is the *mass-weighted* dust temperature, and therefore represents the temperature of the bulk of the ISM (Millard et al. 2020). As described in Millard et al. (2020), mass-weighted dust temperature depends on the mean radiation energy density to the power of $\sim 1/6$ (Scoville et al. 2016), and so vast differences in the ISM environment over the history of the Universe would be required to significantly change the value of T_d . Thus far, there is little evidence to support such differences. Further, simulations of $z = 2$ – 6 galaxies by Liang et al. (2019) showed that the mass-weighted dust temperature of galaxies evolves little over this redshift range, and that a mass-weighted dust temperature of 25 K is a not an unreasonable assumption. One factor to consider is the increasing temperature of the cosmic microwave background (CMB) with redshift (da Cunha et al. 2013) which can become non-negligible at $z > 4$. Using equation (10) from da Cunha et al. (2013), the expected rise in temperature due to the CMB for dust grains with $T_d = 25$ K at $z = 0$ would amount to only 0.4 K at $z = 5$.

Fig. 5 shows how the observed dust masses change for discrete redshifts depending on the assumed dust temperature. We can see that even for an extreme increase in the mass-weighted dust temperature to 35 K, the corresponding decrease in the estimated dust masses is not enough to reconcile the differences in the dust-to-stellar-mass ratio at high redshifts as shown in Fig. 4.

In equation (2), we assumed a value of the dust emissivity spectral index, β , of 1.8, robustly determined by Galactic ISM studies using sub-mm/mm emission from Planck (Planck Collaboration 2016). This is also in agreement with a long-wavelength study of 29 Submillimetre Galaxies (SMGs; rapidly star-forming, highly dust-obscured massive galaxies typically found at $z \sim 2$ – 4) at $z \sim 2.7$ by Chapin et al. (2009), who find an average $\beta = 1.75$. Since the

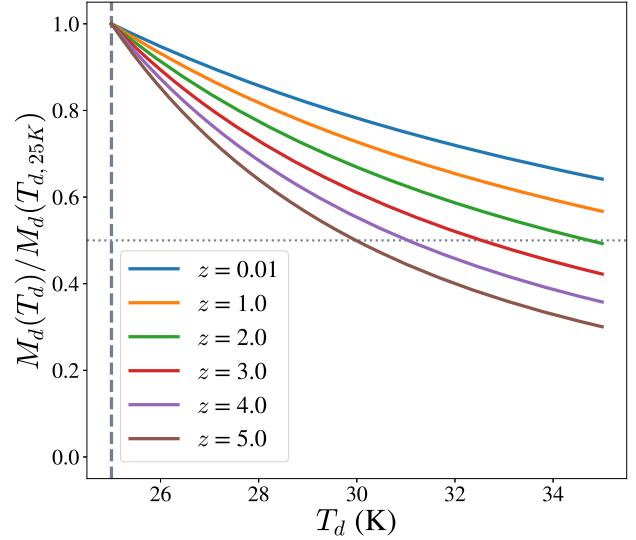


Figure 5. The variation of dust mass with assumed mass-weighted dust temperature for discrete redshifts. Vertical dashed line: the mass-weighted dust temperature assumed in this study (25 K). Horizontal dotted line: 50 per cent reduction in dust mass.

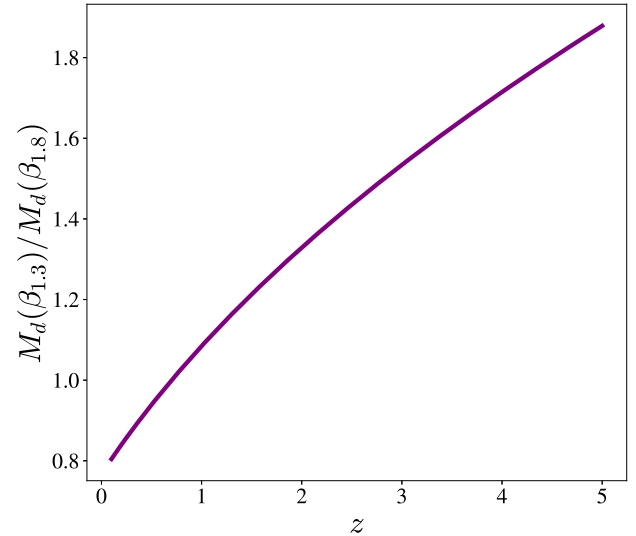


Figure 6. The variation of dust mass with redshift for two assumed values of the dust emissivity spectral index, β .

observational dust masses are inversely proportional to κ_{λ_e} , the choice of β directly impacts the resulting dust masses (equation 1).

Observational studies on extragalactic scales find values of $\beta = 1$ – 2 (e.g. Hildebrand 1983; Dunne & Eales 2001; Chapin et al. 2009; Clements, Dunne & Eales 2010). Although these studies typically converge towards the idea that a value of β towards the upper end of this range is appropriate for dust in galaxies, it is worth examining the impact the choice of a lower β would have on the calculated observational dust masses. Therefore, as an example, we consider the impact on our resulting dust masses had we chosen a significantly lower value of $\beta = 1.3$.

Fig. 6 shows the variation in observational dust mass with redshift for our assumed value of $\beta = 1.8$ and our test value of $\beta = 1.3$. For $z \lesssim 0.7$, we see that a lower value of β would lead to lower dust

masses compared to dust masses estimated using $\beta = 1.8$. However, beyond this redshift, we see that a lower value of β would lead to higher dust mass estimates, only exacerbating the difference between observations and the model seen in Fig. 4.

An important assumption that we have made in estimating the dust masses for TNG100 galaxies is that the fraction of metals locked up in dust grains, ϵ_d , has a constant value, (Section 3.2.1).

By setting its value to 0.5 we have left very little room to explain the discrepancy in Fig. 4 by evolution in ϵ_d . Even incorporating all the metals in a high-redshift galaxy in dust grains would only increase the value of ϵ_d to one, which would only double the high-redshift dust masses, much less than the discrepancy in Fig. 4. Such a scenario is not physical since the difference between the cosmic abundances of elements and the likely chemical composition of dust grains mean that large proportions of plentiful elements such as oxygen must remain in the gas phase (Meyer, Jura & Cardelli 1998). However, as a final check, we use the relationship of ϵ_d with z from Vogelsberger et al. (2020) (valid for $2 \leq z < 9$) and values from Inoue (2003) for $0 < z < 2$ (their Fig. 6) and re-run the analysis in Fig. 4. As before, we perform a quadratic interpolation to the discreet TNG100 results. We note that for the variation of ϵ_d with redshift, this results in a sharp transition at around $z \sim 1.7$ as we move from the $\epsilon_d(z)$ values of Inoue (2003) used for the $z = 1.5$ TNG100 file, to the $\epsilon_d(z)$ values of Vogelsberger et al. (2020) used for the $z \geq 2$ TNG100 files.

As shown in Fig. 4, this variation in ϵ_d with redshift provides a moderately better agreement with the observations at intermediate redshifts ($z \sim 2$) for the three lowest stellar mass bins, where the discrepancy between the observations and simulations (assuming a constant ϵ_d) is the least. For stellar mass bins $\log_{10}(M_{*,\text{bin}}) = 10.25\text{--}10.75$, there is a good agreement between the observations and simulations at intermediate redshifts when using $\epsilon_d(z)$. However, this agreement requires a value of $\epsilon_d(z = 2)$ close to 1, which, as previously explained, is physically unlikely, and is still not enough to explain the discrepancy at the highest redshifts ($z > 2.5$). Out to higher redshifts, and for the higher stellar mass bins, this evolving $\epsilon_d(z)$ does not fix the dichotomy in the evolution.

It is also worth noting that the evolving $\epsilon_d(z)$ leads to worse agreement between the observations and simulations at low redshifts ($z < 2$). The relation from Inoue (2003) gives a $\epsilon_d(z)$ that decreases with redshift. Therefore, the lack of agreement with a $\epsilon_d(z)$ that decreases with redshift, combined with the agreement with the high $\epsilon_d(z)$ from Vogelsberger et al. (2020) at intermediate redshifts and at low stellar masses, possibly indicates a requirement for an increasing $\epsilon_d(z)$ from $z = 0$ out to $z \sim 2\text{--}2.5$, where the star formation rate density of the Universe peaks (Daddi et al. 2007; Madau & Dickinson 2014). Fig. 4 also indicates that any variation in ϵ_d would need to have some dependence on galactic stellar mass. However, this is assuming that the COSMOS stacked observations are correct, and assuming that the only parameter that needs to be changed is ϵ_d .

Another parameter for which ϵ_d could vary is metallicity. To remove the discrepancy in Fig. 4 would require a value greater than the currently assumed value of 0.5 at higher redshifts. Fig. 1 shows that any variation in ϵ_d with metallicity (e.g. Rémy-Ruyer et al. 2014) tends to reduce the value below 0.5, locking even less metals into dust. Therefore an ϵ_d that varies with metallicity cannot explain the discrepancy.

One place where a change in ϵ_d may be important is in explaining the discrepancy between the observed and predicted low-redshift DMFs in Fig. 2 for the *lowest dust mass* galaxies. At low dust masses the TNG100 DMF is higher than the observed DMF. The galaxies with these low dust masses have lower stellar masses than the galaxies

in our COSMOS sample and are therefore more likely to resemble the low-metallicity systems for which there is some evidence that ϵ_d is lower than in the Milky Way (Section 3.2.1). Lowering ϵ_d by a factor of ~ 2 would be enough to resolve the discrepancy in the DMF at low dust mass, in agreement with the scatter in ϵ_d we see in galaxies in the local Universe (Fig. 1). However, this cannot be responsible for the discrepancy in the evolution observed in the low-redshift DMFs (Fig. 2), since models that predict a variation in ϵ_d over $0 < z < 0.5$ produces a difference that is too small to account for the evolution observed in Dunne et al. (2011). For example, Inoue (2003) find a variation of only a factor of 2 in ϵ_d over this range (Fig. 1).

5.2 IllustrisTNG: satellites versus centrals

Previous studies examining the neutral gas content of IllustrisTNG galaxies at $z = 0$ have found that environment is an important influencing factor. For example, Stevens et al. (2019) showed that $z = 0$ TNG100 satellite galaxies are typically a factor of > 3 poorer in HI than a central galaxy of the same stellar mass. In addition, they showed that the inherent neutral gas fractions of TNG100 satellite galaxies show a dip at around $\log_{10}(M_*/M_\odot) \sim 10.75$. They attribute the lack of gas in satellite galaxies to (i) gas lost via AGN feedback (satellites are less able to reattain any gas lost due to their lack of gravitational influence) and (ii) stripping and quenching. Further, Diemer et al. (2019) found a population of TNG100 galaxies, the vast majority of which are classified as satellites, that were devoid of gas. These galaxies were found to largely live in crowded environments, and so have likely been stripped of their gas by a larger halo host.

It is therefore not unreasonable to consider that such unusual gas-poor satellite galaxies may be masking more rapid dust mass evolution in TNG100 galaxies than is shown by the collective population. To test this, we split our simulated galaxy sample into satellites and centrals⁵ and separately examine the dust mass evolution for these two galactic populations (Figs 7 and 8).

Fig. 7 shows the DMFs for the satellite and central TNG100 galaxies. Considering first the central galaxies (Fig. 7; right), we can see that the $z = 0$ TNG100 DMF agrees remarkably well with the $z = 0$ DMF from Beeston et al. (2018) at low and high dust masses, although the knee of the TNG100 DMF sits slightly lower than the observed dust mass function. This shows that at $z = 0$, the dust masses of TNG100 central galaxies are well-modelled by the simulations plus our post-processing recipe.

Now examining the DMFs of the satellite galaxies (Fig. 7; left), we see that in comparison to observations, TNG100 satellite galaxies are particularly devoid of dust. This is not unexpected, considering the lack of neutral gas in satellite galaxies found by previous studies (e.g. Diemer et al. 2019; Stevens et al. 2019).

A comparison of Figs 2 and 7 shows that the excess number of low dust mass sources in the total DMF can be attributed to satellite galaxies with low gas content, probably due to a combination AGN feedback driving gas out of the gravitational influence of satellite galaxies, ram-pressure stripping, and quenching (Stevens et al. 2019).

However, the most striking and important feature of Fig. 7 is that splitting the TNG100 galaxies sample into satellite and centrals does not solve the lack of dust mass evolution over the redshift range $0 < z < 0.5$ in TNG100 galaxies as compared to observations. Both satellites and central galaxies show a lack of evolution. This indicates that the lack of dust mass evolution in TNG100 galaxies can

⁵According to the Subfind classification

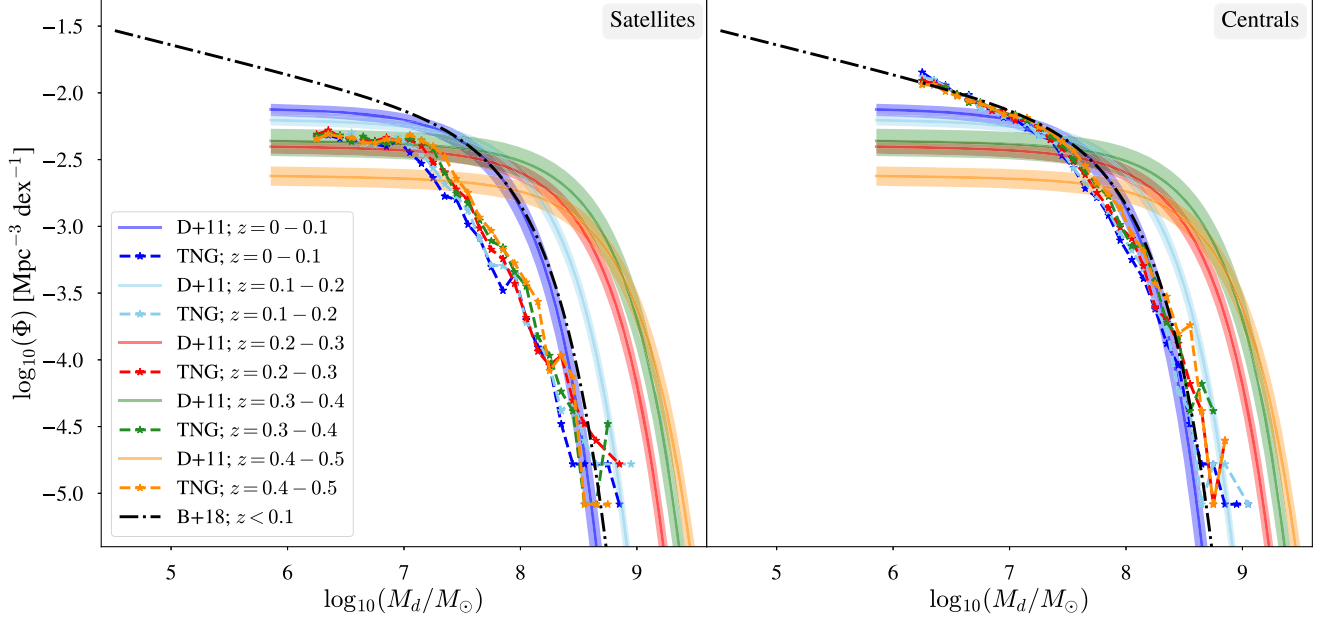


Figure 7. Low-redshift DMFs for IllustrisTNG galaxies (stars and dashed lines) and local DMFs from Beeston et al. (2018) (black dash-dot line) and Dunne et al. (2011) (solid colour lines). The IllustrisTNG sources have been split into two populations: satellite (left) and central galaxies (right). Literature DMFs are the same as in Fig. 2, and as before, have been scaled to IllustrisTNG cosmology and the values of the dust-mass opacity coefficient, κ_{500} , used in this paper.

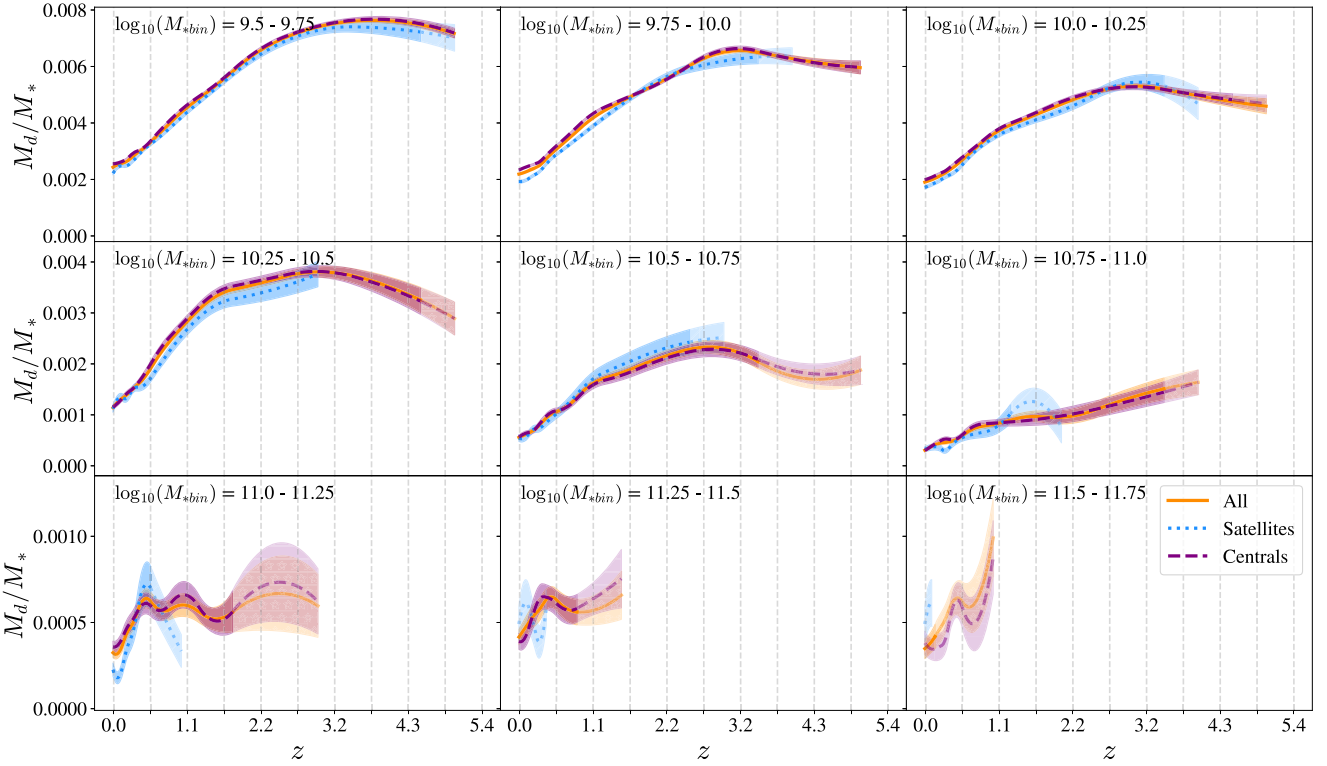


Figure 8. Dust-to-stellar-mass ratios for TNG100 galaxies plotted against redshift. Orange solid line: all galaxies. Purple dashed line: central galaxies. Blue dotted line: satellite galaxies. The paler, hatched areas represent data where the number of TNG100 sources in a given $(M_* - z)$ bin is between 10 and 50. Solid, non-hatched areas represent data where the number of sources in a given $(M_* - z)$ bin is at least 50. The lines are quadratic interpolations to discrete data points calculated at the redshifts of the TNG100 data files (Table 1). The orange line shown here is the same as in Figs B1 and 4.

be attributed to global simulation properties, rather than secondary effects that manifest due to environmental effects, for example.

Next, we explore the dust mass evolution of satellite and central IllustrisTNG galaxies out to high redshifts. Fig. 8 shows the evolution in the dust-to-stellar-mass ratio of TNG100 galaxies for galaxies with stellar masses $\log_{10}(M_*/M_\odot) \geq 9.5$. As shown in Fig. 8, the dust-to-stellar-mass ratio evolution of satellite and central galaxies are largely similar, particularly for stellar masses $\log_{10}(M_*/M_\odot) \leq 11.0$. Above this stellar mass limit (i.e. for galaxies in the three highest stellar mass bins), the dust-to-stellar-mass ratio evolution of satellite galaxies is a little different to that of centrals and the combined sample. However, we caution against inferring too much here, since there are few satellite galaxies in these bins, and the results suffer from low number statistics.

It is interesting to note that in some stellar mass bins, satellite galaxies have slightly higher dust-to-stellar-mass ratios compared to the central galaxies (e.g. $\log_{10}(M_{*,\text{bin}}) = 10.5\text{--}10.75$). In a given stellar mass bin, the average stellar mass of satellite galaxies is slightly lower than the average stellar mass of central galaxies. For galaxies with a similar dust mass, this therefore acts to marginally increase the dust-to-stellar-mass ratio for satellite galaxies.

Further, the similarity in the dust-to-stellar-mass ratio evolution of the two populations with redshift might seem strange when we consider the difference between the DMFs for satellite and central galaxies (Fig. 7). However, in Fig. 8, we are probing the highest stellar mass systems, and there is little difference between the high mass ends of the DMFs for the satellite and central galaxies.

Similar to Fig. 7, the most striking and important feature of Fig. 8 is that splitting the sample into satellite and central galaxies does not solve the lack of dust mass evolution over the redshift range $0 < z < 5$ in TNG100 galaxies as compared to observations.

We now explore global properties of the IllustrisTNG simulation in an attempt to explain the lack of evolution in the dust mass of galaxies over cosmic time as compared to observations. Specifically, in the next sections, we examine metal enrichment prescriptions (Section 5.3), the neutral gas content of galaxies (Section 5.4), and feedback mechanisms (Section 5.5).

5.3 IllustrisTNG metal enrichment prescriptions

As described in Torrey et al. (2019), in the IllustrisTNG simulations, metals are returned to the ISM by both supernovae explosions and asymptotic giant-branch (AGB) stars. In the TNG100 simulations (the data used in this study), individual stars are not resolved; star particles have masses of $m_* \sim 10^6 M_\odot$, and these star particles are assumed to encapsulate a Chabrier (2003) initial mass function IMF (Torrey et al. 2019). At each time-step in the simulation, stellar lifetime tables (Portinari, Chiosi & Bressan 1998) determine which stars leave the main sequence. Upon the death of stars, mass return and metal yield tables are used to determine the mass of metals returned to the ISM (Nomoto et al. 1997; Portinari et al. 1998; Kobayashi et al. 2006; Karakas 2010; Doherty et al. 2014; Fishlock et al. 2014). The returned metal mass is spread over the nearest 64 gas cells.

In IllustrisTNG, galactic winds are probabilistically estimated, based on local SFRs (Torrey et al. 2019; see Pillepich et al. 2018b for details). They are assigned a lower metallicity than the ISM of a galaxy ($Z_{\text{wind}} = 0.4Z_{\text{ISM}}$), to encapsulate the dilution of metal-rich gas by metal-poor gas as the winds drive the metal-rich gas away from their origin point.

In their investigations into the IllustrisTNG mass–metallicity relation, Torrey et al. (2019) found that a significant proportion of the metals of simulated galaxies lie outside of the cool ISM, where we would expect dust to form. They also found that higher-mass IllustrisTNG galaxies are less efficient at retaining metals compared to lower-mass counterparts. We see this reflected in Fig. 4, where the discrepancy between the evolution of dust shown in the observed and simulated galaxies is greater in the bins with higher stellar masses. Therefore, one possible explanation of the low dust masses in the model is that TNG100 ejects too many metals from the cool ISM.

5.4 IllustrisTNG gas mass evolution

Since we derive dust masses for IllustrisTNG galaxies using the neutral gas mass, the lack of evolution in the dust masses might stem from a lack of evolution in the mass of the cool ISM.

Several studies have compared the cold gas mass content of IllustrisTNG galaxies with observations of the local Universe. Stevens et al. (2019) found that the neutral gas fractions of TNG100 galaxies at $z = 0$ agree well with observations, although the most massive central galaxies ($M_* > 10^{10.7} M_\odot$) are somewhat too gas rich, and the gas content of satellite galaxies sharply dips at masses $M_* \sim 10^{10.5} M_\odot$. Similarly, Diemer et al. (2019) showed that the HI and H₂ mass fractions ($M_{\text{HI}+\text{H}_2}/M_*$) of TNG100 galaxies at $z = 0$ largely agree with observations, with some discrepancies in the H₂ mass fraction for the most massive galaxies ($M_* > 2 \times 10^{10} M_\odot$), which they found to be at least a factor of 4 lower compared to observations. However, Diemer et al. (2019) also found that the $z = 0$ TNG100 HI and H₂ mass functions largely agree with observations. This result is echoed in Davé et al. (2020), although they do note that the HI and H₂ mass functions of TNG100 are slightly too high at the high gas mass end.

It is worth noting that in the aforementioned studies (Davé et al. 2020; Diemer et al. 2019; Stevens et al. 2019), comparisons to observational data are not straight-forward – measuring the neutral gas content of galaxies accurately is notoriously difficult. For example, the CO-to-H₂ conversion factor, used to estimate the mass of molecular gas from observations of the CO line, is uncertain and varies from galaxy type to galaxy type (e.g. Bolatto, Wolfire & Leroy 2013; Béthermin et al. 2015; Genzel et al. 2015; Scoville et al. 2016; Popping et al. 2019), and galaxies can contain CO-dark molecular gas (Bolatto et al. 2013). When considering these observational limitations, these studies have shown that in the local universe, the neutral gas content of IllustrisTNG galaxies agrees well with observations.

Accurately measuring the gas content of galaxies at high redshifts is even more difficult. For example, although 21 cm radio emission is efficient at detecting HI gas in galaxies at $z = 0$, it is currently hard to detect 21 cm radiation at $z > \sim 0.2$ (Catinella et al. 2010). Typically beyond the local Universe, gas mass estimates are limited to molecular gas estimates, usually derived using CO as a tracer (e.g. Solomon & Vanden Bout 2005; Coppin et al. 2009; Tacconi et al. 2010; Casey et al. 2011; Bothwell et al. 2013; Carilli & Walter 2013; Tacconi et al. 2013; Combes 2018).

Despite these difficulties, there have been several attempts to compare observations of the evolution in the cool gas in galaxies with predictions by models. Interestingly, Davé et al. (2020) found that the IllustrisTNG HI mass function shows a negative evolution with redshift, and that the H₂ mass function is mostly unevolving with redshift (out to $z < 2$), implying that the gas content of IllustrisTNG galaxies varies little with redshift. They found that the evolution in HI and H₂ was much weaker than predicted by two other simulations: EAGLE (Davé et al. 2019) and SIMBA (Schaye et al. 2015). Similar

results were found by Diemer et al. (2019), who found a very weak evolution in the abundance of HI in TNG100 galaxies between $z = 1-4$, and a moderate evolution in the abundance of H₂ over the redshift range $z = 0-4$, with a peak increase of a factor of 2–3 at around $z = 2$. Recently, Popping et al. (2019) showed that the H₂ masses predicted by TNG100 for galaxies at $z > 1$ were a factor of 2–3 lower than observations (when comparing to observations using a standard CO-to-H₂ conversion factor).

All these studies suggest that TNG100 galaxies beyond $z \sim 1$ are lacking in neutral gas, and that there is little evolution in the gas content of the simulated galaxies with redshift. Here, by using post-processing methods to compare the results of IllustrisTNG with the evolution of dust in galaxies over cosmic time, we get a similar result. The lack of evolution in the dust mass in TNG100 compared to observations could therefore be a reflection of the lack of evolution of the neutral gas content in the simulations that has been seen previously. In the next section, we speculate as to what may be driving the different gas evolution, and therefore dust evolution, of the IllustrisTNG galaxies.

5.5 IllustrisTNG feedback mechanisms

Whether the lack of evolution in the dust predicted by TNG100 is entirely linked to the lack of evolution in the gas content (Section 5.4) or whether it is also linked to the ejection of metals, it is clearly linked to the model ejecting too much material outside a galaxy. It therefore seems likely to be caused by excessive galaxy outflows or feedback.

As discussed in Section 4.2, kinetic feedback from AGN in the form of black hole driven winds becomes the dominant feedback process over other feedback methods (e.g. stellar feedback) above galactic stellar masses of $\log_{10}(M_*/M_\odot) \sim 10.5$, particularly at late times. Kinetic AGN feedback in IllustrisTNG manifests in a two-fold manner – first, it acts to expel star-forming gas from galaxies. Secondly, at these stellar masses, the AGN feedback increases the gas cooling time, preventing radiative cooling and future gas accretion (Terrazas et al. 2020; Zinger et al. 2020). As noted in Davé et al. (2020), at earlier epochs, it is the second of these processes, thermal feedback, which is most important, reducing the amount of cold gas in high-redshift galaxies. The fact that the disagreement between the model predictions and the observations is greatest at the high stellar masses is at least circumstantial evidence that the explanation of the discrepancy might be too vigorous AGN feedback.

5.5.1 Insights from Illustris

Hayward et al. (2020) have recently investigated whether Illustris and IllustrisTNG can replicate the numbers of the rare luminous SMGs, using the star formation rates predicted by the simulations and post-processing estimates of the dust mass. They found that the predicted number counts of SMGs in Illustris was largely comparable with observations, but that IllustrisTNG notably lacked SMGs. Further investigation led to the revelation that the identified IllustrisTNG SMGs are particularly dust-poor, with some SMGs ($\log_{10}(M_*/M_\odot) \approx 11$) at $z \sim 2$ in Illustris having a factor of three higher dust content than the IllustrisTNG counterparts. Hayward et al. (2020) ultimately traced this dichotomy to a lack of gas, as opposed to a lack of metals, in these IllustrisTNG galaxies as compared to Illustris, driven by changes to the feedback model between Illustris and IllustrisTNG – either stellar feedback outflows, and/or AGN feedback. The changes to the feedback model were made to make quenching more efficient and so bring IllustrisTNG galaxies more in line with the observed $z = 0$ colour

bimodality (Weinberger et al. 2017, 2018; Nelson et al. 2018a). However, Hayward et al. (2020) point out that these changes seem to have quenched $z \sim 2-3$ galaxies too early, leading to a direct lack of massive dusty galaxies in IllustrisTNG as compared to observations.

Whilst the results of Hayward et al. (2020) are not directly comparable to this study due to their specific galactic population selection, the general idea that changes to feedback in IllustrisTNG have consequently quenched galaxies too soon or too frequently is in line with the results of this study. If the inability of IllustrisTNG plus our post-processing recipe to match the strong evolution in the dust masses is not caused by cosmic evolution in the properties of the dust grains themselves, it seems most likely that is caused by a lack of cool gas in the high-redshift TNG galaxies, possibly caused by over-enthusiastic AGN feedback.

5.6 Comparison with EAGLE

Recently, Baes et al. (2020) derived DMFs out to $z = 1$ for galaxies from the EAGLE simulation. Dust masses were estimated using modified blackbody fits to synthetic infrared luminosities at FIR wavelengths (160, 250, 350 and 500 μm) generated using a post-processing 3D radiative transfer procedure (Baes et al. 2011; Camps & Baes 2015; Camps et al. 2016, 2018). Similarly to the results of this study, in the local universe ($z < 0.1$), Baes et al. (2020) found that they could reproduce the shape and normalization of the DMF fairly well, getting very good agreement with the DMFs found by Dunne et al. (2011) and Beeston et al. (2018) for dust masses $M_d < 2 \times 10^7 M_\odot$ but predicting too few galaxies at higher masses.

When examining the evolution of the EAGLE DMF up to $z = 1$, Baes et al. (2020) found only a mild evolution in the characteristic mass of the modified Schechter functions fitted to the DMFs, and very little evidence for density evolution. They did, however, find a fairly good agreement with the weak evolution in the cosmic dust mass density found by Driver et al. (2018) for this redshift range.

It is interesting that both the EAGLE and IllustrisTNG simulations show a similar lack of evolution in the DMF as compared to observations, despite two different methods to estimate dust masses in post-processing, which shows that cosmological hydrodynamical simulations are still limited in their ability to reproduce the strong evolution seen in the global properties of dust in galaxies.

6 CONCLUSIONS

In this work, we have compared the observed evolution of the dust in galaxies with the predictions from a model. As our observations, we used previous estimates of the DMF over the redshift range $0 < z < 0.5$ and our own estimates of the mean ratio of dust mass to stellar mass over the redshift range $0 < z < 5$. We made our predictions using the cosmological hydrodynamical simulation IllustrisTNG (TNG100), using a simple post-processing recipe in which half the metals in the cool ISM are locked up in dust grains.

(i) We created ‘local’ DMFs based on galactic dust mass estimates for TNG100 galaxies out to $z = 0.5$, and compared these to previously observed DMFs over the same redshift range. We find the DMFs from TNG100 show little evolution, which is in stark contrast with the strong evolution seen in the empirical DMFs.

(ii) We find that the observed galactic population show a strong evolution in the dust-to-stellar-mass ratio (M_d/M_*) up to $z = 5$. For lower stellar mass bins ($\log_{10}(M_{*,\text{bin}}) < 10.75$), we find that the dust-to-stellar-mass ratio peaks at $z = 3-4$, but we cannot locate the peak for the high-mass galaxies because of the lack of high-mass

galaxies at high redshifts. We find that galaxies with a high stellar mass generally contain less dust than galaxies with a low stellar mass at the same redshift.

(iii) The model predicts much weaker evolution than the observations over the redshift range $0 < z < 5$. At lower stellar masses ($\log_{10}(M_{*,\text{bin}}) < 10.75$) the predicted evolution has a similar redshift dependence to the observed evolution but that for all stellar masses the strength of the evolution is much weaker than for the observations.

(iv) We split our simulated galaxies into two samples, satellites and centrals, and tested to see if the lack of dust mass evolution observed in the full sample was biased due to the dust content of one of these galaxy populations. We found that although satellite galaxies are typically more dust-poor than their central counterparts, both satellites and centrals show a similar lack of dust mass evolution as seen for the collective sample. Splitting the sample into satellite and central galaxies does not solve the lack of dust mass evolution over the redshift range $0 < z < 5$ in TNG100 galaxies as compared to observations.

(v) We find that it is very difficult to bring the observed and model dust mass evolution into agreement by changes in the assumptions underpinning our observations. The obvious ways of doing this are: (i) a drastic evolution in the dust-mass opacity coefficient with redshift; (ii) a non-constant dust-to-metals ratio; (iii) an extreme increase in the mass-weighted dust temperature with redshift; or (iv) an extremely high value of $\beta \gg 2$. The third and fourth of these are inconsistent with observations. Therefore the only possible ways, from the observational side, of making the observations and theory consistent would be to assume that the properties of the dust itself change drastically with redshift. This would require the dust-mass opacity coefficient to be much higher at high redshifts than at low redshifts. There is no particular reason to think this should happen but we cannot rule this out. We also show that the results of our study are robust against changes to our assumption about the fraction of metals bound up in dust grains: varying ϵ_d with metallicity cannot account for the discrepancy we observe. An ϵ_d that varies with redshift, however, can reduce, though not eliminate, the discrepancy if (i) it increases with redshift (ii) is a strong function of stellar mass and (iii) is extremely high (≥ 0.8 would be needed in the higher redshift bins).

(vi) Although determining the root cause of the discrepancies between simulations and observations is difficult, we attribute the differences to one or more of the following in the models: (i) excessive galactic winds driving metals out of the cool ISM where we expect dust to form; (ii) a lack of evolution in the neutral gas content of galaxies with redshift; (iii) kinetic feedback from AGN expelling gas from galaxies.

(vii) We note that a lack of evolution in the dust content of galaxies as compared to observations is not limited to IllustrisTNG, but has also been in the $z < 1$ DMFs calculated for the EAGLE simulation, despite using a different post-processing technique for estimating dust masses.

Previous studies of neutral gas ($\text{HI} + \text{H}_2$) have concluded that the neutral gas content in IllustrisTNG does not show sufficient evolution with redshift. However, at high redshifts, H_2 and (more so) HI are hard to measure. It is much easier to measure the dust masses of high-redshift galaxies than their gas masses. By using IllustrisTNG with a simple post-processing technique we have been able to predict the dust masses of high-redshift galaxies. We find that the observed evolution is much stronger than the predicted evolution. If the discrepancy is not produced by cosmic evolution in the properties of interstellar dust itself, the most likely explanation seems to be that

TNG does not predict strong enough evolution in the neutral gas content of galaxies.

ACKNOWLEDGEMENTS

We thank Claudia Lagos, Dylan Nelson, and Annalisa Pillepich for helpful comments on the draft manuscript. We also thank Matthieu Bethermin for a productive and thorough referee report, which significantly improved this paper. JSM, HLG and RAB acknowledge support from the European Research Council in the form of Consolidator Grant CosmicDust (ERC-2014-CoG-647939, PI: H. L. Gomez). SAE and MWLS gratefully acknowledge the support of a Consolidated Grant (ST/K00926/1) from the UK Science and Technology Facilities Council (STFC). JSM would also like to thank N Parry and O Millard-Parry for their support.

The James Clerk Maxwell Telescope is operated by the East Asian Observatory on behalf of The National Astronomical Observatory of Japan; Academia Sinica Institute of Astronomy and Astrophysics; the Korea Astronomy and Space Science Institute; the Operation, Maintenance and Upgrading Fund for Astronomical Telescopes and Facility Instruments, budgeted from the Ministry of Finance (MOF) of China and administrated by the Chinese Academy of Sciences (CAS), as well as the National Key R&D Program of China (No. 2017YFA0402700). Additional funding support is provided by the Science and Technology Facilities Council of the United Kingdom and participating universities in the United Kingdom and Canada (ST/M007634/1, ST/M003019/1, and ST/N005856/1). The James Clerk Maxwell Telescope has historically been operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the United Kingdom, the National Research Council of Canada and the Netherlands Organization for Scientific Research and data from observations undertaken during this period of operation is used in this manuscript. This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. The data used in this work were taken as part of Program ID M16AL002.

This research has made use of Astropy, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013; Price-Whelan et al. 2018). Physical parameters in IllustrisTNG were calculated using COLOSSUS, a public, open-source PYTHON package for calculations related to cosmology, the large-scale structure of matter in the universe, and the properties of dark matter haloes (Diemer 2018). This research also used the PYTHON libraries SciPy (Virtanen et al. 2020), NumPy (van der Walt, Colbert & Varoquaux 2011), Matplotlib (Hunter 2007), and pandas (McKinney 2010). We thank Dr Siyu Gao for the development of curlyBrace.

DATA AVAILABILITY

The data used in this study are largely publicly available and can be shared upon reasonable request to the corresponding author.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Ahn C. P. et al., 2014, *ApJS*, 211, 17
- Andrews S. K., Driver S. P., Davies L. J. M., Kaffle P. R., Robotham A. S. G., Wright A. H., 2017, *MNRAS*, 464, 1569
- Asano R. S., Takeuchi T. T., Hirashita H., Inoue A. K., 2013, *Earth Planets Space*, 65, 213
- Astropy Collaboration, 2013, *A&A*, 558, A33

- Baes M., Verstappen J., De Looze I., Fritz J., Saitly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, *ApJ*, 196, 22
- Baes M. et al., 2020, *MNRAS*, 494, 2912
- Beeston R. A. et al., 2018, *MNRAS*, 479, 1077
- Béthermin M. et al., 2015, *A&A*, 573, A113
- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, 51, 207
- Boselli A. et al., 2010, *PASP*, 122, 261
- Bothwell M. S. et al., 2013, *MNRAS*, 429, 3047
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Camps P., Baes M., 2015, *Astron. Comput.*, 9, 20
- Camps P., Trayford J. W., Baes M., Theuns T., Schaller M., Schaye J., 2016, *MNRAS*, 462, 1057
- Camps P. et al., 2018, *ApJS*, 234, 20
- Carilli C. L., Walter F., 2013, *ARA&A*, 51, 105
- Casey C. M. et al., 2011, *MNRAS*, 415, 2723
- Catinella B. et al., 2010, *MNRAS*, 403, 683
- Chabrier G., 2003, *PASP*, 115, 763
- Chapin E. L. et al., 2009, *MNRAS*, 398, 1793
- Charlot S., Fall S. M., 2000, *ApJ*, 539, 718
- Chawner H. et al., 2019, *MNRAS*, 483, 70
- Chawner H. et al., 2020, *MNRAS*, 493, 2706
- Ciesla L. et al., 2015, *A&A*, 576, A10
- Cigan P. et al., 2019, *ApJ*, 886, 51
- Clark C. J. R., Schofield S. P., Gomez H. L., Davies J. I., 2016, *MNRAS*, 459, 1646
- Clemens M. S. et al., 2013, *MNRAS*, 433, 695
- Clements D. L., Dunne L., Eales S., 2010, *MNRAS*, 403, 274
- Combes F., 2018, *A&AR*, 26, 5
- Cool R. J. et al., 2013, *ApJ*, 767, 118
- Coppin K. E. K. et al., 2009, *MNRAS*, 395, 1905
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- da Cunha E., Charlot S., Elbaz D., 2008, *MNRAS*, 388, 1595
- da Cunha E. et al., 2013, *ApJ*, 766, 13
- Daddi E. et al., 2007, *ApJ*, 670, 156
- Davies L. J. M. et al., 2015, *MNRAS*, 447, 1014
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, *MNRAS*, 497, 146
- De Cia A., Ledoux C., Savaglio S., Schady P., Vreeswijk P. M., 2013, *A&A*, 560, A88
- De Looze I., Barlow M. J., Swinyard B. M., Rho J., Gomez H. L., Matsuura M., Wesson R., 2017, *MNRAS*, 465, 3309
- De Looze I. et al., 2019, *MNRAS*, 488, 164
- De Vis P. et al., 2017a, *MNRAS*, 464, 4680
- De Vis P. et al., 2017b, *MNRAS*, 471, 1743
- De Vis P. et al., 2019, *A&A*, 623, A5
- Diemer B., 2018, *ApJS*, 239, 35
- Diemer B. et al., 2018, *ApJS*, 238, 33
- Diemer B. et al., 2019, *MNRAS*, 487, 1529
- Doherty C. L., Gil-Pons P., Lau H. H. B., Lattanzio J. C., Siess L., 2014, *MNRAS*, 437, 195
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Dole H. et al., 2006, *A&A*, 451, 417
- Draine B. T., 2009, in Henning T., Grün E., Steinacker J., eds, ASP Conf. Ser. Vol. 414, Cosmic Dust - Near and Far. Astron. Soc. Pac., San Francisco, p. 453
- Driver S. P. et al., 2011, *MNRAS*, 413, 971
- Driver S. P. et al., 2018, *MNRAS*, 475, 2891
- Dunne L., Eales S. A., 2001, *MNRAS*, 327, 697
- Dunne L., Eales S., Edmunds M., Ivison R., Alexander P., Clements D. L., 2000, *MNRAS*, 315, 115
- Dunne L., Eales S. A., Edmunds M. G., 2003a, *MNRAS*, 341, 589
- Dunne L., Eales S., Ivison R., Morgan H., Edmunds M., 2003b, *Nature*, 424, 285
- Dunne L. et al., 2011, *MNRAS*, 417, 1510
- Dwek E., Galliano F., Jones A. P., 2007, *ApJ*, 662, 927
- Eales S. et al., 2009, *ApJ*, 707, 1779
- Eales S. et al., 2010, *PASP*, 122, 499
- Eales S. et al., 2012, *ApJ*, 761, 168
- Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, *PASP*, 110, 761
- Fishlock C. K., Karakas A. I., Lugaro M., Yong D., 2014, *ApJ*, 797, 44
- Fixsen D. J., Dwek E., Mather J. C., Bennett C. L., Shafer R. A., 1998, *ApJ*, 508, 123
- Geach J. E. et al., 2017, *MNRAS*, 465, 1789
- Genel S. et al., 2014, *MNRAS*, 445, 175
- Genzel R. et al., 2015, *ApJ*, 800, 20
- Gomez H. L. et al., 2012, *ApJ*, 760, 96
- Hayward C. C. et al., 2020, preprint ([arXiv:2007.01885](https://arxiv.org/abs/2007.01885))
- Hildebrand R. H., 1983, *QJRAS*, 24, 267
- Holland W. S. et al., 2013, *MNRAS*, 430, 2513
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Inoue A. K., 2003, *PASJ*, 55, 901 (I03)
- James A., Dunne L., Eales S., Edmunds M. G., 2002, *MNRAS*, 335, 753
- Karakas A. I., 2010, *MNRAS*, 403, 1413
- Karim A. et al., 2011, *ApJ*, 730, 61
- Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, *ApJ*, 653, 1145
- Lagos C. d. P. et al., 2019, *MNRAS*, 489, 4196
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Lee N. et al., 2015, *ApJ*, 801, 80
- Le Fèvre O. et al., 2013, *A&A*, 559, A14
- Liang L. et al., 2019, *MNRAS*, 489, 1397
- Lilly S. J. et al., 2007, *ApJS*, 172, 70
- Li Q., Narayanan D., Davé R., 2019, *MNRAS*, 490, 1425
- McKinney W., 2010, in van der Walt S., Millman J., eds, Proceedings of the 9th Python in Science Conference. SciPy2010, p. 56
- McKinnon R., Torrey P., Vogelsberger M., 2016, *MNRAS*, 457, 3775
- McKinnon R., Kannan R., Vogelsberger M., O'Neil S., Torrey P., Li H., 2019, preprint ([arXiv:1912.02825](https://arxiv.org/abs/1912.02825))
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Madden S. C. et al., 2013, *PASP*, 125, 600
- Marinacci F. et al., 2018, *MNRAS*, 480, 5113
- Matsuura M. et al., 2011, *Science*, 333, 1258
- Meyer D. M., Jura M., Cardelli J. A., 1998, *ApJ*, 493, 222
- Michałowski M. J. et al., 2017, *MNRAS*, 469, 492
- Millard J. S. et al., 2020, *MNRAS*, 494, 293 (Paper I)
- Morgan H. L., Edmunds M. G., 2003, *MNRAS*, 343, 427
- Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
- Nelson D. et al., 2018a, *MNRAS*, 475, 624
- Nelson D. et al., 2018b, *MNRAS*, 477, 450
- Nelson D. et al., 2019, *Comput. Astrophys. Cosmol.*, 6, 2
- Nomoto K., Iwamoto K., Nakasato N., Thielemann F. K., Brachwitz F., Tsujimoto T., Kubo Y., Kishimoto N., 1997, *Nucl. Phys. A*, 621, 467
- Pilbratt G. L. et al., 2010, *A&A*, 518, L1
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Planck Collaboration, 2016, *A&A*, 594, A13
- Popping G., Somerville R., Galametz M., 2017, *MNRAS*, 471, 3152
- Popping G. et al., 2019, *ApJ*, 882, 137
- Portinari L., Chiosi C., Bressan A., 1998, *A&A*, 334, 505
- Price-Whelan A. M. et al., 2018, *AJ*, 156, 123
- Puget J.-L., Abergel A., Bernard J.-P., Boulanger F., Burton W. B., Desert F.-X., Hartmann D., 1996, *A&A*, 308, L5
- Roman-Duval J. et al., 2019, *ApJ*, 871, 151
- Rémy-Ruyer A. et al., 2014, *A&A*, 563, A31
- Salpeter E. E., 1974, *ApJ*, 193, 579
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Schulz S., Popping G., Pillepich A., Nelson D., Vogelsberger M., Marinacci F., Hernquist L., 2020, *MNRAS*, 497, 4773
- Scoville N. et al., 2007, *ApJS*, 172, 1
- Scoville N. et al., 2014, *ApJ*, 783, 84
- Scoville N. et al., 2016, *ApJ*, 820, 83

Scoville N. et al., 2017, *ApJ*, 837, 150
 Shen X. et al., 2020, *MNRAS*, 495, 4747
 Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *MNRAS*, 452, 575
 Simpson J. M. et al., 2019, *ApJ*, 880, 43
 Solomon P. M., Vanden Bout P. A., 2005, *ARA&A*, 43, 677
 Springel V., 2010, *MNRAS*, 401, 791
 Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
 Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
 Springel V. et al., 2018, *MNRAS*, 475, 676
 Stevens A. R. H. et al., 2019, *MNRAS*, 483, 5334
 Tacconi L. J. et al., 2010, *Nature*, 463, 781
 Tacconi L. J. et al., 2013, *ApJ*, 768, 74
 Tacconi L. J. et al., 2018, *ApJ*, 853, 179
 Terrazas B. A. et al., 2020, *MNRAS*, 493, 1888
 Torrey P., Vogelsberger M., Sijacki D., Springel V., Hernquist L., 2012, *MNRAS*, 427, 2224
 Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, *MNRAS*, 438, 1985
 Torrey P. et al., 2019, *MNRAS*, 484, 5587
 Valiante R., Schneider R., Bianchi S., Andersen A. C., 2009, *MNRAS*, 397, 1661
 van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22

Virtanen P. et al., 2020, *Nat. Methods*, 17, 261
 Vlahakis C., Dunne L., Eales S., 2005, *MNRAS*, 364, 1253
 Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031
 Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518
 Vogelsberger M. et al., 2014b, *Nature*, 509, 177
 Vogelsberger M. et al., 2020, *MNRAS*, 492, 5167 (V20)
 Weinberger R. et al., 2017, *MNRAS*, 465, 3291
 Weinberger R. et al., 2018, *MNRAS*, 479, 4056
 Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, *ApJ*, 754, L29
 Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, *MNRAS*, 399, 574
 Wiseman P., Schady P., Bolmer J., Krühler T., Yates R. M., Greiner J., Fynbo J. P. U., 2017, *A&A*, 599, A24
 Yajima H., Nagamine K., Thompson R., Choi J.-H., 2014, *MNRAS*, 439, 3073
 Zafar T., Watson D., 2013, *A&A*, 560, A26
 Zhukovska S., Dobbs C., Jenkins E. B., Klessen R. S., 2016, *ApJ*, 831, 147
 Zinger E. et al., 2020, *MNRAS*, 499, 768

APPENDIX A: STELLAR MASS DISTRIBUTIONS

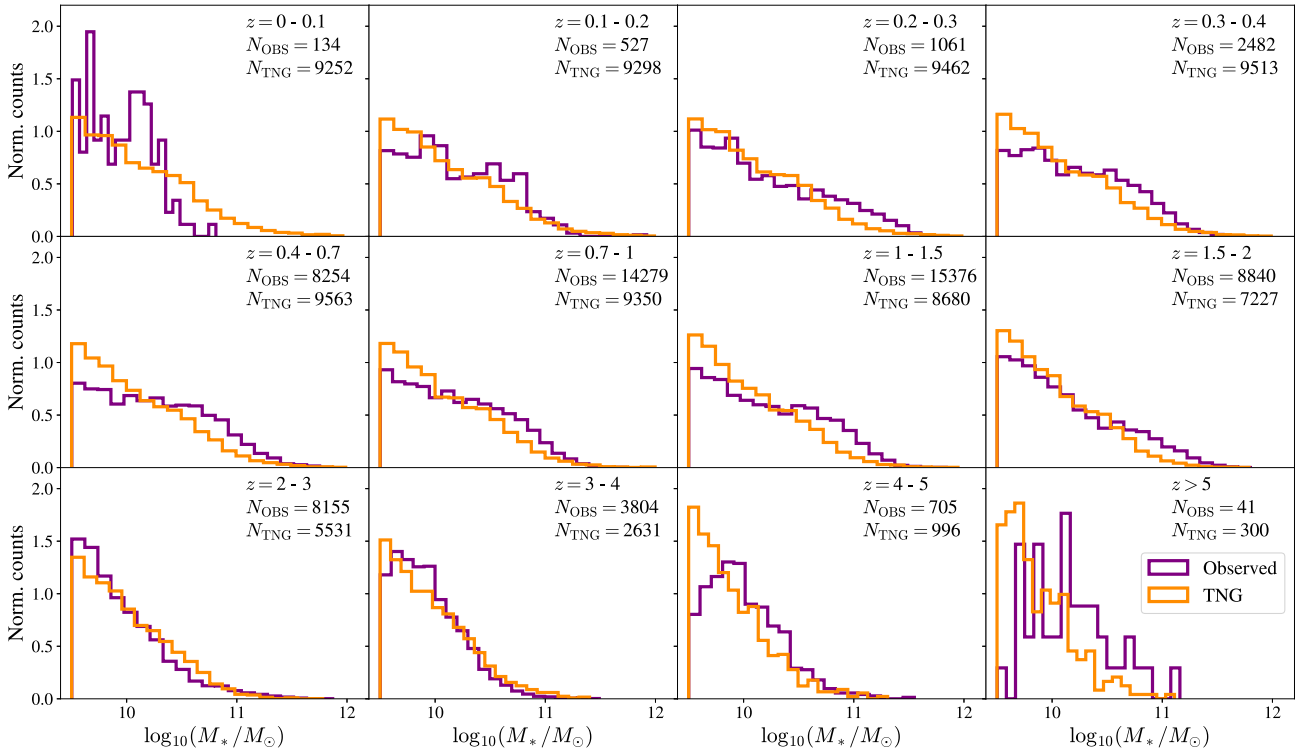


Figure A1. Distributions of the stellar masses of galaxies in different redshift bins for TNG100 and MAGPHYS-COSMOS sources. TNG100 stellar masses are calculated using particles and cells within $2r_{0.5}$. Purple: MAGPHYS-COSMOS sources used in stacking analysis to determine average dust properties. Orange: TNG100 galaxies.

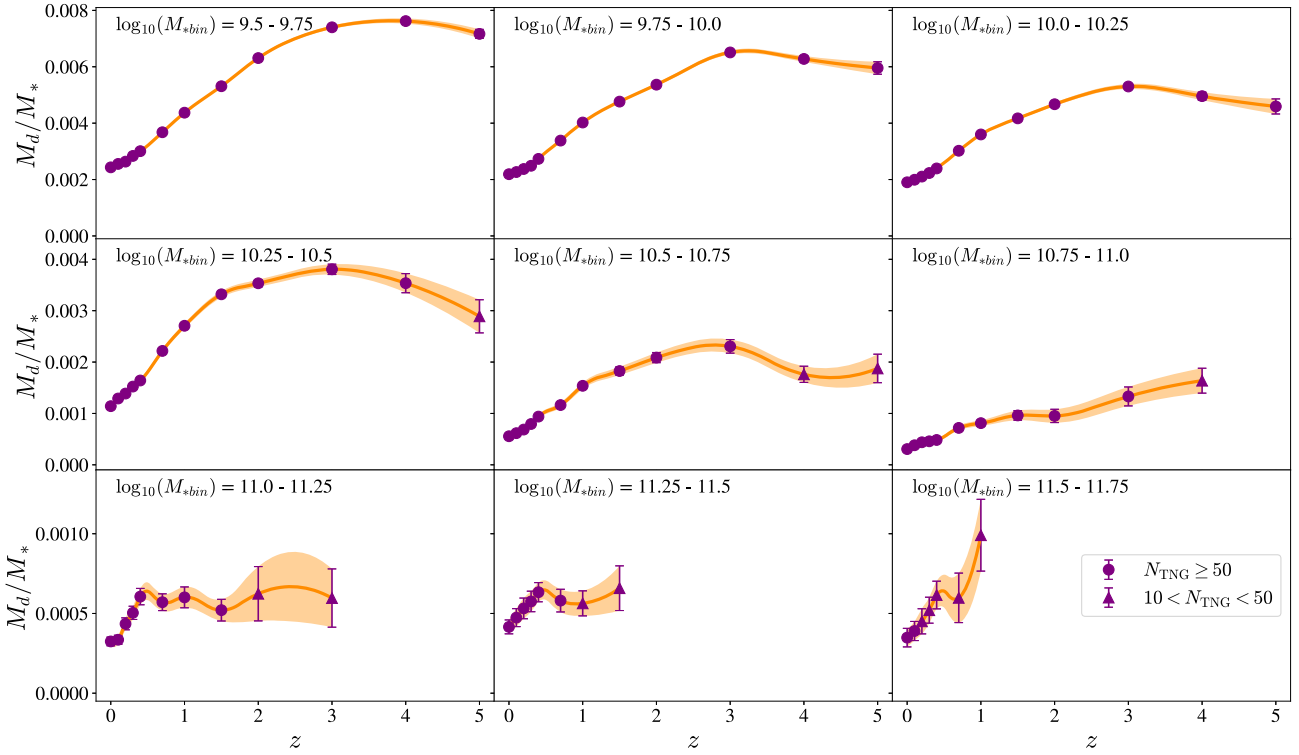
APPENDIX B: TNG100
DUST-TO-STELLAR-MASS RATIOS


Figure B1. Dust-to-stellar-mass ratios obtained from TNG100 simulations in redshift and stellar mass bins. Circles: at least 50 sources in the $(M_* - z)$ bin. Triangles: between 10 and 50 sources in the $(M_* - z)$ bin. The line is a quadratic interpolation to the data. Note that the scale on the y-axis has been chosen to make clear the evolution that is predicted by TNG100 and is not the same as the scale used in Fig. 4.

This paper has been typeset from a \LaTeX file prepared by the author.