

# TNG SAM: Bridging Hydrodynamical Complexity and Semi-Analytic Efficiency to Model Galaxy Formation

Osase Omoruyi,<sup>1\*</sup> Bryan Terrazas,<sup>2</sup> Yossi Oren,<sup>3</sup> Austen Gabrielpillai,<sup>4,5</sup> Viraj Pandya,<sup>4</sup> Rachel S. Somerville,<sup>6</sup> Lars Hernquist<sup>1</sup>

<sup>1</sup>Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA

<sup>2</sup>Department of Physics & Astronomy, Oberlin College, Oberlin, OH, 44074, USA

<sup>3</sup>School of Physics and Astronomy, Tel Aviv University, Ramat Aviv 69978, Israel

<sup>4</sup>Columbia University, Department of Astronomy, New York, NY 10027, USA

<sup>5</sup>The City University of New York, 365 5th Ave, New York, NY 10016, USA

<sup>6</sup>Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

All cosmological models of galaxy formation weigh physical accuracy against computational efficiency. Computationally expensive hydrodynamical simulations provide detailed insights into the co-evolution of dark matter, gas, stars, and black holes, but still rely on simplified subgrid models for small-scale processes (e.g., star formation). In contrast, semi-analytic models (SAMs) achieve efficiency by relying entirely on subgrid models, limiting their physical fidelity. In this work, we leverage the strengths of the Santa Cruz SAM and the IllustrisTNG 100 Mpc hydrodynamical simulation to develop the TNG SAM. Calibrated to emulate the underlying flow of baryons in stellar feedback-dominated TNG galaxies ( $\sim 10^{10} M_{\odot} < M_{200} < 10^{12} M_{\odot}$ ), the TNG SAM introduces several key updates to the Santa Cruz SAM framework: 1) a prescription for the efficiency of halo gas (re-)accretion, 2) a revised cooling recipe for gas transitioning from the hot halo to the galaxy based on the cooling time, eliminating the traditional “cold mode” vs. “hot mode” dichotomy, 3) the incorporation of both galactic and halo-scale outflows, 4) a star formation prescription that depends on how efficiently cold gas forms stars, and 5) metallicity-dependent mass loading factors that describe the circulation of metals between galaxies and their surroundings. These updates not only enable the TNG SAM to reproduce TNG’s underlying flow of gas and metals from the scale of the galaxy to the halo, but also global galaxy (e.g., stellar mass) and halo (e.g. baryon fraction) properties within  $\sim 30\%$  accuracy up to  $z = 6$ . This work demonstrates that, with proper calibration, SAMs can effectively capture the complex physics of galaxy formation seen in hydrodynamical simulations, offering a flexible and physically motivated framework for studying galaxy evolution across the large cosmological volumes targeted by future surveys.

**Key words:** galaxies: evolution – galaxies: formation – galaxies: halos – methods: numerical

## 1 INTRODUCTION

Observations across the electromagnetic spectrum have revealed a Universe rich in dark matter, stars, gas, and dust, sparking numerous questions regarding their formation and evolution. The  $\Lambda$  Cold Dark Matter ( $\Lambda$ CDM) model of cosmological structure formation proposes that the Universe consists of a complex filamentary network molded by gravity and populated with dark matter halos. Within these halos, baryonic matter falls to the center of the gravitational potential well, condenses, and forms galaxies (White & Rees 1978). To study this process from cosmological to stellar scales, numerous studies have used numerical simulations to predict and correlate observable galaxy properties—such as stellar mass, cold gas mass, and rate of star formation—with dark matter halo characteristics (see Somerville

& Davé (2015) for a comprehensive review). With the advent of powerful observatories like *JWST* and the upcoming *Nancy Grace Roman Space Telescope*, our ability to observe galaxies in the earliest days of the Universe and across larger cosmic volumes is rapidly improving, necessitating the development of more sophisticated and flexible cosmological models of galaxy formation.

Modeling how galaxies form and evolve requires investigating how they interact with their surrounding environment. As dark matter halos grow, galaxies accrete gas from the intergalactic medium (IGM) into the circumgalactic medium (CGM), connecting the large-scale cosmic web with the galaxies at the centers of these halos. Gravity continues to drive flows of gas from the CGM into the interstellar medium (ISM), fueling star formation. As stars form, evolve, and eventually die, the resulting feedback from stellar winds, supernovae (SNe), and black hole activity heats and/or expels material back into the CGM or IGM. This matter can later be re-accreted, fueling new

\* E-mail: osase.omoruyi@gmail.com

star formation. This cycle of gas accretion, star formation, feedback, and re-accretion—known as the “baryon cycle”—is fundamental to shaping how galaxies evolve.

To model the baryon cycle within the  $\Lambda$ CDM paradigm, two primary techniques have emerged: numerical hydrodynamical simulations and semi-analytic models (SAMs). Improving upon dark-matter and gravity-only simulations to include baryonic physics, numerical hydrodynamical simulations explicitly solve equations of gravity, (magneto)hydrodynamics, and thermodynamics. This comprehensive approach enables them to self-consistently model the coupled evolution of dark matter, gas, stars, and black holes, providing detailed information on each component’s spatial distribution, temperature, and kinematics. However, simulating large cosmological volumes requires sacrificing capturing the detailed physics of key processes occurring below their resolution limits, such as star formation and black hole feedback. As a result, these simulations rely on phenomenological “subgrid recipes” that contain adjustable parameters calibrated to reproduce a given set of observations. Additionally, the computational expense of hydrodynamical simulations also limits their ability to explore variations in subgrid models across vast volumes.

In contrast, SAMs offer a computationally efficient alternative by treating galaxies as unresolved entities and relying solely on “subgrid” recipes to simulate their formation and evolution. Built on merger trees extracted from dark matter-only N-body simulations or constructed using semi-analytic approaches based on the extended Press-Schechter formalism (Bond et al. 1991; Bower 1991; Lacey & Cole 1993) SAMs bridge the gap between the hierarchical growth of dark matter halos and observed galaxy populations by employing simplified analytical models of key physical processes relevant to galaxy evolution (White & Frenk 1991; Cole 1991; Lacey & Silk 1991). In a SAM, a galaxy is typically represented as a collection of mass reservoirs, such as the stellar disk, stellar bulge, cold gas disk, hot halo gas, and ejected material. As a result, SAMs essentially function as “flow models,” solving ordinary differential equations to track the flow of baryons between each component, with the flows governed by the physical prescriptions included in the model. The specific processes included in a given SAM—and the level of detail used to model them—lie at the discretion of the model’s creator. This flexibility has led to a range of implementations, from comprehensive frameworks like Santa Cruz (Somerville et al. 2008b), SAGE (Croton et al. 2006), GALFORM (Lacey et al. 2016), and SHARK (Lagos et al. 2018), to modular platforms that support user-defined prescriptions and calibration strategies (e.g., Benson 2012). It also allows SAMs to be tailored to specific scientific questions and computational resources, making them a valuable tool for understanding the “big picture” of the complexities shaping galaxy evolution.

Despite their different approaches, both SAMs and large-volume hydrodynamical simulations share a reliance on phenomenological prescriptions to simulate fundamental processes in the baryon cycle. This has led to the practice of “fine-tuning” or calibrating these models to reproduce a selection of observations, introducing additional uncertainty when comparing and interpreting results. Early comparisons of SAMs and hydrodynamical simulations (e.g. Benson et al. 2001; Helly et al. 2003; Yoshida et al. 2002; Saro et al. 2010; Stringer et al. 2010; Hirschmann et al. 2012) found general agreement in predicting properties such as the overall distribution of galaxies and their cold gas mass fractions, but differences in properties like the efficiency of gas cooling, star formation histories, and the merger history of specific galaxies. More detailed comparisons (e.g. Stringer et al. 2010; Monaco et al. 2014; Guo et al. 2016; Mitchell et al. 2018; Côté et al. 2018) similarly found that SAMs can ap-

proximate key galaxy properties such as stellar mass functions and specific star formation rates to a reasonable degree of accuracy when compared to hydrodynamical simulations. However, they also highlighted significant differences in the treatment of gas cooling, star formation efficiencies, galaxy sizes, and the evolution of star formation rate density. These results suggest that discrepancies between the two methods may arise from specific parameterizations and models used, rather than inherent limitations of each approach.

Recognizing SAMs and hydrodynamical simulations as complementary, rather than competing, techniques, can significantly enhance their predictive power. For example, the detailed physical insights from hydrodynamical simulations can refine the subgrid models in SAMs, as demonstrated by Pandya et al. (2020), who compared the FIRE-2 hydrodynamic simulations (Hopkins et al. 2018) with the Santa Cruz SAM (SC SAM) for halos with masses between  $10^{10}$  and  $10^{12} M_{\odot}$ . They found substantial agreement between FIRE-2 and the SC SAM for stellar and cold gas masses, but significant disagreements in the properties of the hot halo gas — a measurement not widely available for calibrating or validating models, particularly for Milky Way and lower mass halos. By adjusting the SAM to account for the suppression of halo gas accretion via stellar winds, particularly in dwarf galaxies, Pandya et al. (2020) reproduced the reduced halo gas accretion efficiencies of the FIRE-2 complementary galaxies remarkably well.

Similarly, Côté et al. (2018) refined GAMMA, a SAM aimed at understanding the chemical evolution of low-mass galaxies and the origins of metal-poor stars, by calibrating it against the observable properties of the most massive galaxy in the Wise et al. (2011) high-redshift hydrodynamical simulation. They found that implementing a non-uniform mixing model was key to reproducing the observable quantities in the hydrodynamical simulation. Lastly, in a more extensive effort to incorporate insights from hydrodynamical simulations into semi-analytic models, Pandya et al. 2023 used FIRE-2’s predictions for mass and energy flows in a given set of halos to calibrate their SAPPHIRE SAM, which emphasizes the role of the CGM and its energy content in governing galaxy evolution. These calibrations allowed their model to successfully emulate the detailed baryon cycle found in FIRE-2 halos.

Of course, SAMs can also enhance the predictive power of hydrodynamical simulations. For instance, the IllustrisTNG model for galactic-scale, star formation-driven kinetic winds incorporates several refinements over its predecessor in Illustris, described in detail in Section 2.1. Among these, the TNG model introduced a scaling of wind velocity with the local dark matter velocity dispersion,  $\sigma_{\text{DM}}$ , and a redshift-dependent factor to better match observed stellar mass functions and luminosity functions across redshift (Henriques et al. 2013). These modifications—rooted in SAM-driven findings—enhanced the realism of feedback prescriptions in IllustrisTNG, showcasing the bidirectional synergy between these modeling approaches.

In this work, we focus on the former direction — improving the fidelity of SAMs with insights from hydrodynamical simulations — by comparing the predictions of IllustrisTNG (hereafter TNG) with the Santa Cruz SAM, both of which have demonstrated success in reproducing a wide range of observations. Building on the initial comparisons between the two simulations made by Gabrielpillai et al. (2022), we modify the Santa Cruz SAM’s galaxy formation model to better emulate the detailed physical processes found in the IllustrisTNG simulations, with a specific focus on mirroring TNG’s treatment of stellar feedback. To avoid the added complexity and modeling uncertainties associated with AGN feedback, we limit our analysis to lower-mass halos, where stellar-driven processes domi-

nate the baryon cycle. By calibrating the SAM with measurements of galaxy- and halo-scale inflow and outflow rates for a subset of TNG galaxies (Oren et al., 2025, in prep), we aim to emulate the underlying baryon cycle modeled in TNG, resulting in the creation of a new “TNG SAM.” This endeavor serves three primary goals: (1) to simplify and distill TNG’s complex feedback and baryon cycling behaviors into a more interpretable semi-analytic framework, (2) to facilitate direct comparisons between SAMs and hydrodynamical simulations, clarifying their respective strengths and limitations, and (3) to ultimately extend the predictive power of SAMs to volumes beyond the reach of hydrodynamical simulations.

This paper is organized as follows: In Section 2, we provide an overview of the IllustrisTNG hydrodynamical simulations and the Santa Cruz SAM, while Section 3 details the methods used to compare their respective outputs. Section 4 outlines the modifications made to the SC SAM’s galaxy formation framework that underpin the development of the new “TNG SAM.” The results of these modifications, namely the performance of the TNG SAM in reproducing key global and flow properties of the galaxies modeled in the TNG Universe, are presented in Section 5. In Section 6, we analyze how each modification contributes to the TNG SAM’s success, and also discuss the model’s current limitations and potential areas for further refinement. Finally, Section 7 provides a concise summary of our results and their implications for developing the next generation of SAMs.

## 2 MODEL DESCRIPTIONS

In this section, we describe the two models used: the IllustrisTNG hydrodynamical simulations and the Santa-Cruz semi-analytic model. We summarize their key parameters in Table 1 and outline the methods used to compare the outputs of the Santa Cruz SAM with those of TNG.

### 2.1 The IllustrisTNG Simulations

The IllustrisTNG simulations are a suite of gravitomagnetohydrodynamical simulations that model the physical processes governing the formation and evolution of galaxies across different cosmological volumes (Springel et al. 2018; Weinberger et al. 2017; Pillepich et al. 2018b; Nelson et al. 2018). The suite includes nine simulations, with box sizes of 50, 100, and 300 Mpc<sup>3</sup>, each run at three different resolutions (1 being the highest and 3 the lowest; Marinacci et al. 2018; Naiman et al. 2018; Springel et al. 2018; Nelson et al. 2019; Pillepich et al. 2019). Each simulation has a companion dark matter-only (DMO) and full-physics (FP) run. The cosmological parameters used in each simulation are taken from Planck Collaboration et al. (2016), with matter density  $\Omega_{M,0} = 0.3089$ , baryon density  $\Omega_{b,0} = 0.0486$ , dark energy density  $\Omega_{\Lambda,0} = 0.6911$ , Hubble constant  $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , power spectrum normalization factor  $\sigma_8 = 0.8159$ , and spectral index  $n_s = 0.9667$ .

TNG builds upon the original Illustris model (Vogelsberger et al. 2013; Torrey et al. 2014), which uses the AREPO code (Springel 2010) to solve coupled equations of self-gravity and magnetohydrodynamics (Pakmor et al. 2011; Pakmor & Springel 2013). To achieve improved consistency with observations, TNG refined the original Illustris models for active galactic nuclei (AGN) feedback, galactic winds, and magnetic fields (Weinberger et al. 2017; Pillepich et al. 2018a).

Given that this paper focuses on low mass halos  $M_{\text{halo}} < 10^{12} M_{\odot}$ , TNG’s treatment of star formation and stellar feedback, particularly

the treatment of supernova-driven galactic winds (Pillepich et al. 2018a), is most relevant for our study. In both the TNG and Illustris models, stars form in dense gas regions that meet a threshold for star formation, following a subgrid implementation of the Kennicutt-Schmidt law (Kennicutt 1998). Specifically, in the TNG model, gas above a critical density enters a two-phase ISM regime, where cold, star-forming clouds coexist with a hot, ionized medium. In this regime, the local star formation rate density scales as  $\dot{\rho}_* \propto \rho^{1.5}$ , where  $\rho$  is the cold gas volume density (Springel & Hernquist 2003; Diemer et al. 2018).

Building on the original Illustris simulation (Vogelsberger et al. 2013), TNG models supernova-driven galactic winds that are launched isotropically from star-forming gas cells in the dense ISM. The winds are modeled as collisionless particles, with the kinetic energy release rate determined by a mass-loading function that depends on the gas-phase metallicity and the dark matter velocity dispersion,  $\sigma_{DM}$ . This velocity dispersion is calculated using a weighted kernel over the nearest 64 DM particles, and the wind velocity is further modulated by  $\sigma_{DM}$  and redshift. These outflows are powerful enough to eject mass from the ISM (Nelson et al. 2019) and, as we discuss in Section 4, can push gas entirely out of the halo.

Several physical processes in TNG—such as star formation, stellar and AGN feedback, black hole seeding, and accretion—are implemented using sub-grid recipes calibrated to align with various observations, similar to the methodology used in SAMs. However, the specific observations and the precision of calibration differ. For TNG, calibration targets include the cosmic star formation rate density as a function of redshift, stellar mass functions, the relationship between black hole mass and stellar mass, the hot gas fraction in galaxy clusters, and the galaxy stellar mass versus radius relation (Pillepich et al. 2018a). Additionally, the distribution of galaxy optical colors in stellar mass bins was used to refine AGN feedback parameters, as detailed in Nelson et al. (2018). While these calibration choices overlap with those used for the SC SAM, described in the section below, there are differences in the specific observations and calibration techniques used.

In this paper, we use both the hydrodynamical (full-physics) and dark matter only (N-body) outputs of the intermediate-sized box ( $L_{\text{box}} = 75 \text{ Mpc h}^{-1}$ ) at its highest resolution, TNG100-1.

### 2.2 The Santa Cruz semi-analytic model

The Santa Cruz SAM for galaxy formation, first introduced in Somerville & Primack (1999), and subsequently updated in Somerville et al. (2008a, 2012), Porter et al. (2014), Popping et al. (2014), and Somerville et al. (2015), traces the evolution of baryons within galaxies by partitioning them into four distinct reservoirs: the cold gas in the galaxy disk, the hot gas associated with the dark matter halo, the ejected gas reservoir, and the stellar component, the latter of which is further divided into disk and bulge populations. To model the mass and energy transfer between these reservoirs, the Santa Cruz SAM incorporates physically and observationally motivated prescriptions for a variety of baryonic processes. These include gas cooling, star formation, stellar feedback, gas recycling, chemical enrichment, and the growth of supermassive black holes and their associated feedback. In this work, we use the latest version of the Santa Cruz SAM published in Somerville et al. (2015) and recently used in Gabrielpillai et al. (2022) and Yung et al. (2023). Since we limit our focus to central galaxies with  $M_{\text{halo}} < 10^{12} M_{\odot}$ , we turn all AGN feedback “off” in the SAM. Below, we briefly summarize the key processes modeled in the SAM, and refer the reader to Somerville et al. (2008a) and Somerville et al. (2015) for

	TNG100 Pillepich et al. (2018a); Springel & Hernquist (2003)	SC SAM Somerville et al. (2008a, 2015)	TNG SAM this paper
<i>Box Size</i>	$75 h^{-1} \text{ Mpc}$	$75 h^{-1} \text{ Mpc}$	$75 h^{-1} \text{ Mpc}$
<i>Particle Resolution</i>	$m_{\text{DM}} = 5.06 \times 10^6 M_{\odot}/h$	$\text{mass}_{\text{min}} = 8.85 \times 10^8 M_{\odot}$	$\text{mass}_{\text{min}} = 8.85 \times 10^8 M_{\odot}$
<i>Halo Gas Accretion</i>	...	NFW	$f_{\text{in}}^{\text{CGM}} \propto f(M_{\text{halo}}, z)$
<i>Hot Halo Gas Re-incorporation</i>	...	$\chi_{\text{re-infall}} = 0.1$	bundled into $f_{\text{in}}^{\text{CGM}}$
<i>Hot Halo Gas Cooling</i>	...	$r_{\text{cool}} < r_{\text{vir}}$ : cold mode accretion $r_{\text{cool}} > r_{\text{vir}}$ : hot mode accretion	$t_{\text{cool}} \propto f(M_{\text{halo}}, z)$
<i>Star Formation</i>	Kennicutt-Schmidt; gas above density threshold forms stars; $\Sigma \text{SFR} \propto \rho^{1.5}$ ; $\tau_{*,0} = 2.1 \text{ Gyr}$ ; Chabrier IMF	multi-phase gas partitioning into neutral, ionized, and molecular components (Gnedin & Kravtsov 2011); H <sub>2</sub> -based star formation recipe (Bigiel et al. 2008); $\tau_{*,0} = 1.0 \text{ Gyr}$ ; Chabrier IMF	$\tau_* \propto f(M_{\text{halo}}, z)$ $\dot{m}_* = M_{\text{cold}}/\tau_*(M_{\text{halo}}, z)$ ; Chabrier IMF
<i>Chemical Enrichment</i>	delayed recycling; metals produced by AGB stars ( $1-8 M_{\odot}$ ), SNII ( $8-100 M_{\odot}$ ), and SNIa; yields for each source determined from rescaled yield tables; enrichment events discretized when mass fraction $> 0.0001$ or star age $< 100 \text{ Myr}$	instantaneous recycling; fixed yield $y = 1.2$ ; $f_{\text{recycle}} = 0.43$	instantaneous recycling; fixed yield $y = 2.0$ ; $f_{\text{recycle}} = 0.43$
<i>Stellar Feedback</i>	originates from star-forming gas particles; Injection velocity $\propto [350 \text{ km s}^{-1}, f(\sigma_{\text{DM}}, H(z))]$ ; $\eta_{\text{ISM}} \propto f(v_w, e_w)$	originates from scale of galaxy; SNe ejection efficiency $\epsilon_{\text{SN}} = 1.7$ , ejection scale $V_{\text{eject}} = 110 \text{ km/s}$ ; gas deposited in ejected reservoir	originates from scale of galaxy; $\eta_{\text{ISM}} \propto f(v_w, e_w, M_{\text{halo}}, z)$ ; gas deposited in hot halo; leftover SNe energy drives halo outflows (Henriques et al. 2015); $\eta_{\text{halo}} \propto f(\eta_{\text{ISM}}, M_{\text{halo}}, z)$ ; gas deposited in ejected reservoir
<i>Metal Circulation</i>	metal loading of wind particles: $\gamma_w = 0.4$	metal flows carry the same metallicity as the source reservoir	metal flows carry metallicity $\propto f(M_{\text{halo}}, z)$ of the source reservoir
<i>Calibrations</i>	cosmic SFR density as a function of redshift; galaxy stellar mass vs radius function at $z = 0$ ; stellar-to-halo mass relation, BH mass versus stellar mass relation; hot gas fraction in galaxy clusters	stellar-to-halo mass ratio; stellar mass function; stellar mass–metallicity relation; cold gas fraction vs stellar mass relation for disc-dominated galaxies; black hole mass versus bulge mass relation	<i>TNG100</i> : $f_{\text{in,CGM}}$ ; $t_{\text{cool}}$ ; $\tau_*$ ; $\eta_{\text{ISM}}$ ; $\eta_{\text{halo}}$ ; $\zeta^{\text{in/out}}_{\text{ISM}}$ ; $\zeta^{\text{in/out}}_{\text{halo}}$

**Table 1.** Comparison of the different approaches to modeling galaxy formation taken by TNG100, the Santa Cruz SAM, and the newly developed TNG SAM. Key aspects include simulation box size, resolution, halo gas accretion and cooling, star formation prescriptions, chemical enrichment, stellar feedback, metal circulation, and calibration strategies.

details.

### 2.2.1 Gas Cooling and Accretion

For any given halo, the SAM estimates the dark matter accretion rate,  $\dot{m}_{\text{DM}}$ , for each halo by using the virial mass provided by the halo merger tree and calculating its rate of change between consecutive snapshots. Prior to reionization, the SAM assumes that gas accre-

tion  $\dot{m}_{\text{acc}}$  closely follows the accretion of dark matter, scaled by the universal baryon fraction  $f_b$  as:

$$\dot{m}_{\text{acc}} = f_b \dot{m}_{\text{DM}}, \quad (1)$$

where  $f_b = 0.1573$  (Planck Collaboration et al. 2016).

Following reionization, the photoionizing background suppresses gas accretion into halos, with the fraction of baryons able to collapse determined by  $f_{\text{coll}}$ , a function of halo mass and redshift (Okamoto et al. 2008). The SAM also accounts for the recycling of gas pre-

viously expelled by stellar feedback, described in detail in Section 2.2.2. The rate of gas re-accretion is given by:

$$\dot{M}_{\text{CGM,in,recycled}} = \chi_{\text{re-infall}} \left( \frac{M_{\text{ejected}}}{t_{\text{dyn}}} \right), \quad (2)$$

where  $M_{\text{ejected}}$  represents the total mass of gas in the ejected reservoir, and  $\chi_{\text{re-infall}}$  specifies the fraction of this gas that can return to the hot halo at each time step. Typically, in the previously published SC SAMs,  $\chi_{\text{re-infall}} = 0.1$ , meaning that the ejected gas is recycled over  $\sim 10$  dynamical times. The dynamical time is defined as  $t_{\text{dyn}} \equiv R_{\text{vir}}/V_{\text{vir}}$ , where  $V_{\text{vir}}$  is the circular velocity of the halo at the virial radius. Additionally, when a halo becomes a subhalo, the SAM assumes its CGM is immediately transferred to the host halo, adding to the host's hot gas reservoir.

Following the framework of White & Frenk (1991), the SAM assumes that the CGM is uniformly at the virial temperature of the halo at each time step. The radiative cooling time, which determines how quickly the gas loses thermal energy via radiation, is calculated using the cooling function  $\Lambda(T_{\text{vir}}, Z_h)$  (Sutherland & Dopita 1993), where  $T_{\text{vir}}$  is the virial temperature and  $Z_h$  is the metallicity of the hot halo gas. The gas density profile is modeled as a singular isothermal sphere, and the cooled gas mass within the radius  $r_{\text{cool}}$ —the radius within which all gas can cool within the cooling time  $t_{\text{cool}}$ —determines the ISM mass accretion rate.

If  $r_{\text{cool}}$  is smaller than the virial radius  $r_{\text{vir}}$ , the SAM applies a standard cooling flow model, with  $t_{\text{cool}}$  calculated following Somerville et al. (2008a) under the assumption of an isothermal and isobaric gas density profile. When  $r_{\text{cool}}$  exceeds  $r_{\text{vir}}$ , the cooling rate matches the gas accretion rate into the halo, representing “cold/fast/filamentary” mode accretion. In these cases, the SAM ignores radiative cooling predictions and sets the ISM accretion rate equal to the halo gas accretion rate. Variations in this cooling model, such as adjustments to the cooling time definition or changes to the gas density profile, can impact the ISM accretion rate by a factor of 2–3 (Somerville et al. 2008b).

## 2.2.2 Star Formation and Stellar Feedback

In the most recently used Santa Cruz SAM, gas accreted into the ISM is partitioned into H I, H<sub>2</sub>, H II, and metals, with their respective mass surface densities tracked in radial disk annuli (Popping et al. 2014; Somerville et al. 2015). The default prescription for the star formation rate (SFR) surface density is based on the molecular hydrogen gas phase alone, accounting for a higher conversion efficiency above a critical  $H_2$  surface density (Bigiel et al. 2008; Narayanan et al. 2012):

$$\dot{\Sigma}_{\text{SFR}} = A_{\text{SF}} \left( \frac{\Sigma_{H_2}}{10 M_{\odot} \text{pc}^{-2}} \right) \left( 1 + \frac{\Sigma_{H_2}}{\Sigma_{H_2, \text{crit}}} \right)^{N_{\text{SF}}}, \quad (3)$$

where  $A_{\text{SF}}$ ,  $N_{\text{SF}}$  and  $\Sigma_{H_2, \text{crit}}$  are free parameters. The molecular gas surface density  $\Sigma_{H_2}$  is estimated using the metallicity-dependent partitioning scheme of Gnedin & Kravtsov (2011).

In the SAM, stellar feedback ejects cold gas from the ISM, with the ejection rate modeled as a power law:

$$\dot{m}_{\text{eject}} = \epsilon_{\text{SN}} \left( \frac{200 \text{ km/s}}{V_{\text{disk}}} \right)^{\alpha_{\text{rh}}} \dot{m}_{*}, \quad (4)$$

where the free parameters  $\epsilon_{\text{SN}}$  and  $\alpha_{\text{rh}}$  modulate the efficiency and dependence on the disk's circular velocity  $V_{\text{disk}}$ , and  $\dot{m}_{*}$  is the star formation rate.  $V_{\text{disk}}$  is approximated as the circular velocity of the uncontracted halo at twice the Navarro-Frenk-White (NFW) scale radius  $r_s$  (Navarro et al. 1996). All ejected gas is either expelled from

the halo entirely or deposited in the “ejected” reservoir, depending on whether the halo's circular velocity falls below a critical threshold. A fraction of this gas is allowed to return to the hot halo on dynamical timescales, as described in Section 2.2.1.

## 2.2.3 Metal Production

The Santa Cruz SAM models metal production using the instantaneous recycling approximation, a commonly employed approach in SAMs. When a mass of stars,  $dm_{*}$ , forms, it generates a corresponding mass of metals,  $dM_Z = y \cdot dm_{*}$ , which are instantly mixed with the cold gas in the disk. The yield,  $y$ , is typically calibrated to match observational data, such as the normalization of the stellar metallicity–mass relation from Gallazzi et al. 2005, as shown in Somerville et al. 2015.

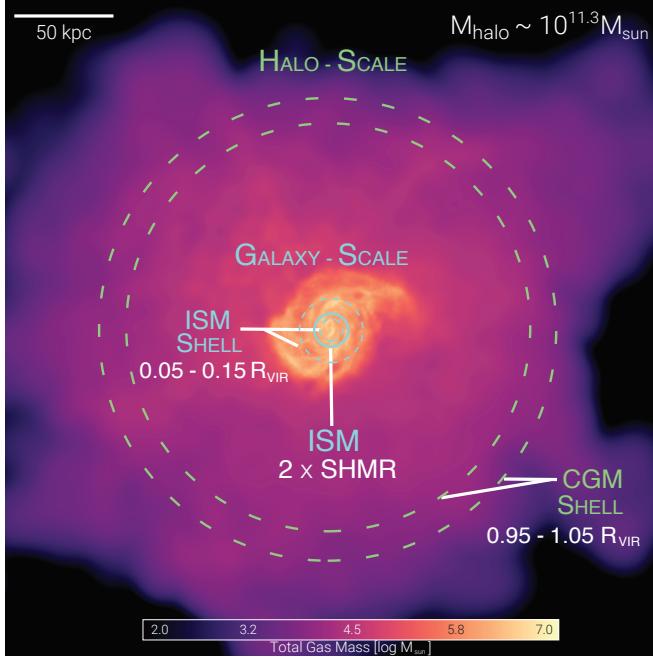
The metallicity of the cold gas,  $Z_{\text{ISM}}$ , evolves over time, with newly formed stars inheriting this metallicity. Supernova feedback drives both gas and metals out of the disk, transferring them to the hot halo or ejecting them into an external reservoir at rates proportional to the mass of outflowing gas. These ejected metals are not lost permanently; they can be re-accreted into the halo, contributing to ongoing chemical enrichment over time.

## 2.2.4 Observational Calibration

The SC SAM is calibrated by adjusting free parameters to match key observations at  $z \sim 0$ , such as the stellar-to-halo mass relation, stellar mass function, stellar mass-metallicity relation, cold gas fraction versus stellar mass for disk-dominated galaxies, and the black hole mass vs. bulge mass relation. Observational constraints for these calibrations are sourced from Rodríguez-Puebla et al. (2017), Bernardi et al. (2013), Gallazzi et al. (2005), Peebles et al. (2014), Calette et al. (2018), and McConnell & Ma (2013). Additional cross-checks involve the cold gas phase mass-metallicity relation and the  $H_2$  mass function, using constraints from Obreschkow et al. (2009), Keres et al. (2003), Anderson et al. (2014), Zahid et al. (2013), and Boselli et al. (2014). The calibration considers the sensitivity of gas fraction, stellar metallicity, and star formation efficiency, fine-tuning the parameters to balance these observational constraints while adjusting for the effects of AGN and SNe feedback to match both high-mass and low-mass galaxy populations. For a more detailed description of the observations used for calibration, refer to (Gabrielpillai et al. 2022) and Appendix B in Yung et al. (2023).

## 2.2.5 Modifications to the Santa Cruz SAM

The version of the Santa Cruz SAM presented in this work revises how the SAM computes the rate of gas condensation from the CGM into the ISM. When  $r_{\text{cool}} > r_{\text{vir}}$ , the original Santa Cruz SAM disregards the radiative cooling prediction and equates the ISM accretion rate to the halo gas accretion rate. This approach, however, results in unrealistically low hot halo gas masses for dwarf galaxies when compared to predictions from hydrodynamical simulations (Pandya et al. 2020). To rectify this issue, we follow Pandya (2021) and adopt a modified approach inspired by Guo et al. (2011): when  $r_{\text{cool}} > r_{\text{vir}}$ , the rate of cooling is limited solely by the freefall time of the hot halo



**Figure 1.** A sample  $M_{\text{halo}} \sim 10^{11.3} M_{\odot}$  central subhalo (ID: 609911) from TNG100 is shown to illustrate how TNG’s outputs are compared to the SAM. The map is color-coded by total gas mass, with brighter regions indicating higher mass concentrations. In TNG, we define the interstellar medium (ISM) as all material within twice the stellar half mass radius (solid blue circle), and the circumgalactic medium as all material within the central galaxy’s group (not shown). To analyze baryon flows on galaxy and halo scales, we measure inflow and outflow rates across the dashed shells at  $0.05\text{--}0.15r_{\text{vir}}$  (galaxy scale) and  $0.95\text{--}1.05r_{\text{vir}}$  (halo scale).

gas, such that:

$$\dot{M}_{\text{ISM}}^{\text{in}} = \begin{cases} \frac{M_{\text{CGM}}}{t_{\text{dyn}}} \frac{r_{\text{cool}}}{r_{\text{vir}}} & \text{when } r_{\text{cool}} < r_{\text{vir}} \\ \frac{M_{\text{CGM}}}{t_{\text{dyn}}} & \text{when } r_{\text{cool}} \geq r_{\text{vir}}. \end{cases} \quad (5)$$

While this modification no longer predicts extremely low CGM masses in dwarf galaxies (See Figure A1), it may instead overestimate CGM masses by underestimating the total gas inflow into the galaxy. When  $r_{\text{cool}} > r_{\text{vir}}$ , our new version of the model assumes a hot accretion mode from the entire mass within  $r_{\text{vir}}$ . However, this approach does not take into account the contribution from cold accretion in streams or filaments from the IGM. Thus, our model represents an upper bound on the actual accretion rate.

### 3 COMPARING THE OUTPUT BETWEEN TNG AND THE SC SAM

#### 3.1 Matching Galaxies Between TNG FP and SC SAM Run on TNG-DMO

The precise identification of halos and construction of merger trees forms the backbone of all SAMs. Thus, to facilitate a direct comparison between the Santa Cruz SAM and TNG, we extract merger trees from a dark matter only simulation (TNG-DMO) run with the same initial conditions as the Full Physics TNG simulation (FP TNG). The SC SAM employs the ROCKSTAR halo finder and CONSISTENT TREES, a gravitationally consistent merger tree algorithm developed by Behroozi et al. (2012). TNG, however, utilizes a different halo finder based on the friends-of-friends (FoF) algorithm (Davis et al. 1985) to identify “groups,” and the SUBFIND algorithm (Springel

et al. 2001) to identify substructure. These merger trees are not compatible with the Santa Cruz SAM.

To create bijective matches between the two sets of halo catalogs, Gabrielpillai et al. (2022) ran the ROCKSTAR halo finder and CONSISTENT TREES on TNG100-1-DM. They then matched halos identified by ROCKSTAR and SUBFIND catalogs with SubLink, a software tool developed by Rodriguez-Gomez et al. (2015) that uses a merit function for matching subhalos. This produced successful bijective matches between the ROCKSTAR and FoF/SUBFIND halos for central galaxies, achieving a 99% match rate at  $z = 0$ . The fraction of bijective matches for non-central subhalos, however, is often substantially lower, with a match rate from 50–70%. As a result, we limit our analysis to central galaxies only.

Although 99% of central galaxies have bijective matches, a direct comparison between the halo mass of the central galaxies between the two catalogs reveals slight offsets up to 20% for low mass halos (see Figure 2 in Gabrielpillai et al. 2022). These discrepancies are likely primarily influenced by baryonic physics rather than differences in halo-finding algorithms. For a more detailed description of the halo finder, merger tree algorithms, and method used to create these bijective matches, we refer the reader to Section 3 of Gabrielpillai et al. (2022).

#### 3.2 Selecting Comparable Galaxy and Halo Properties

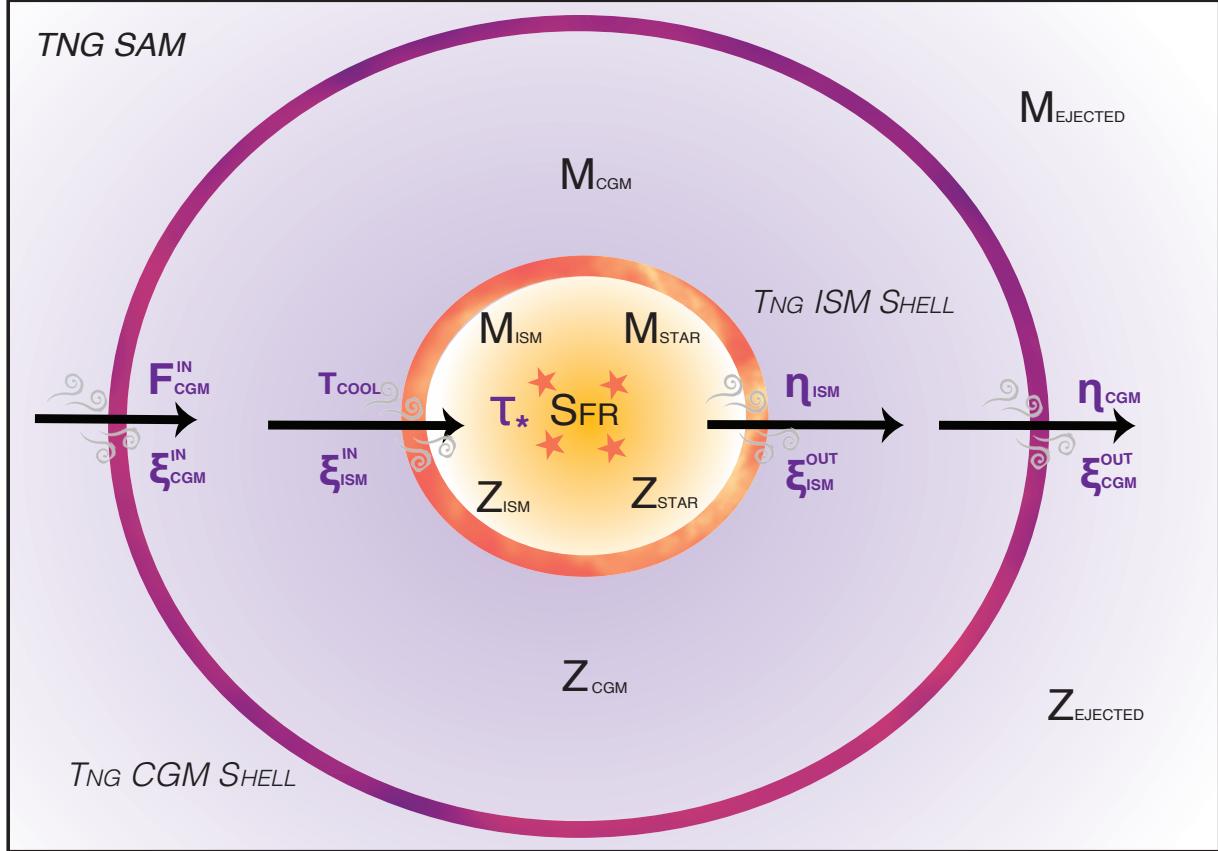
We define two physical scales for comparing TNG and the SAM:

- (i) galaxy, and
- (ii) halo,

as illustrated in Figure 1. Properties at the “galaxy scale” describe the stellar mass ( $M_{\star}$ ), cold gas mass ( $M_{\text{ISM}}$ ), their respective metallicities ( $Z_{\star}$  and  $Z_{\text{ISM}}$ ), and the star formation rate (SFR) of the central galaxy within the halo. Properties at the “halo scale” include the virial mass ( $M_{\text{halo}}$ ), the mass of the hot halo ( $M_{\text{CGM}}$ ), the metallicity of the gas in the hot halo ( $Z_{\text{CGM}}$ ), and the overall baryon fraction ( $f_{\text{baryon}}$ ). We define the baryon fraction as the sum of  $M_{\text{ISM}}$ ,  $M_{\star}$ , and  $M_{\text{CGM}}$  divided by the total halo mass. Virial masses and radii are computed using the spherical overdensity definition of Bryan & Norman (1998), where  $M_{\text{halo}}$  is the total mass enclosed within  $r_{\text{vir}}$  such that the mean density of dark matter haloes is  $\Delta_c$  times the critical density at a given redshift.

The SAM naturally segments and outputs all quantities defined above. For TNG, we measure galaxy-scale properties within twice the stellar half mass radius, and halo-scale properties within the entire host group, excluding material within twice the stellar half radius of all group members. The corresponding field names used in both the SAM and TNG outputs are listed in Table 1.

To enable reliable comparison between TNG and the SAM, we restrict the bulk of our analysis to well-resolved TNG halos. Following the general guidelines of Gabrielpillai et al. (2022), we require halos to contain at least 100 particles to be considered well-resolved. We also require that the root halo be at least 100 times more massive than the smallest resolvable progenitor to accurately capture its merger history. For TNG100-1-Dark, with a particle mass resolution of  $8.9 \times 10^6 M_{\odot}$ , this results in a conservative minimum root halo mass of  $8.9 \times 10^8 M_{\odot}$ . Since the SAM’s resolution is determined by the N-body simulation used to extract merger trees, we adopt this minimum root halo mass as our resolution threshold. For the hydrodynamical TNG100-1 simulation, each gas or star particle has a mass of  $1.4 \times 10^6 M_{\odot}$ . We therefore restrict our sample to galaxies with at least  $\sim 100$  stellar particles  $M_{\star} \geq 1.4 \times 10^8 M_{\odot}$ , and 100 gas



**Figure 2.** Illustration of the TNG SAM, calibrated to reproduce the baryon cycle of  $\sim 400$  central galaxies in the TNG100 simulation. The diagram shows the main baryonic reservoirs—cold gas ( $M_{\text{ISM}}$ ), stars ( $M_{\text{star}}$ ), hot halo gas ( $M_{\text{CGM}}$ ), and ejected gas ( $M_{\text{ejected}}$ )—and the flows of gas and metals between them. Arrows represent baryon flow directions, with governing parameters labeled in purple (e.g.,  $t_{\text{cool}}$  regulates cooling from the hot halo to the ISM, while  $\eta_{\text{ISM}}$  and  $\eta_{\text{halo}}$  control gas outflows). Metallicity enrichment is tracked for each reservoir as  $Z_{\text{ISM}}$ ,  $Z_{\text{star}}$ ,  $Z_{\text{CGM}}$ , and  $Z_{\text{ejected}}$ . Each parameter was calibrated using inflow and outflow rates measured at boundaries defined in Figure 1.

particles  $M_{\text{gas}} \geq 1.4 \times 10^8 M_{\odot}$ . The gas particle requirement naturally sets a lower limit on the SFR. Following, Donnari et al. (2019), we adopt a minimum SFR of  $1.0 \times 10^{-3} M_{\odot} \text{ yr}^{-1}$ . Applying these criteria yields a final sample of 19,650 galaxies with halo masses ranging from  $\sim 4 \times 10^{10} - 10^{12} M_{\odot}$  at  $z = 0$ .

### 3.2.1 Flow Cycle Measurements in TNG

To establish a common language for describing the baryon cycle between TNG and the SC SAM, we draw on the analysis recently carried out by Oren et al. 2025 (hereafter YO25). YO25 extracted measurements of gas mass, metal, and energy inflows and outflows at ISM and halo scales for a randomly selected subset of  $\sim 9,522$  central galaxies in TNG100-1 across 10 redshifts ranging from  $z = 0$  to  $z = 10$ . At each redshift, YO25 selected halos in bins of 0.3 dex in virial mass, spanning  $M_{\text{vir}} = 10^{10} M_{\odot}$  to the most massive halo at that redshift.

Using an Eulerian approach, YO25 calculated instantaneous gas flows by defining volumetric shells of thickness  $\Delta r \sim 0.05 r_{\text{vir}}$  from  $0.05r_{\text{vir}}$  to  $2r_{\text{vir}}$ . The mass and velocity of the gas particles crossing these boundaries were tracked between consecutive simulation outputs, directly measuring the inflow and outflow rates. For further details on the methodology used to derive these measurements, we refer the reader to YO25.

To minimize the impact of resolution effects on our results, we divided the YO25 “flow sample” into “resolved” and “unresolved” categories using the same resolution criteria applied to the full TNG100

sample. While these resolution criteria were originally developed to assess the reliability of global galaxy properties (such as  $M_{\text{star}}$  and SFR), we apply them here as a practical filter to exclude galaxies that are likely to be poorly resolved overall. The resolved subsample comprises  $\sim 400$  galaxies per redshift, with the remaining galaxies classified as unresolved. For each subsample, we define the halo boundary from  $0.95 - 1.05 r_{\text{vir}}$ , and the galaxy boundary from  $0.05 - 0.15 r_{\text{vir}}$ , both illustrated as dashed circles in Figure 1. We volumetrically summed the inflow and outflow rates within each boundary to obtain the gas and metal flow rates shown in Sections 4 and 5.

Both the resolved and unresolved samples are used in this work to ensure a comprehensive analysis across all halo masses. While we primarily rely on the resolved sample for calibrating the TNG SAM, we also make use of the unresolved halos to inform our extrapolation to lower masses, where many key phases of early galaxy evolution occur and resolution limitations become more severe (see Section ??).

## 4 TRANSLATING TNG’S DETAILED PHYSICS INTO THE SC SAM’S GALAXY FORMATION MODEL

To translate the complex physics of a hydrodynamical simulation like TNG into the simplified framework of the SC SAM, we focus on modeling the baryon cycle—how gas and metals circulate between the galaxy and its surrounding halo—over time, as summarized in Figure 2. We achieve this using the YO25 subsample of TNG galaxies described in Section 3.2.1, which provides inflow and outflow rates of gas and metals for each galaxy in their sample. Treating this

subset as representative of the entire TNG central galaxy population, we derived physical parameters that govern the evolution of these baryon flows, namely halo (re)-accretion efficiency, cooling time, star formation, and mass and metal enrichment. For the star formation efficiency, we use the full sample of  $\sim 20,000$  TNG galaxies. These parameters, shown in Figure 3 as functions of halo mass and redshift, form the foundation for calibrating the SC SAM to capture the baryon cycling processes found in TNG.

To implement the scaling relations in the SAM, we adopt a two-pronged approach that leverages both the “resolved” and “unresolved” populations of the [YO25](#) sample. For halos within the well-resolved mass range of  $10.5 < \log M_{\text{halo}} < 12$ , we directly apply the median scaling relations derived from the resolved sample, as shown by the solid lines in Figure 3. When halos fall below this mass range, we extrapolate the scaling relations to lower-mass halos ( $\log M_{\text{halo}} < 10.5$ ) by analyzing trends in the complete [YO25](#) sample, which includes both resolved and unresolved populations down to  $\log M_{\text{halo}} \sim 10$ , and by incorporating the full TNG100 sample, which resolves halos down to  $\log M_{\text{halo}} \sim 8$ . Specifically, we analyze the behavior of halos in the range  $10 < \log M_{\text{halo}} < 10.5$  (and extend this analysis to lower masses when using the full TNG100 sample) and use the resulting relations (shown as dashed lines in Figure 3) to model all halos with  $\log M_{\text{halo}} < 10.5$ .

This approach allows the SAM to capture baryon cycling in unresolved halos, even though these systems are not directly included in the [YO25](#) flow sample. However, extrapolating into this low-mass regime naturally introduces larger uncertainties in the predicted evolution of these halos. We discuss the implications and limitations of this method in greater detail in Section 6.2.

The newly calibrated SAM applies these scaling relations across halo mass and time, performing linear interpolation to determine the relations at intermediary halo masses and redshifts. While the physical processes governing gas flows are not expected to fundamentally change with redshift, TNG shows that the behavior clearly evolves over time, likely driven by underlying physical mechanisms that remain poorly understood. We bundle this ignorance into our interpolation approach, providing a pragmatic approximation of TNG’s redshift-dependent behavior. We also provide analytic functions of halo mass and redshift that capture each relation within 30% accuracy in Table A1. This approach, instead of simply directly inputting the median flow rates from TNG, enables the SAM to compute flow rates in a self-consistent manner according to the provided physical parameters, creating a more physically motivated and flexible SAM. We refer to this newly calibrated model as the ‘TNG SAM.’

## 4.1 Halo Accretion and Cooling

### 4.1.1 IGM - CGM Accretion and Recycling

In the traditional SAM framework, the rate at which gas accretes onto the CGM is tightly coupled to the dark matter accretion rate, as detailed in Section 2.2.1. In the TNG SAM, we introduce the parameter  $f_{\text{in,CGM}}$  to regulate total gas inflow into the CGM:

$$f_{\text{in,CGM}} \equiv \dot{M}_{\text{in,CGM}} / (f_b \dot{M}_{\text{halo}}), \quad (6)$$

where  $f_{\text{in,CGM}}$  measures the efficiency of gas accretion relative to the expected baryonic inflow, without distinguishing between pristine and recycled gas. Here,  $\dot{M}_{\text{in,CGM}}$  is the gas inflow rate into the CGM,  $f_b$  is the cosmic baryon fraction, and  $\dot{M}_{\text{halo}}$  is the net mass accretion rate of the halo in TNG. We use the *net* accretion rate, rather than the total inflow rate, because it aligns with how the SAM calculates  $\dot{M}_{\text{halo}}$ . Specifically, the SAM estimates halo growth by finite differencing

the total halo mass between consecutive timesteps, capturing the net change in halo mass from both smooth accretion and mergers.

When  $f_{\text{in,CGM}} = 1$ , the inflow rate matches the expected baryon-to-dark matter ratio. Values below 1 indicate reduced inflow, while values above 1 suggest that gas recycling contributes to enhanced inflow. Figure 3, panel *i.*, shows that in TNG,  $f_{\text{in,CGM}}$  fluctuates at or above unity across halo mass and redshift, indicating that the gas inflow generally meets or exceeds the expected baryonic fraction. We further explore the implications of this behavior in Section 6.1.1. We note that the halos in our analysis are well above the mass where we expect the metagalactic photoionizing background to reduce baryonic accretion into halos.

### 4.1.2 CGM-ISM Cooling

At the scale of the galaxy, the SC SAM, like most conventional SAMs, relies on the cooling radius  $r_{\text{cool}}$  to determine the rate at which gas cools into the ISM, as described in Section 2.2.1. In the traditional model, the cooling time  $t_{\text{cool}}$ —the time required for gas to radiate away its thermal energy—is estimated using

$$t_{\text{cool}}(r) \equiv \frac{3}{2} \frac{(n_e + n_p)kT(r)}{n_e n_H \Lambda(kT(r), Z)}, \quad (7)$$

where  $n_e$  and  $n_p$  are the electron and proton densities,  $T(r)$  is the gas temperature, and  $\Lambda$  is the cooling function of [Sutherland & Dopita \(1993\)](#).

In hydrodynamical simulations, this equation is often approximated as the time required to radiate away the thermal energy  $t_{\text{cool}} \sim E_{\text{th}}/\Lambda n^2$ , where  $E_{\text{th}}$  is the thermal energy of the gas (e.g., Equations 21 and 27 in [Pandya et al. 2023](#)). In this framework, cooling follows a “cold” mode ( $r_{\text{cool}} > r_{\text{vir}}$ ) when the gas cools efficiently and falls directly into the galaxy, with the cooling time scaling with the dynamical time. Conversely, in “hot” mode accretion ( $r_{\text{cool}} < r_{\text{vir}}$ ), the gas is shock-heated, leading to significantly longer cooling times.

Since the TNG SAM does not track the thermal energy of the hot halo gas, we approximate the cooling time using

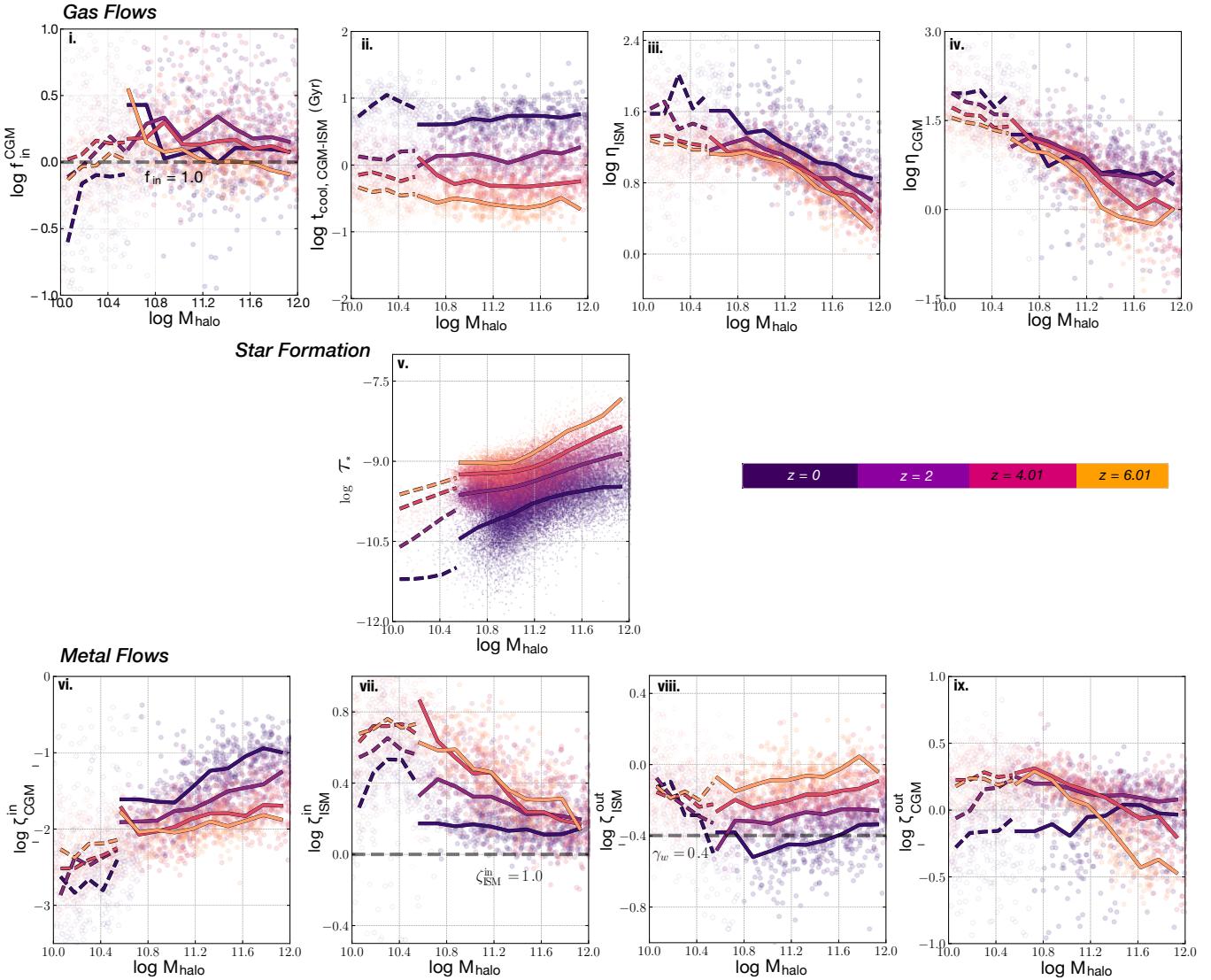
$$t_{\text{cool}} \equiv \frac{M_{\text{CGM}}}{\dot{M}_{\text{cool,CGM-ISM}}}, \quad (8)$$

where  $\dot{M}_{\text{cool,CGM-ISM}}$  describes the rate that gas cools from the hot halo into the ISM. Rather than tracking the time for gas to radiate away its thermal energy, as in the standard definition of  $t_{\text{cool}}$ , this formulation instead measures how quickly gas transitions from the CGM to the ISM.

Figure 3, panel *ii.*, shows the resulting cooling times as a function of halo mass and redshift. We find that the cooling times consistently exceed the dynamical time across all halo masses explored, implying that the “cold” vs. “hot” mode dichotomy does not accurately describe the rate of cooling in TNG, discussed further in Section 6.1.2. In the new TNG SAM, we use the cooling times found in TNG to determine the rate at which gas accretes onto the galaxy by rearranging Equation 8.

## 4.2 Star Formation

In TNG, star formation is treated using a subgrid model. Specifically, TNG adopts the two-phase ISM model of [Springel & Hernquist \(2003\)](#), where an effective equation of state is adopted for the ISM based on a picture in which dense, cold clouds are embedded in a hot tenuous medium, with the two phases in rough pressure equilibrium.



**Figure 3.** Overview of the physical parameters measured from the subset of  $\sim 400$  TNG100 galaxies described in Section 3.2.1, organized by relations regulating gas flows, star formation, and metal flows. Each panel shows a relation plotted as a function of halo mass at multiple redshifts ( $z = 0, 2, 4, 6$ ). Circles represent individual galaxies, colored by redshift, with faded points indicating unresolved systems. Solid and dashed lines show medians for resolved and unresolved galaxies, respectively. The parameters shown are: (i) CGM inflow efficiency ( $f_{\text{in,CGM}}$ ), (ii) CGM - ISM cooling time ( $t_{\text{cool}}$ ), (iii - iv) ISM and CGM outflow mass-loading factors ( $\eta_{\text{ISM}}, \eta_{\text{halo}}$ ), (v) star formation efficiency ( $\tau_*$ ), (vi - vii) CGM/ISM inflow metallicity loadings ( $\zeta_{\text{CGM}}^{\text{in}}, \zeta_{\text{ISM}}^{\text{in}}$ ), and (viii-ix) ISM/CGM outflow metallicity loadings ( $\zeta_{\text{ISM}}^{\text{out}}, \zeta_{\text{CGM}}^{\text{out}}$ ). Analytic functions fitted to these median relations, each accurate to within  $\sim 30\%$ , are implemented directly in the TNG SAM, with coefficients provided in Table A1.

For gas volume densities greater than a critical value  $\rho > \rho_{\text{th}}$ , star formation is assumed to take place at a rate

$$\frac{d\rho_*}{dt} = \rho_c/t_* \quad (9)$$

where  $\rho_c$  is the density of gas in the cold phase, as computed by the subgrid model. The gas consumption time is then given by  $t_* = t_0^*(\rho/\rho_{\text{th}})^{1/2}$ . The parameters  $t_0^*$  and  $\rho_{\text{th}}$  are tuned to reproduce the global observed Kennicutt-Schmidt (KS) relation, given by Kennicutt (1998):

$$\Sigma_{\text{SFR}} = (2.5 \pm 0.7) \times 10^{-4} \left( \frac{\Sigma_{\text{gas}}}{M_{\odot}, \text{pc}^{-2}} \right)^{1.4 \pm 0.15} M_{\odot}, \text{yr}^{-1}, \text{kpc}^{-2}. \quad (10)$$

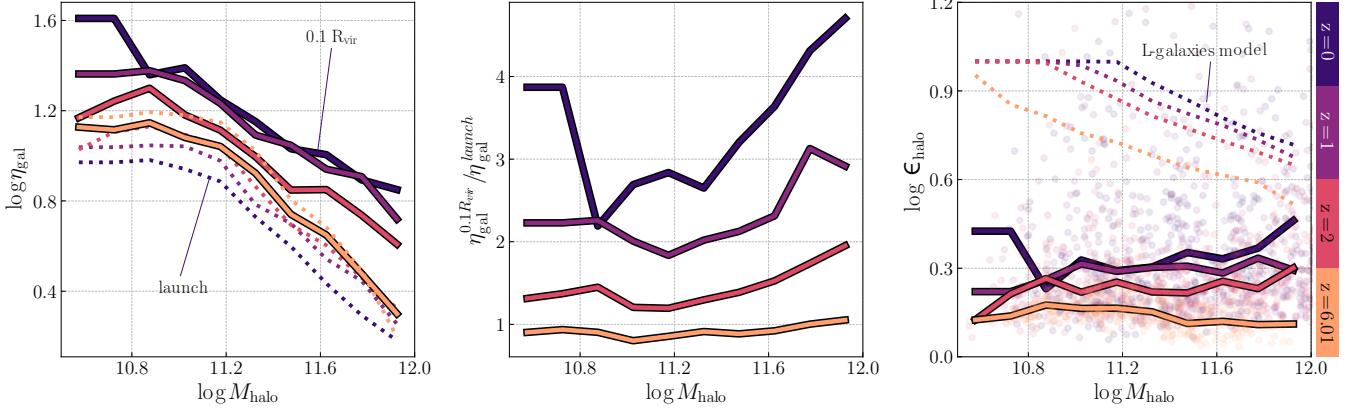
In the TNG SAM, we do not implement the KS relation, even though it is available as an option in the SC SAM's framework. This decision stems from the fact that the SC SAM and TNG predict markedly different galaxy sizes—particularly at higher redshifts

(see Section 6.1.3)—which prevents a consistent use of a surface-density-based star-formation law. Instead, we model star formation directly through the star-formation efficiency  $\tau_*$ , defined as

$$\tau_* = \frac{\dot{m}_*}{M_{\text{cold}}}. \quad (11)$$

We measure  $\tau_*$  as a function of halo mass and redshift using the full TNG100 galaxy sample (see Figure 3 panel v.), and calibrate the star formation rate with:

$$\dot{m}_* = \tau_*(M_{\text{halo}}, z) \cdot M_{\text{cold}}. \quad (12)$$



**Figure 4.** Functions that govern the TNG SAM’s stellar feedback-driven outflows at the galaxy (left panels) and halo (right panel) scales across redshift. The left panel shows that the mass loading factor  $\eta_{\text{ISM}}$  (solid lines), defined as the ratio of the outflow rate to the star formation rate, almost always exceeds the launch value from star-forming gas cells in TNG (dotted lines), suggesting additional mass entrainment as winds reach the galaxy’s radius. To account for this, the TNG SAM scales the launch value by the ratio of galaxy-scale to launch-scale values, which vary with halo mass and redshift, as shown in the middle panel. At the halo scale, the outflow efficiency is captured by  $\epsilon_{\text{halo}}$ , defined analogously as the outflow rate through the halo boundary divided by the star formation rate, and is modeled using the dimensionless parameter  $\epsilon_{\text{halo}}$ . The values of  $\epsilon_{\text{halo}}$  adopted in the L-galaxies SAM (dotted lines) are much higher than those directly derived from TNG (solid lines) using  $\eta_{\text{halo}}$  and  $\eta_{\text{ISM}}$ , which are used as inputs in the TNG SAM.

### 4.3 Stellar Feedback

#### 4.3.1 Galactic Winds and Galaxy-scale Outflows

To emulate TNG’s treatment of stellar feedback, the TNG SAM incorporates the galactic wind model from Pillepich et al. 2018b. Inspired by the results of semi-analytic models, the TNG model launches wind particles with an initial speed that scales with the local dark matter velocity dispersion  $\sigma_{\text{DM}}$ . It also includes a redshift-dependent factor and a minimum wind velocity, such that the wind velocity  $v_w$  scales as:

$$v_w = \max \left[ \kappa_w \sigma_{\text{DM}} \left( \frac{H_0}{H(z)} \right)^{1/3}, v_{w,\min} \right] \quad (13)$$

where  $\kappa_w = 7.4$  and  $v_{w,\min} = 350 \text{ km s}^{-1}$  in the fiducial TNG model. This form ensures a physically plausible scaling of wind velocities with redshift and halo mass, while preventing unphysically large mass loading factors in low-mass halos.

The TNG model further refines wind generation by linking the wind mass loading factor  $\eta_w$  to the specific energy available for wind generation  $e_w$ . This energy is primarily determined by the energy released from Type II supernovae (SNII), with a fraction  $\tau_w$  being thermal. The wind energy itself is a function of the metallicity of the star-forming gas, leading to a metallicity-dependent galactic wind model.

The wind mass loading factor  $\eta_w$ , defined as the ratio of the mass outflow rate  $\dot{m}_{\text{ISM,out}}$  to the star formation rate  $\dot{m}_*$ , quantifies the efficiency of these winds in removing gas from the ISM, and is related to  $v_w$  and the wind energy,  $e_w$ , as:

$$\eta_w = \frac{2}{v_w^2} e_w (1 - \tau_w) \quad (14)$$

where  $e_w$  is given by:

$$e_w = \bar{e}_w \left[ f_{w,Z} + \frac{1 - f_{w,Z}}{1 + (Z/Z_{w,\text{ref}})^{\gamma_{w,Z}}} \right] \times N_{\text{SNII}} E_{\text{SNII}} 10^{51} \text{ erg M}_{\odot}^{-1}. \quad (15)$$

Here,  $\bar{e}_w$  is the average energy per unit stellar mass formed,  $f_{w,Z}$  modulates the metallicity dependence,  $Z$  represents the gas metallicity,  $Z_{w,\text{ref}}$  is a reference metallicity,  $\gamma_{w,Z}$  controls the strength of the

metallicity dependence,  $N_{\text{SNII}}$  denotes the number of Type II supernovae, and  $E_{\text{SNII},51}$  is the energy released per supernova in units of  $10^{51}$  erg.

In the TNG model,  $\eta_w$  is implemented at the scale of the launched wind particle. To implement this model within the TNG SAM, we instead apply  $\eta_w$  at the boundary of the ISM. The left panel of Figure 4 shows that the value of  $\eta_w$  at the scale of the galaxy (solid lines) can differ significantly from the value at launch (dashed lines), with actual wind outflow rates consistently higher than the launch rates at  $z < 2$ , indicating that the winds entrain additional material as they propagate outwards, and lower at  $z \sim 6$ , indicating that the winds are stalling, perhaps due to the higher density and pressure of the CGM. To account for this in the SAM, we model the ratio of  $\eta_w$  at the site of the launched wind particle to  $\eta_w$  at the scale of the galaxy in TNG, shown in the middle panel of Figure 4.

In the TNG SAM, the material ejected from the ISM is always deposited in the CGM. To prevent unrealistically large galaxy-scale mass loading factors, we cap  $\eta_w$  at a maximum value of 100. This upper limit ensures that lower mass halos do not produce outflow efficiencies far beyond those seen in the YO25 sample of TNG galaxies. Moreover, because the TNG model already imposes a ceiling on wind velocities, applying a corresponding constraint on the mass loading factor provides a natural way to maintain consistency between the two models.

#### 4.3.2 Halo-scale Outflows

In TNG, galactic winds can not only expel gas from the ISM but also entrain additional material from the CGM, ejecting it into the IGM. In the SC SAM, however, stellar feedback ejects gas directly from the ISM into an ejected reservoir, bypassing the CGM. The ejected gas only reconnects with the CGM indirectly when a portion of it is re-accreted into the hot halo at later times. In the new TNG SAM, we explicitly account for the CGM-IGM outflow channel with the parameter  $\eta_{\text{halo}}$ , defined as the ratio of the mass outflow rate  $\dot{m}_{\text{CGM,out}}$  to the star formation rate  $\dot{m}_*$ . Complementing the existing  $\eta_w$  at the galaxy scale,  $\eta_{\text{halo}}$  quantifies the efficiency with which stellar feedback ejects gas from the halo.

Few SAMs include a CGM-IGM outflow channel. The TNG SAM model for  $\eta_{\text{halo}}$  is inspired by the halo-scale outflow prescription in

the L-galaxies SAM (Henriques et al. 2015, H15). In the L-galaxies framework, excess supernova energy first reheats gas in the ISM and ejects it into the CGM. If enough energy remains, it can further expel hot halo gas into the IGM, with:

$$\frac{1}{2}\Delta m_{\text{eject}}V_{\text{vir}}^2 = \Delta E_{\text{SN}} - \Delta E_{\text{reheat}}, \quad (16)$$

where  $\Delta E_{\text{reheat}} = \frac{1}{2}\Delta m_{\text{reheat}}V_{\text{vir}}^2$ .

Assuming the quantity  $\Delta m_{\text{eject}}/\Delta m_*$  is equivalent to our  $\eta_{\text{halo}} \equiv \dot{m}_{\text{CGM-IGM}}/\dot{m}_*$ , we derive the expression for  $\eta_{\text{halo}}$  as:

$$\eta_{\text{halo}} = \frac{2\epsilon_{\text{halo}}V_{\text{SN}}^2}{V_{\text{vir}}^2} - \eta_{\text{ISM}} \quad (17)$$

where  $V_{\text{SN}} = 630$  km/s is a constant and  $\epsilon_{\text{halo}}$  is a dimensionless efficiency parameter capped at unity and is given by:

$$\epsilon_{\text{halo}} = \epsilon_{\text{halo},0} \left[ 0.5 + \left( \frac{V_{\text{max}}}{V_{\text{eject}}} \right)^{-\beta_2} \right]. \quad (18)$$

In H15, the best-fit values for  $\epsilon_{\text{halo}}$  are  $\epsilon_{\text{halo},0} = 0.62$ ,  $\beta_2 = 0.80$ , and  $V_{\text{eject}} = 100$  km/s. However, as shown in the right panel of Figure 4, applying these values in TNG (dashed lines) results in  $\epsilon_{\text{halo}}$  values significantly higher than what we see in TNG (solid lines), overestimating the efficiency by up to a factor of 10. To correct this, the TNG SAM instead parameterizes  $\epsilon_{\text{halo}}$  based on its actual dependence on halo mass and redshift in TNG (solid lines). Additionally, to prevent the halo-scale outflow efficiency from reaching unphysically large values in low-mass halos, we cap  $\eta_{\text{halo}}$  at 100, consistent with the calibration applied at the galaxy scale.

In the TNG SAM, all gas ejected from the halo does not reaccrete. Instead, the  $f_{\text{in,CGM}}$  parameter accounts for a fraction of this gas returning to the halo. However, the reaccretion process is not explicitly modeled, as the distinction between pristine and recycled gas cannot be resolved with the Eulerian method used by YO25.

#### 4.4 Metal Enrichment and Circulation

In TNG, metals play a dual role: they trace star formation and chemical enrichment while also regulating galactic winds and cooling rates. The TNG model explicitly tracks the gradual release of metals from asymptotic giant branch (AGB) stars, core-collapse supernovae, and Type Ia supernovae (SNIa), allowing the simulation to follow their redistribution within the ISM and CGM. In contrast, the SC SAM employs a simplified, instantaneous approach to metal production and circulation, described in Section 2.2.3.

To better replicate TNG’s metal flows in the SAM, we introduce metal-enrichment factors that quantify the fraction of metals transported into and out of reservoirs relative to the total gas flow. Following Peeples & Shankar (2011), we define the metal enrichment factor for outflows as:

$$\zeta_{\text{out}} = \frac{\dot{M}_{Z,\text{out}}}{\dot{M}_{\text{out}} \cdot Z_{\text{gas}}}, \quad (19)$$

where  $\dot{M}_{Z,\text{out}}$  is the rate at which metals are outflowing from the ISM/CGM,  $\dot{M}_{\text{out}}$  is the rate at which all gas is outflowing from the ISM/CGM, and  $Z_{\text{gas}}$  describes the metallicity of the gas reservoir from which the outflows originated.

Traditionally, the SC SAM (like most SAMs) assumed  $\zeta_{\text{out}} = 1$  for both the ISM and CGM, meaning that the metal content of the outflowing gas was directly proportional to the metallicity of the ISM/CGM. However, panels *viii.* and *ix.* in Figure 3 show that this assumption is not what is seen in TNG. At the CGM scale,  $\zeta_{\text{out,CGM}}$

often deviates from unity, decreasing significantly at higher redshifts. At the ISM scale,  $\zeta_{\text{out,ISM}}$  consistently falls below unity, indicating that the winds eject fewer metals relative to the average metallicity of the ISM. This result is not too surprising, for the fiducial TNG model sets the wind metal loading factor  $\gamma_w = 0.4$  (Vogelsberger et al. 2013), indicating that the winds are under-enriched compared to the average metallicity of the ISM. However, as the winds reach the scale of the galaxy, they entrain additional metal mass, as evidenced by the fraction of metals outflowing consistently being greater than 0.4 (the grey dashed line in panel *viii.*).

In addition to calibrating metal enrichment factors for gas outflows, we also created inflow enrichment factors to capture the metallicity of gas entering the ISM and CGM from the CGM and IGM, respectively. For the ISM, we define  $\zeta_{\text{ISM}}^{\text{in}}$  as:

$$\zeta_{\text{ISM}}^{\text{in}} = \frac{\dot{M}_{Z,\text{ISM}}^{\text{in}}}{\dot{M}_{\text{ISM}}^{\text{in}} \cdot Z_{\text{CGM}}}, \quad (20)$$

where  $\dot{M}_{Z,\text{ISM}}^{\text{in}}$  is the metal inflow rate into the ISM,  $\dot{M}_{\text{ISM}}^{\text{in}}$  is the total gas inflow rate, and  $Z_{\text{CGM}}$  is the metallicity of the CGM. For the CGM, since the SAM does not track the metallicity of the IGM, we define  $\zeta_{\text{CGM}}^{\text{in}}$  as the fraction of metals entering the CGM compared to the total rate of gas flowing into the CGM:

$$\zeta_{\text{CGM}}^{\text{in}} = \frac{\dot{M}_{Z,\text{CGM}}^{\text{in}}}{\dot{M}_{\text{CGM}}^{\text{in}}}. \quad (21)$$

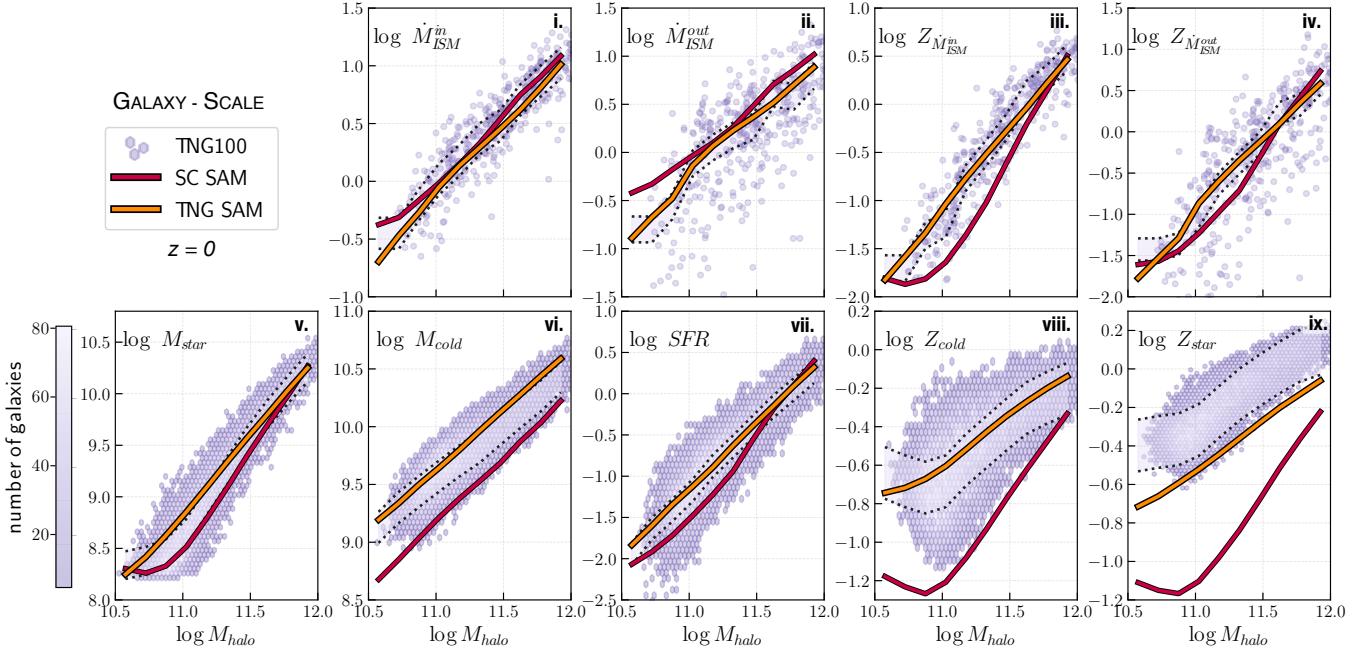
Notably, Figure 3 panel *vii.* shows that the fraction of metals accreting into the ISM in TNG often exceeds or lies below unity across redshift, whereas the SC SAM (like most/all SAMs) assumes a fixed value of unity. This indicates that the gas returning to the ISM is more metal-rich than the average CGM, suggesting efficient cooling of metal-enriched gas.

While neither the SAM nor TNG explicitly track the metallicity of the IGM, the non-zero value of  $\zeta_{\text{CGM}}^{\text{in}}$  implies a non-negligible inflow of metals into the CGM. This suggests that a fraction of the metals ejected via feedback can escape the halo and subsequently be re-accreted, highlighting the importance of considering metal recycling on both galactic and halo scales.

To improve the SAM’s agreement with TNG for metals, we also increased the stellar yield  $y$  from the fiducial value of  $1.2 Z_{\odot}$  in the SC SAM to  $1.5 Z_{\odot}$  to better match the aggregate stellar metallicity found in TNG galaxies. This adjustment ensures that the total metal content locked into stars aligns with the normalization of the stellar metallicity–mass relation in TNG. While TNG does not assume a fixed yield, Torrey et al. 2019 estimate an effective global yield of  $y_{\text{global}} \approx 0.05$ , implying that roughly 5% of the stellar mass formed is returned to the ISM as metals. In comparison, our adopted yield of  $1.5 Z_{\odot}$ , assuming  $Z_{\odot} = 0.02$ , corresponds to a metal production of 3% by mass. Although this is somewhat lower than TNG’s global value, the comparison is not truly one-to-one: TNG computes metal production self-consistently from stellar evolution models, including delayed enrichment from AGB stars, Type Ia supernovae, and core-collapse SNe, whereas the SAM assumes a fixed, instantaneous yield deposited directly into the ISM at the time of star formation.

## 5 REPRODUCING TNG100’S RESULTS OVER 12 GYR

As discussed in Section 1, hydrodynamical simulations and SAMs can model gas and metal flows differently yet still produce similar global predictions for galaxies. The primary goal of the TNG SAM is to use the calibrations presented in Section 4 to not only match TNG’s global properties but also replicate its underlying gas and



**Figure 5.** Comparison of baryon flow rates (panels *i* - *iv*.) and global properties (panels *v* - *ix*.) at the scale of the galaxy for the TNG SAM (orange), SC SAM (brown), and TNG100 (purple). Each panel shows the median (50th percentile) predictions from the SAMs compared to TNG100, with scatter points representing the YO25 TNG subsample and histograms representing the full TNG100 dataset. The dotted black lines and light purple shading mark deviations within  $\pm 30\%$  of TNG100's median. The SC SAM predictions mostly lie outside this range for baryon flows but remain within  $\pm 30\%$  for global properties such as  $M_{\text{star}}$ ,  $M_{\text{ISM}}$ , and SFR. The recalibrated TNG SAM, calibrated to both observations and TNG's baryon cycle, significantly improves upon the SC SAM, with most predictions agreeing within  $\pm 30\%$  of TNG100's median. The largest improvement is in the metallicities of stars and cold gas, where agreement improves from  $\sim 70\%$  with the SC SAM to 20–40% with the TNG SAM.

metal flow cycles. In this section, we evaluate how well the TNG SAM reproduces both the large-scale gas and metal flows and the global galaxy and halo properties in TNG. We also compare the TNG SAM to the fiducial SC SAM at  $z = 0$ , highlighting how differences in their galaxy formation models impact their predictions.

### 5.1 The Galaxy Scale

Figure 5 panels *i*. - *iv*. compares the galaxy-scale inflow and outflow rates of gas and metals at  $z = 0$  among TNG, the TNG SAM, and the fiducial SC SAM.. For gas inflow into the ISM, the TNG SAM closely follows TNG's inflow trends, achieving a median difference of 18% across the halo mass range  $10.5 < \log M_{\text{halo}} < 12$ . While the SC SAM also shows reasonable agreement, it tends to overpredict inflow rates at the lower end of this mass range, due to its shorter cooling times.

For gas outflows from the ISM, the TNG SAM reproduces TNG100's more moderate outflow rates, with a median difference of 11%. By comparison, the SC SAM predicts significantly higher outflow rates—up to 60% greater than TNG—particularly for lower-mass halos. For metal flows, the TNG SAM continues to perform well, capturing inflow and outflow rates with median differences of 14% and 9%, respectively, across halo masses. In contrast, the SC SAM underestimates both metal inflow and outflow rates by  $\sim 35$  –  $60\%$ , for reasons discussed further in Section 6.1.5

The TNG SAM's ability to reproduce gas flows into and out of galaxies within  $\sim 20\%$  of TNG's values translates into similarly strong agreement for global galaxy properties. Panels *v*. - *ix*. of Figure 5 compares key global properties—stellar mass, cold gas mass, star formation rate, and the metallicities of the stellar population and cold gas—between TNG, the TNG SAM, and the SC SAM. For stellar mass, the TNG SAM closely follows TNG, with average dis-

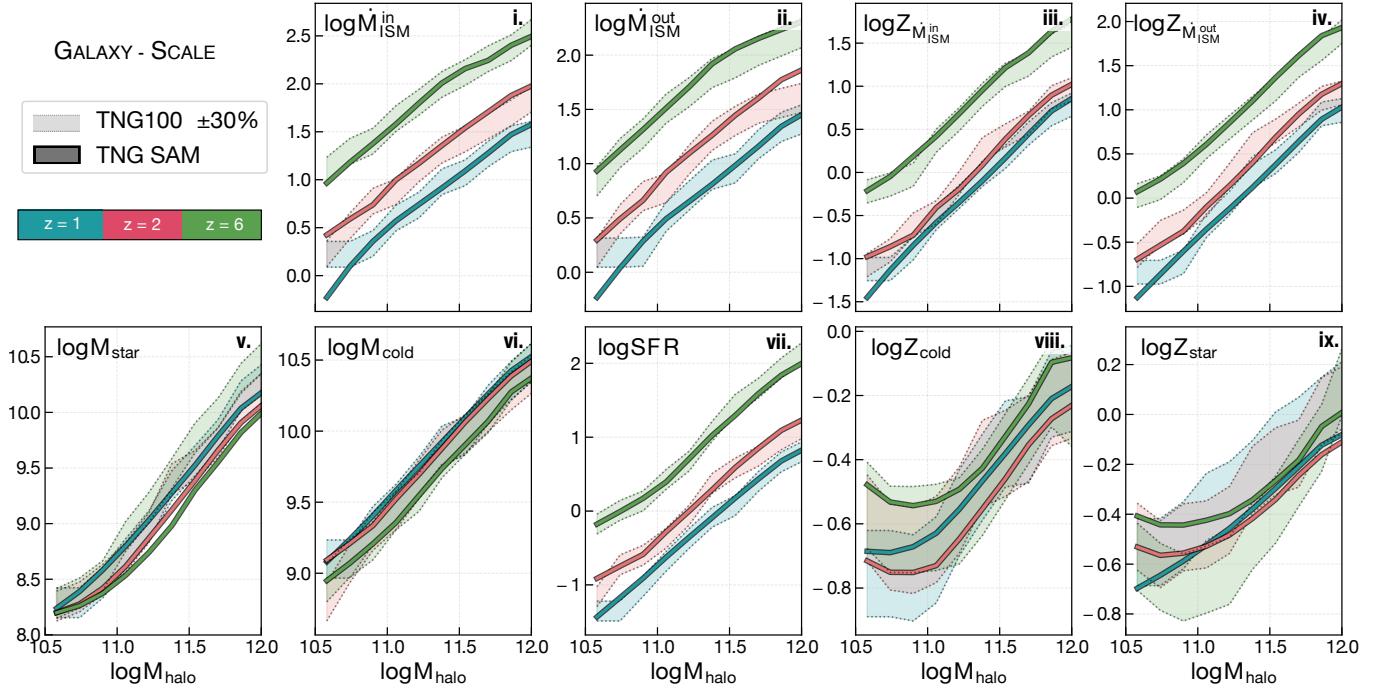
crepancies reduced to 11%. The SC SAM also performs reasonably well, agreeing with TNG within 30% across the mass range, despite differing treatments of gas accretion and star formation. This again demonstrates that different baryon cycling processes can yield similar global galaxy properties, raising the question of which physical processes actually govern galaxy evolution in the real universe.

For the cold gas mass, the TNG SAM performs moderately well, with most deviations within 25%. In contrast, the SC SAM underpredicts TNG cold gas masses by  $\sim 50\%$ . Star formation rates show excellent agreement between TNG and the TNG SAM, with differences typically within 6%. However, the SC SAM predicts systematically lower SFRs than TNG across the mass range shown.

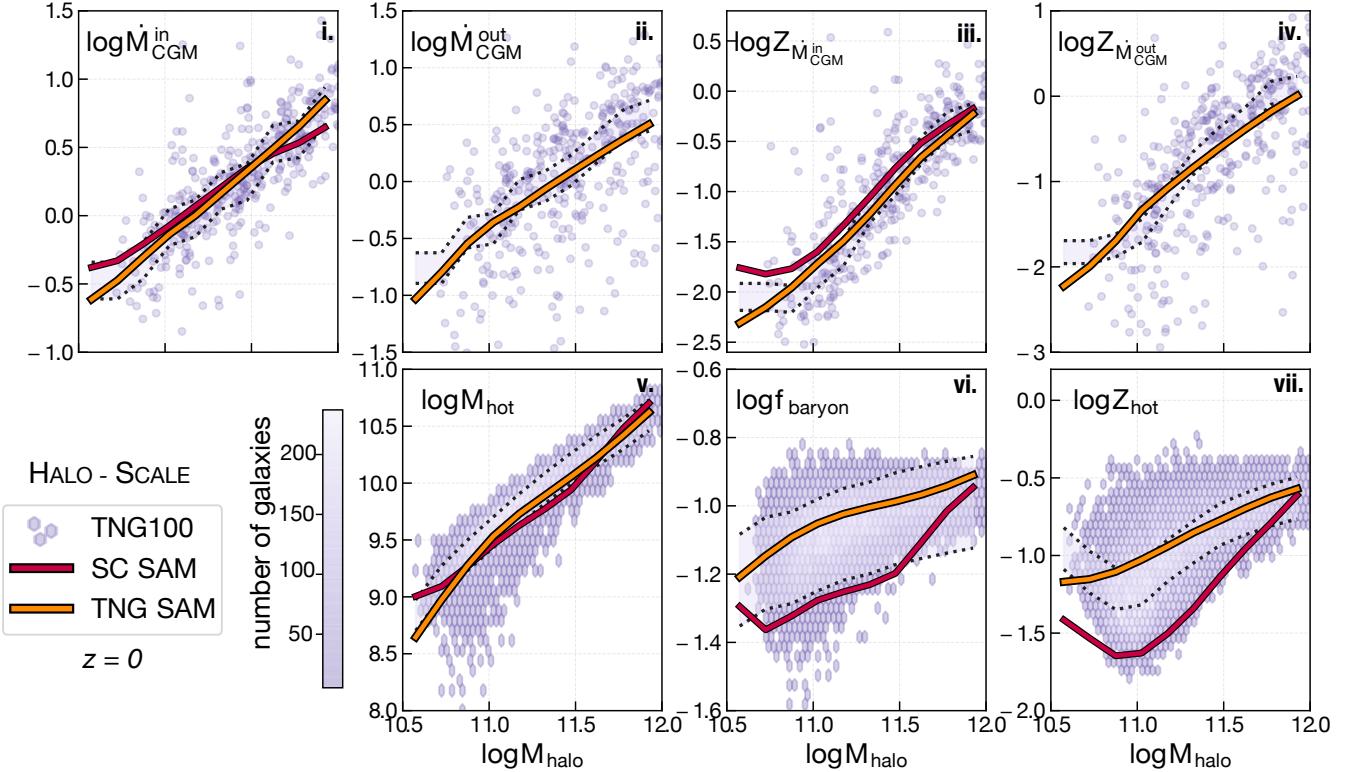
Metallicities in the TNG SAM show impressive agreement with TNG compared to the fiducial SC SAM model. The SC SAM underpredicts stellar and cold gas metallicities by  $\sim 70$  –  $80\%$ , whereas the TNG SAM matches TNG's predictions within 8% for cold gas metallicities and 40% for stellar metallicities.

To evaluate how well the TNG SAM's agreement with TNG at  $z = 0$  holds across cosmic time, Figure 6 presents the same properties at  $z = 1, 2$ , and 6. For gas inflows and outflows in the ISM, the TNG SAM generally remains within 30% of TNG, with deviations increasing to about 50% for lower-mass halos. Metal inflows and outflows show greater consistency, with differences remaining within 20% across the halo mass range and redshifts explored.

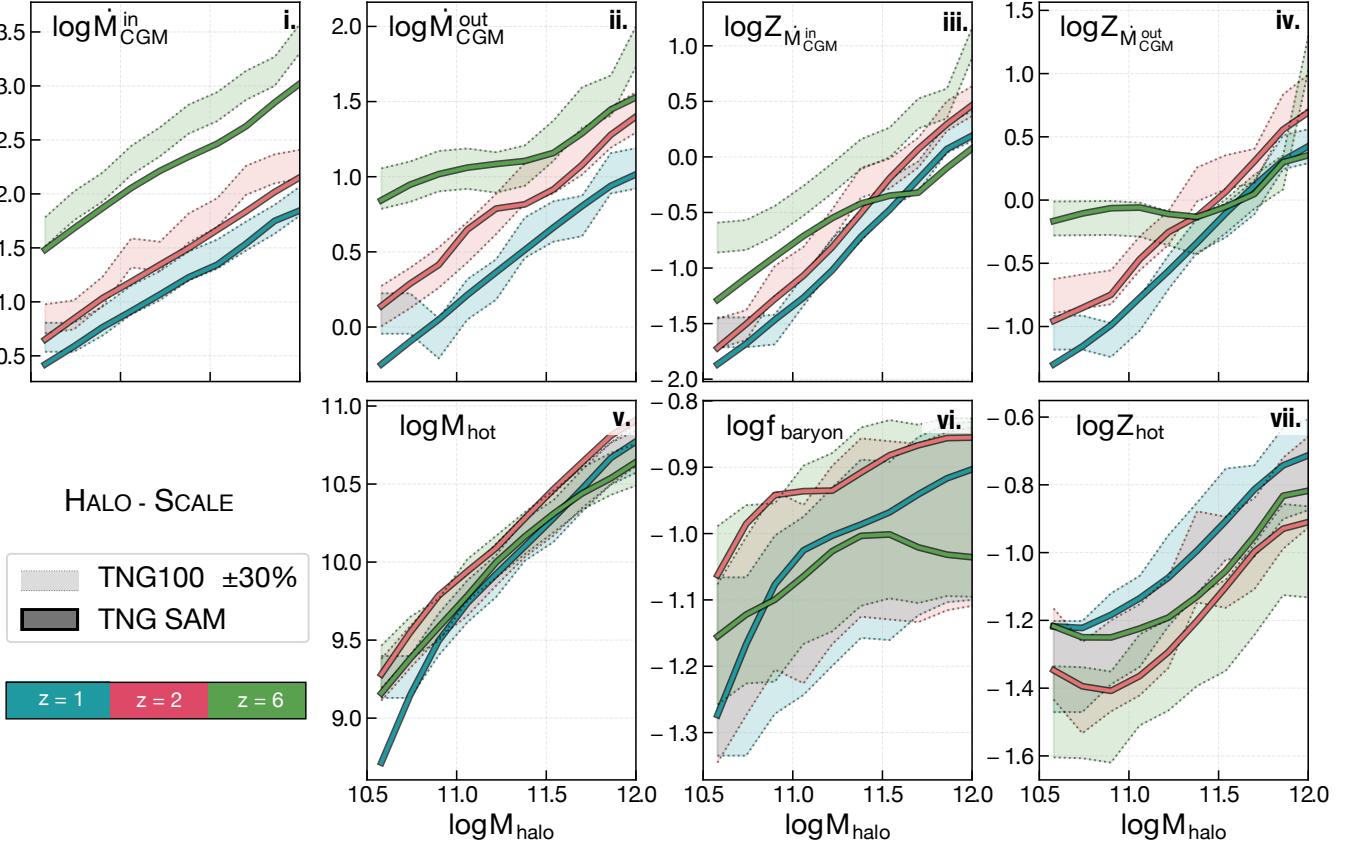
For global properties, the TNG SAM generally tracks TNG's stellar mass, cold gas mass, and star formation rate within 20% across time, with slightly larger deviations at earlier epochs. For example, the TNG SAM shows a maximum deviation of 60% for stellar mass at early times for higher-mass halos. For the metal populations, the TNG SAM consistently provides good ( $< 30\%$ ) agreement with TNG for the cold gas metallicities. Interestingly, the agreement for the stellar metallicities improves at higher redshifts, with deviations decreasing



**Figure 6.** Median galaxy-scale flow rates (panels *i* – *iv*.) and global properties (panels *v*. – *ix*.), comparing the TNG SAM (solid lines) and TNG100 at  $z = 1, 2, 6$ , with blue, red, and green lines representing each redshift. Shaded regions show  $\pm 30\%$  of TNG’s distribution, with dotted gray lines marking the upper and lower bounds. The TNG SAM closely replicates TNG’s gas flow rates, remaining within 30% of the median values. Similar agreement is found for global galaxy properties, although the agreement for  $M_{\text{star}}$  worsens to  $\sim 60\%$  at  $z = 6$ .



**Figure 7.** The same comparison made in Figure 5, now at the halo scale at  $z = 0$ . The SC SAM fluctuates in and out of the  $\pm 30\%$  range of TNG100’s distribution for flow properties and remains mostly within 35% for global properties such as  $M_{\text{CGM}}$  and  $f_{\text{baryon}}$ , while still underpredicting the metallicity of the hot gas. In contrast, the TNG SAM reproduces both global and flow properties within 30% of TNG100’s median values.



**Figure 8.** Median halo-scale flow rates (*i.* – *iv.*) and global properties (*v.* – *vii.*), comparing the TNG SAM and TNG100 at  $z = 1, 2, 6$ . The same color scheme used for galaxy-scale properties in Figure 6 is applied here. The TNG SAM predicts all halo-scale global and flow properties within 30%, except for  $\dot{M}_{\text{CGM}}^{\text{in}}$  at  $z = 6$ , where the agreement falls to 50%.

in the early universe. The largest differences remain observed at  $z = 0$ , where the deviation reaches up to 40%.

## 5.2 The Halo Scale

Figure 7 assesses how well the TNG SAM and SC SAM reproduce halo-scale gas and metal flows in TNG at  $z = 0$ . For gas and metal inflows into the halo (panels *i.* – *iv.*), the TNG SAM generally matches TNG’s predictions within 20%. In contrast, the SC SAM, which uses a more simplified model for gas accretion, predicts inflow rates up to 45% higher than TNG across halo masses. This discrepancy is particularly evident for metal inflows, where the SC SAM’s more efficient recycling leads to higher predicted rates, while the TNG SAM aligns more closely with TNG’s gradual metal flows, maintaining accuracy within 30% across the mass range displayed.

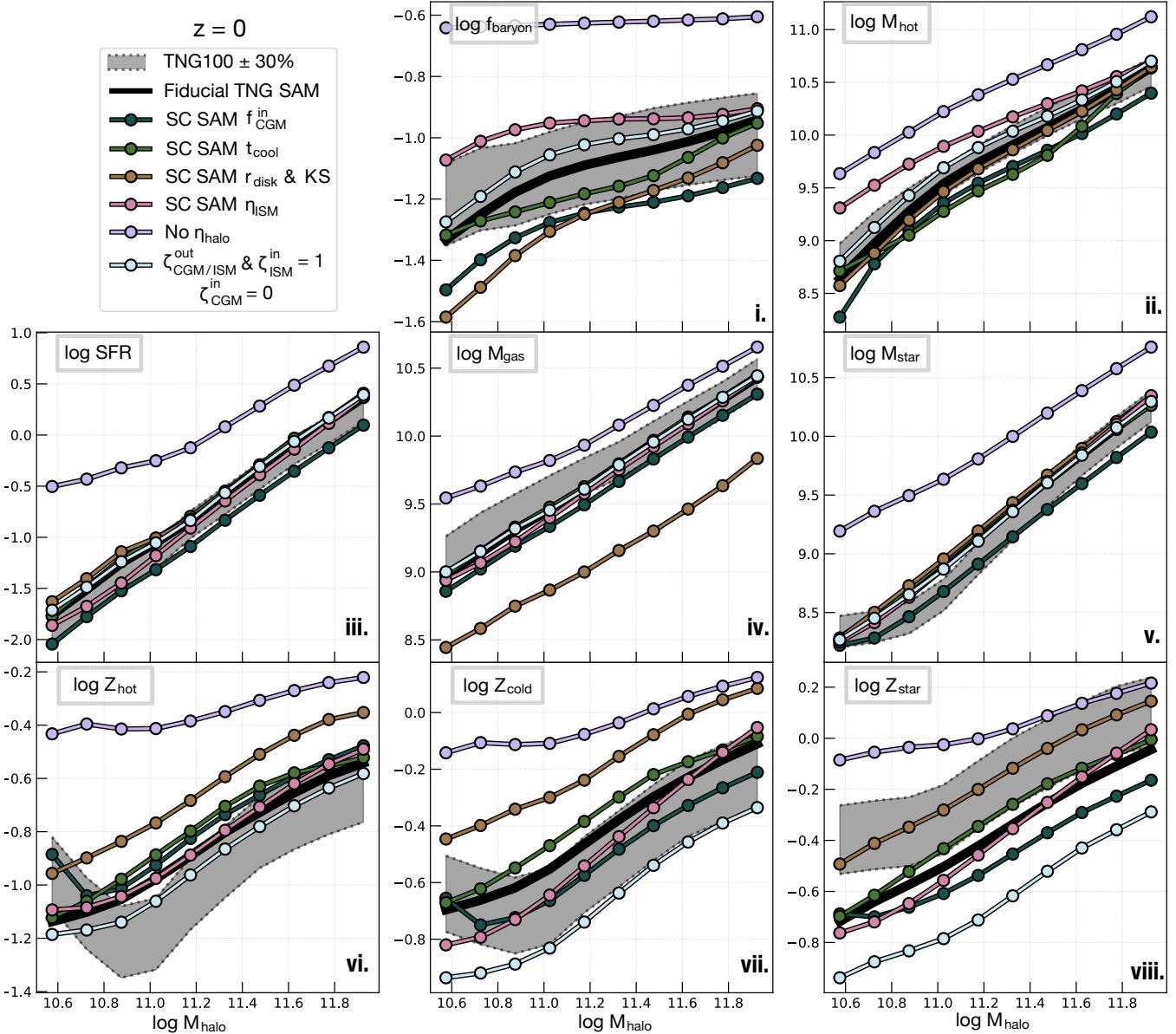
One of the most notable changes in the TNG SAM is the explicit inclusion of gas and metal expulsion from the CGM to the ejected reservoir, a channel not included in the SC SAM and thus not shown in Figures 7 and 8. While these outflows are generally an order of magnitude weaker than the inflowing component, their rates are comparable to gas outflows from the ISM, highlighting their importance in regulating the baryon cycle. The TNG SAM’s gas and metal outflow rates closely match TNG’s predictions, with deviations of ~ 20% and 30%, respectively.

The TNG SAM’s ability to reproduce gas flows in the CGM within ~ 30% of TNG’s values translates into a similarly strong agreement for global halo properties. Figure 7 panels *v.* – *vii.* compares key halo-scale quantities—hot gas mass, baryon fraction, and the metallicity of the hot gas reservoir—across TNG, the SC SAM, and the TNG

SAM. The TNG SAM closely matches TNG’s hot halo mass, with deviations generally within 20% across the halo mass range. Despite lacking halo-scale outflows, the SC SAM predicts only slightly lower hot gas masses, again demonstrating how different baryon cycling processes can still produce similar global outcomes.

For the overall baryon fraction, the TNG SAM matches TNG within 10% accuracy, while the SC SAM underpredicts the baryon fraction by ~ 35% at low to intermediate halo masses, consistent with its slight underestimation of the hot halo gas mass. Similarly, the TNG SAM closely reproduces TNG’s predictions for hot gas metallicities, with deviations within 12%. In contrast, the SC SAM significantly underpredicts hot gas metallicities by ~ 65%.

Figure 8 examines how the TNG SAM’s agreement with TNG evolves over time at the scale of the halo. For gas and metal inflows and outflows, the TNG SAM generally stays within 30% of TNG’s values, except at  $z = 6$ , where inflow rates are underpredicted by 50 – 60%. This discrepancy arises from the simplified way the SAM calculates dark matter accretion rates, which we discuss further in Section 6.2.2. Nevertheless, for global halo properties, the TNG SAM maintains good agreement with TNG over time, with deviations for the baryon fraction and hot gas mass staying within 30%. For the metallicity of the hot gas reservoir, the TNG SAM consistently achieves strong agreement (< 30%).



**Figure 9.** To show the importance of each calibration made to the TNG SAM, we evaluate how reverting individual model components to their original SC SAM formulations impacts the TNG SAM’s ability to reproduce TNG’s global results at the galaxy and halo scales at  $z = 0$ . The solid black line represents the 50th percentile of the fully calibrated TNG SAM, while the shaded gray region marks  $\pm 30\%$  of the TNG100 population, with dotted lines indicating the upper and lower boundaries. Solid lines with circles represent the TNG SAM’s performance when specific calibrations are reverted, with each line colored according to the removed new model component.

## 6 DISCUSSION

### 6.1 Why do the SAM and TNG Agree?

The TNG SAM’s ability to reproduce the median galaxy- and halo-scale properties of  $\sim 20,000$  TNG galaxies mostly within  $\sim 10\text{--}30\%$  highlights the effectiveness of recalibrating SAMs using insights from hydrodynamical simulations. Below, we examine the specific physical modifications that facilitated this agreement for the galaxy- and halo-scale properties compared. Figures 9 and 10 summarize the effect of each modification, showing the TNG SAM’s performance before and after these adjustments at  $z = 0$  and their impact on each of the compared properties.

#### 6.1.1 Efficient Halo (Re-)accretion: Strong Gas Recycling and Weak Preventive Feedback

The efficiency of gas accretion into halos, regulated by  $f_{\text{in,CGM}}$ , plays a significant role in the TNG SAM’s ability to replicate TNG’s baryon cycle. As shown in Figure 3,  $f_{\text{in,CGM}}$  in TNG often exceeds unity, particularly at lower redshifts, indicating that gas inflow frequently meets or surpasses the expected baryon-to-dark matter ratio. In contrast,  $f_{\text{in,CGM}}$  falls below unity in some other hydrodynamical simulations like FIRE (Pandya et al. 2020; Pandya 2021) and EAGLE (Mitchell et al. 2020; Mitchell & Schaye 2022; Wright et al. 2020), reflecting stronger “preventive” feedback that heats the IGM and limits gas accretion onto halos. TNG’s elevated  $f_{\text{in,CGM}}$  suggests that

ejected gas re-enters halos relatively rapidly, mitigating the impact of any preventive feedback that may be present on the halo scale.

With  $f_{\text{in,CGM}}$  remaining high across all halo masses explored, the question becomes: what's driving such efficient gas recycling in TNG? For low-mass halos ( $10.5 < \log M_{\text{halo}} < 11$ ), stellar feedback plays the primary role in ejecting the gas that gets recycled, as demonstrated by YO25. As halo mass increases, AGN feedback becomes increasingly significant, and YO25 notes that for these halos, the number of galaxies dominated by AGN vs supernovae feedback is roughly balanced. Despite this transition,  $f_{\text{in,CGM}}$  remains consistently elevated, indicating that both stellar and AGN feedback are effective at driving gas beyond the galaxy and into the CGM, but neither fully unbinds the gas from the larger potential surrounding the halo. As a result, a significant fraction of the ejected material is fairly quickly recycled back into the halo over time.

In the TNG SAM, the function  $f_{\text{in,CGM}}$  replaces the model for the return of ejected gas, governed by the parameter  $f_{\text{return}} = 0.1$  used in the SC SAM. Most SAMs also rely on static values for gas recycling (e.g., 0.64 in GALFORM, 1.0 in L-galaxies), although the parameterization of the re-accretion time in terms of halo mass and dynamical time varies (see e.g. H15). However, static return fractions clearly oversimplify the dynamic accretion and recycling processes seen in hydrodynamical simulations like TNG. Without  $f_{\text{in,CGM}}$ , the TNG SAM underpredicts the CGM gas inflow rate by roughly 45% (Figure 10, panel *i*).

Part of this discrepancy stems from how inflows are defined and measured. The SC SAM's  $\dot{M}_{\text{CGM}}^{\text{in}}$  output includes only smooth accretion and the re-accretion of previously ejected gas – it does not account for gas delivered to the CGM by merging satellites, whose hot or ejected gas reservoirs are stripped and incorporated into the central halo. In contrast, the TNG flow sample measures gas inflow directly across a thin spherical shell, capturing smooth accretion, re-accretion, and merger-driven contributions alike. As a result, the SC SAM should naturally underpredict the CGM inflow rate relative to TNG. However, the magnitude of this discrepancy, coupled with a corresponding underprediction of the hot gas mass by  $\sim 45\%$  (Figure 9, panel *ii*), suggests that missing merger-driven inflows alone cannot account for the full deficit. The lower inflow rate leads to insufficient hot gas buildup, suppressing cold gas accretion, star formation, and stellar mass growth across the galaxy population.

Despite the effectiveness of  $f_{\text{in,CGM}}$  in the SAM, uncertainties remain regarding the precise nature of gas recycling in TNG. TNG's mesh-based approach makes it difficult to track whether accreted gas is pristine or recycled, unlike particle-based simulations like EAGLE, where gas flows can be explicitly traced. In EAGLE, for instance, Mitchell & Schaye (2022) found that the halo recycling efficiency increases monotonically with halo mass and redshift. A tracer particle analysis of the flows in TNG could provide valuable insights into whether preventive feedback still plays a role in first-time gas accretion in low-mass halos, although recycling is clearly the dominant process once it begins.

Looking ahead, the next generation of SAMs would greatly benefit from incorporating more flexible and physically motivated recycling models that are guided by the explicit tracking of gas flows observed in hydrodynamical simulations. While the  $f_{\text{in,CGM}}$  parameter offers a useful step in replacing the ad hoc static recycling fractions used in most SAMs, further refinement—such as directly modeling the recycling efficiency as it evolves with halo mass and redshift could also be helpful. For instance, based on fits to multiple hydrodynamical simulations SAGE (Croton et al. 2016) dynamically adjusts the reincorporation rate so that gas is returned to the halo more efficiently in massive halos.

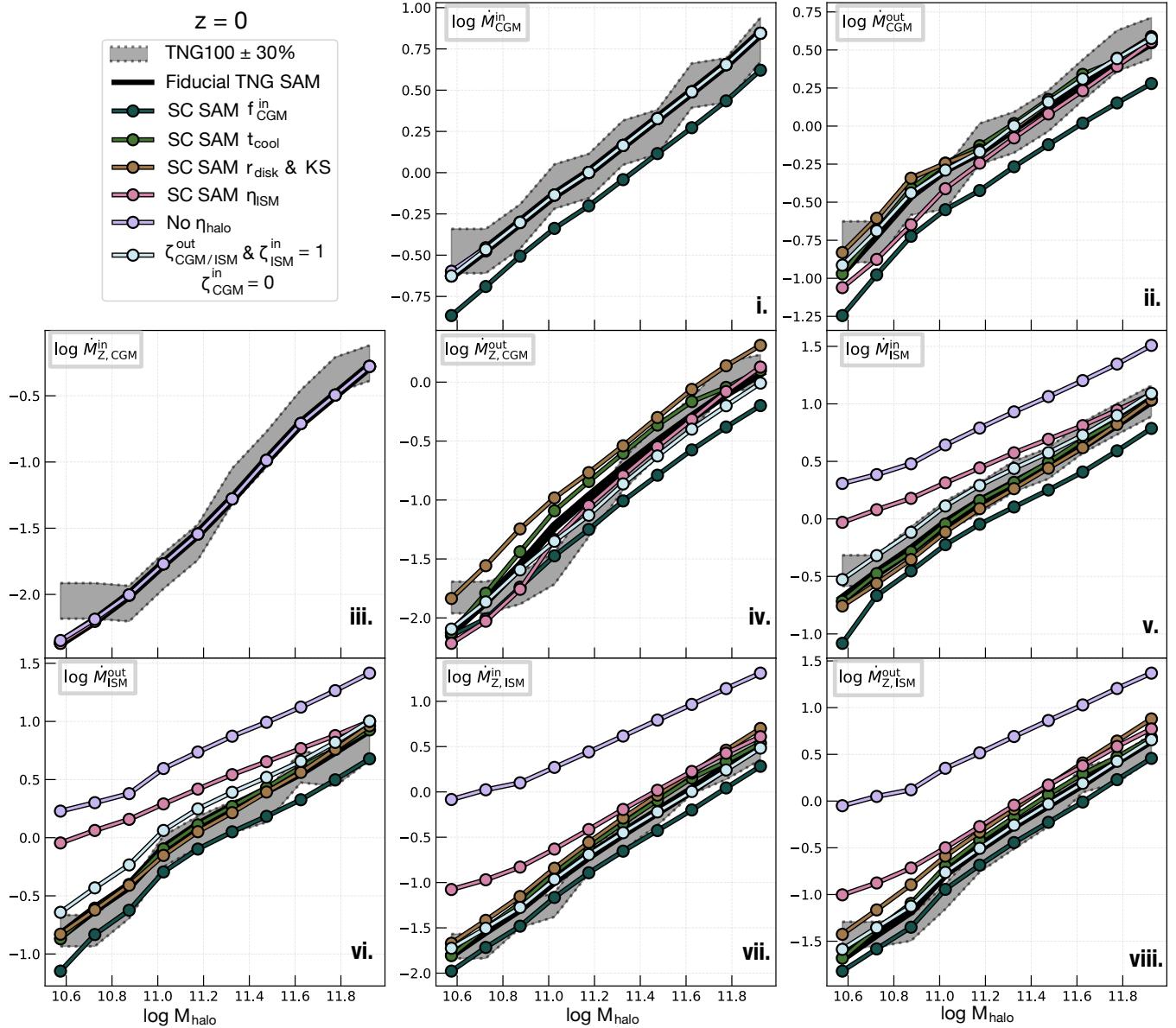
### 6.1.2 Revised Cooling Model: Limitations of the Cold-Mode vs. Hot-Mode Dichotomy

Another significant improvement to the TNG SAM came from revising the cooling model for the CGM onto the central galaxy. The overwhelming majority of SAMs published in the literature (e.g., Galacticus (Benson 2012), Shark (Lagos et al. 2018), GAEA (Hirschmann et al. 2016), Morgana (Monaco et al. 2007)) classify gas accretion as either “cold-mode” or “hot-mode” based on the ratio of the cooling radius to the virial radius. In these models, radiative cooling is governed by well-established processes such as collisional excitation, ionization, recombination, and bremsstrahlung, with the assumption that the hot gas is shock-heated to the virial temperature of the host halo and that chemical abundances are well-mixed throughout the gas. As cooling depends on the gas density and metallicity, denser gas at the center of the halo is assumed to cool faster than outer gas, resulting in an inside-out cooling pattern. While this simplified approach captures the basic mechanisms of gas cooling, it has been shown to lead to systematic discrepancies when compared to cosmological hydrodynamic simulations, with SAMs systematically underpredicting gas accretion rates onto low-mass halos and overpredicting them for massive halos (Lu et al. 2011).

In TNG, we find the traditional cooling model has the opposite trend for low-mass halos. The left panels of Figure 11 show that cooling times in TNG often exceed the expected timescales for both cold- and hot-mode accretion. To distinguish between cold and hot accretion modes, we compare the cooling radius  $r_{\text{cool}}$ —defined as the radius within which the gas cooling time equals the halo dynamical time—to the halo virial radius  $r_{\text{vir}}$ . If  $r_{\text{cool}} \geq r_{\text{vir}}$ , the halo is classified as cold-mode dominated, implying gas can cool rapidly without shock-heating (White & Frenk 1991). Conversely, if  $r_{\text{cool}} < r_{\text{vir}}$ , the halo is considered hot-mode dominated, with gas shock-heating near the virial radius and cooling inefficiently thereafter. Even in halos where the cooling radius exceeds the virial radius (the cold-mode regime), the cooling time frequently surpasses the dynamical time (dashed line). If cooling were dominated by these rapid, cold-mode flows, we would expect significantly shorter cooling times and higher accretion ISM rates than those found in TNG.

However, the TNG model incorporates additional local factors—including gas density, temperature, element-based metal line cooling, and radiation from the UV background and nearby AGN—that regulate radiative cooling rates. As a result, CGM gas cools more gradually, more closely resembling the delayed cooling expected from hot-mode accretion rather than a distinct cold-mode process. Instead of two separate cooling channels, cooling in TNG seems to follow a single mode. Although identifying the exact refinements in TNG's radiative cooling model that lead to disagreement with SAMs is beyond the scope of this paper, such an investigation could provide valuable insights for refining SAMs in the future.

Interestingly, this behavior is seen in other hydrodynamic simulations besides TNG. The right panel of Figure 11 compares cooling times in TNG and FIRE, where each yellow triangle represents an individual FIRE halo taken from Figure 4.22 of Pandya (2021). Both simulations yield cooling times generally longer than those expected under the classic cold-mode accretion model, even in halos where the cooling radius exceeds the virial radius. This similarity persists despite the two simulations modeling feedback very differently: FIRE uses high-resolution, explicit feedback to capture bursty, localized stellar-driven outflows for a limited number of halos, while TNG applies sub-grid prescriptions that distribute feedback energy over larger volumes across thousands of halos. This result suggests that



**Figure 10.** Continuation of Figure 9. Visualization of how individual calibration adjustments to the TNG SAM impact the flows of gas and metals into and out of galaxies and their surrounding hot halos.

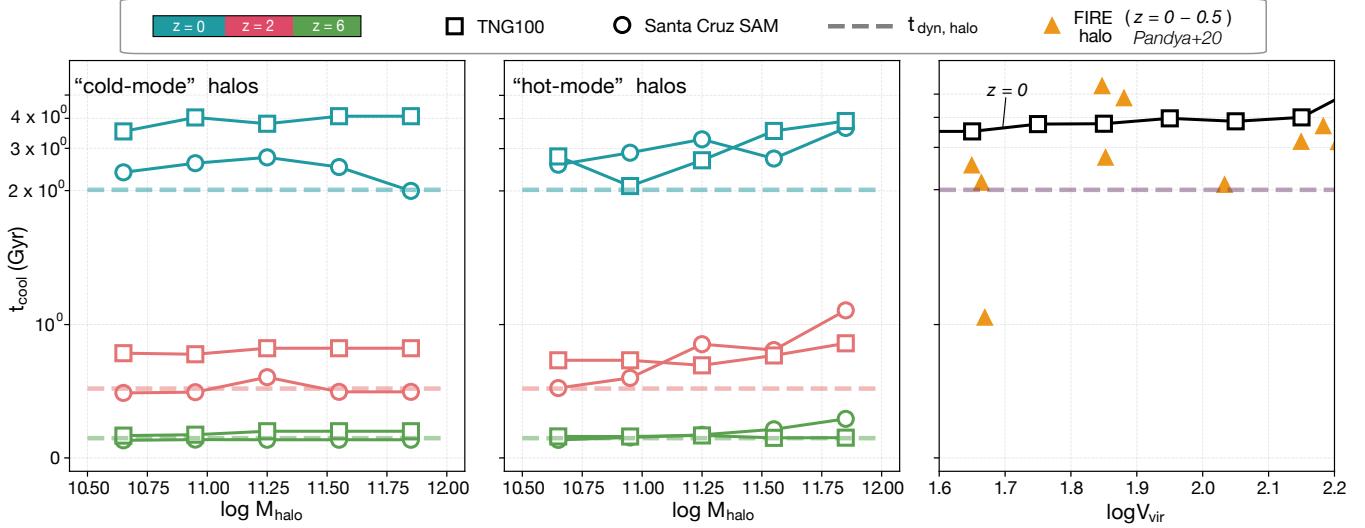
the traditional cooling model used in SAMs may miss key aspects of gas dynamics in realistic galaxy environments.

To address this shortcoming, we revised the TNG SAM’s cooling model to align with the cooling times in TNG. This adjustment significantly improves the SAM’s predictions of halo-scale properties compared to the traditional cooling model, as shown in panels *i.* and *ii.* of Figure 9. In the traditional model, the shorter cooling times lead to faster depletion of the hot gas reservoir, reducing the total hot gas mass by  $\sim 60\%$  and shifting more gas into the ISM. While the instantaneous inflow rate into the ISM appears roughly equivalent between the two models (Figure 10 panel *v.*), this reflects the lower available hot gas mass rather than the true long-term cooling behavior.

While this revision preserves a more realistic hot halo, better aligning with TNG’s predictions, it is important to note that our definition of  $t_{\text{cool}}$  serves as an empirical depletion timescale for CGM-to-ISM gas transfer (see Section 4.1.1), rather than a physically motivated radiative cooling model. It captures the net rate at which hot gas

transitions into the ISM, bundling the effects of radiative cooling, feedback, and other local processes into a single effective timescale, rather than isolating the individual contributions from each process.

Moving forward, refining cooling models to incorporate empirical constraints and better capture radiative processes will likely require not only explicitly tracking mass flows but also modeling energy flows within SAMs. Pandya et al. (2023), Carr et al. (2023), and Voit et al. (2024b,a) demonstrated that accounting for the transfer of feedback energy to the CGM using energy-tracking ordinary differential equations naturally slows cooling rates, producing timescales consistent with those observed in FIRE-2 halos. Given that both TNG and FIRE-2 halos exhibit slow cooling times, such an energy-based framework may also achieve cooling times that better align with TNG’s results.



**Figure 11.** Cooling times ( $t_{\text{cool}}$ ) for TNG (squares) and the SC SAM (circles). The left and middle panels show the time evolution of  $t_{\text{cool}}$  for hot-mode ( $r_{\text{cool}} < r_{\text{vir}}$ ) and cold-mode ( $r_{\text{cool}} > r_{\text{vir}}$ ) halos, respectively, at  $z = 0, 2, 6$ . The dashed lines represent the dynamical time. In TNG,  $t_{\text{cool}}$  generally remains above the dynamical time for both modes, highlighting the limitations of the traditional binary accretion classification in SAMs. The right panel compares  $t_{\text{cool}}$  as a function of virial velocity ( $V_{\text{vir}}$ ) in the median TNG100 population (black squares), and individual FIRE halos (orange triangles,  $z = 0 - 0.5$ ; Figure 4.22 in Pandya 2021). Despite differing methodologies, both TNG and FIRE predict  $t_{\text{cool}}$  values generally longer than the dynamical time (dashed line).

### 6.1.3 The Star Formation Efficiency Drives ISM Agreement

Despite the fundamental difference in spatial scales at which star formation recipes are applied in TNG—locally at the scale of  $10^6 M_{\odot}$  gas cells—and in the TNG SAM—globally across the entire galaxy—the TNG SAM’s calibration to the full TNG00 sample’s star formation efficiency as a function of halo mass and redshift successfully reproduces the global ISM properties very well. The cold gas mass, stellar mass and star formation rate all closely follow TNG’s trends within 30%. In contrast, Figure 9 panels *iii*. - *v*. show that the fiducial SAM’s Bigiel et al. (2008)-inspired star formation recipe overestimates the star formation rate and stellar mass by 35%, which depletes the cold gas reservoir by  $\sim 80\%$ . At face value, the TNG SAM’s agreement isn’t particularly unexpected given that the SAM’s star formation rate is set directly by the star formation efficiency. However, this agreement does not imply that any global star-formation law will perform equally well.

Although not shown, we find that even if the TNG SAM applies a global Kennicutt-Schmidt star formation recipe calibrated to match the global star formation rate density vs gas density in TNG galaxies, the SAM still struggles to reproduce the correct star formation rate, stellar mass, and cold gas mass within 30% across cosmic time. This mismatch is mostly driven by the discrepancy between disk sizes predicted by the SAM and those produced by TNG.

Most SAMs, including the SC SAM, determine disk sizes through angular momentum conservation under the assumption of an exponential disk. While this method achieves reasonably good agreement with observed radial disk sizes as a function of stellar mass up to  $z \sim 2$  in the SC SAM (Somerville et al. 2008b), it does not align with observations as well as TNG’s predictions (Genel et al. 2018). Recent work (e.g., Karmakar et al. 2023) has shown that TNG matches the observed size–mass relation across a range of redshifts. In contrast, the SC SAM tends to systematically underestimate disk sizes for low-mass galaxies and at higher redshifts (see Appendix ?? for a direct comparison).

Given the current uncertainties in predicting disk sizes in SAMs, calibrating the star formation rate using the star formation efficiency offers a more robust and flexible alternative. Better disk size models

will likely be needed before surface-density-based star-formation prescriptions can be reliably applied in future SAMs.

### 6.1.4 ISM and Halo-Scale Outflows Modulate the Stellar Mass and the Mass of the Hot Halo

While galaxy-scale outflows are well-established as incredibly important regulators of galaxy evolution in numerical simulations, recent studies have underscored the importance of outflows at the halo scale as well (Wright et al. 2024; Pandya et al. 2021; Mitchell et al. 2020). Consistent with these findings, we observe that both galaxy and halo-scale outflows are essential for accurately modeling the gas content in the TNG SAM.

Incorporating TNG’s galactic winds model into the TNG SAM, which explicitly ties the wind velocity to the metallicity of the cold gas, significantly improved the model’s ability to reproduce the magnitude of the galaxy-scale outflows found in TNG. Although this added complexity, it was essential for avoiding the  $\sim 30\%$  overestimate in outflow rates produced when using the SC SAM’s outflow model within the TNG SAM framework (Figure 10 panel *vi*.). The SC SAM’s model expels too much gas from the ISM into the CGM, leading to excessive buildup of CGM mass and artificially high  $M_{\text{CGM}}$  and  $f_{\text{baryon}}$  values relative to TNG.

Although not obvious in panel *v*., we also find that correctly extrapolating  $\eta_{\text{ISM}}$  at lower halo masses plays an important role in predicting the stellar mass. If  $\eta_{\text{ISM}}$  is too low in lower-mass halos, the stellar mass exceeds TNG’s predictions by nearly 40% because more cold gas remains available for star formation. The difference relative to the SC SAM is not immediately obvious because the SC SAM expels even more gas from low-mass halos, with  $\eta_{\text{ISM}}$  exceeding 1000 in  $10^8 M_{\odot}$  halos at  $z = 0$ . Whether such values are realistic remains a subject of debate, but we find good agreement with TNG by capping  $\eta_{\text{ISM}}$  at 100 in low-mass halos.

The most critical update to the TNG SAM’s feedback model was the explicit inclusion of halo-scale outflows, which are not modeled in the SC SAM or in most traditional SAMs (e.g., Croton et al. 2006, 2016; Lacey et al. 2016; Hirschmann et al. 2016), where gas is ejected directly from the ISM into an external reservoir without first pass-

ing through the halo. More recent SAMs (e.g., [Henriques et al. 2015](#); [Benson 2012](#); [Lagos et al. 2018](#)) explicitly model gas cycling through the CGM before ejection. Similarly, in the TNG SAM, gas first leaves the ISM and enters the CGM, where it can later be expelled into the ejected reservoir/IGM, forming the CGM-to-IGM outflow channel. This two-step process provides a more realistic depiction of how gas is cycled and expelled, as opposed to the simplified ISM-to-ejected reservoir flow used in the SC SAM. Neglecting halo-scale outflows, as shown in Figure 9 panels *i.* & *ii.*, results in substantial overpredictions of  $M_{\text{CGM}}$  and  $f_{\text{baryon}}$  by over 200%. These overestimates propagate into the ISM, causing  $M_{\text{ISM}}$ , SFR, and  $M_{\star}$  (panels *iii.* - *v.*) to be overestimated by 100–300%.

### 6.1.5 Metal Cycling Efficiencies Improve Metallicity Predictions

Incorporating metallicity-weighted mass-loading factors greatly enhanced the TNG SAM’s ability to replicate the evolution of cold gas, stellar, and hot halo metallicities found in TNG100. Without these refinements, the TNG SAM underpredicts the cold gas and stellar metallicities by up to  $\sim 70\%$ , as shown in Figure 9 panels *vii.* & *viii.*. While the metallicity of the CGM is less affected (panel *vi.*), discrepancies up to 20% persist, consistent with the TNG SAM.

Introducing metal cycling efficiencies mostly improves the TNG SAM’s predictions for the cold gas metallicities. The agreement for stellar metallicities, however, shows a more complex pattern. At higher redshifts, the stellar and cold gas metallicities in TNG are nearly equal, consistent with the instantaneous recycling approximation used in the TNG SAM, which assumes that metals produced by stars are immediately mixed into the surrounding gas. This leads to improved agreement for stellar metallicities at early times. However, at lower redshifts, this approximation breaks down in TNG, where time delays in metal transport and mixing become more significant. As a result, the TNG SAM increasingly underpredicts stellar metallicities relative to TNG at later times. Despite these challenges, the introduction of metallicity-weighted mass-loading factors compensates for some of the model’s simplifications, allowing the TNG SAM to better track the movement of metals between the ISM, CGM, and IGM.

Most SAMs in the literature use instantaneous recycling models for metal production, although some have incorporated more sophisticated multi-element galactic chemical evolution models that relax the instantaneous recycling approximation (e.g., [Arrigoni et al. 2010](#); [Yates et al. 2012](#); [Kobayashi et al. 2007](#); [Hirschmann et al. 2012](#)). However, nearly all SAMs, including the Santa Cruz SAM, also assume that metals are exchanged between baryonic reservoirs in direct proportion to the exchanged gas mass, without accounting for metal-enhanced or metal-depleted inflows and outflows. A few SAMs, such as SAG ([Collacchioni et al. 2018](#)) and GALFORM ([Lagos et al. 2013](#)), have experimented with incorporating metal enrichment factors. However, these models found little impact on reproducing the observed stellar mass-metallicity relation, perhaps because their supernova feedback models are not tied to the metallicity of the gas. [YO25](#) and our work show that both inflows and outflows can be either metal-enhanced or metal-depleted relative to the reservoir of origin, and these metal enrichment factors can have a complex dependence on various halo properties. Incorporating these metal enrichment factors into SAMs is critical for reproducing the metal cycle in realistic simulations.

## 6.2 Limitations and Future Directions

Although Section 5 shows that the median galaxy population produced by the TNG SAM agrees quite well with TNG given the physical updates made, there are still several limitations to our approach, discussed below.

### 6.2.1 Modeling Individual Galaxies with Median-Based Calibrations

The TNG SAM is calibrated using median relations that reflect the overall behavior of galaxies within TNG rather than tailoring the model to fit individual galaxies or halos. To evaluate how well the TNG SAM’s median-based calibrations replicate the detailed growth histories of individual TNG galaxies, we directly matched subhalos between TNG and the SAM to perform an object-by-object comparison.

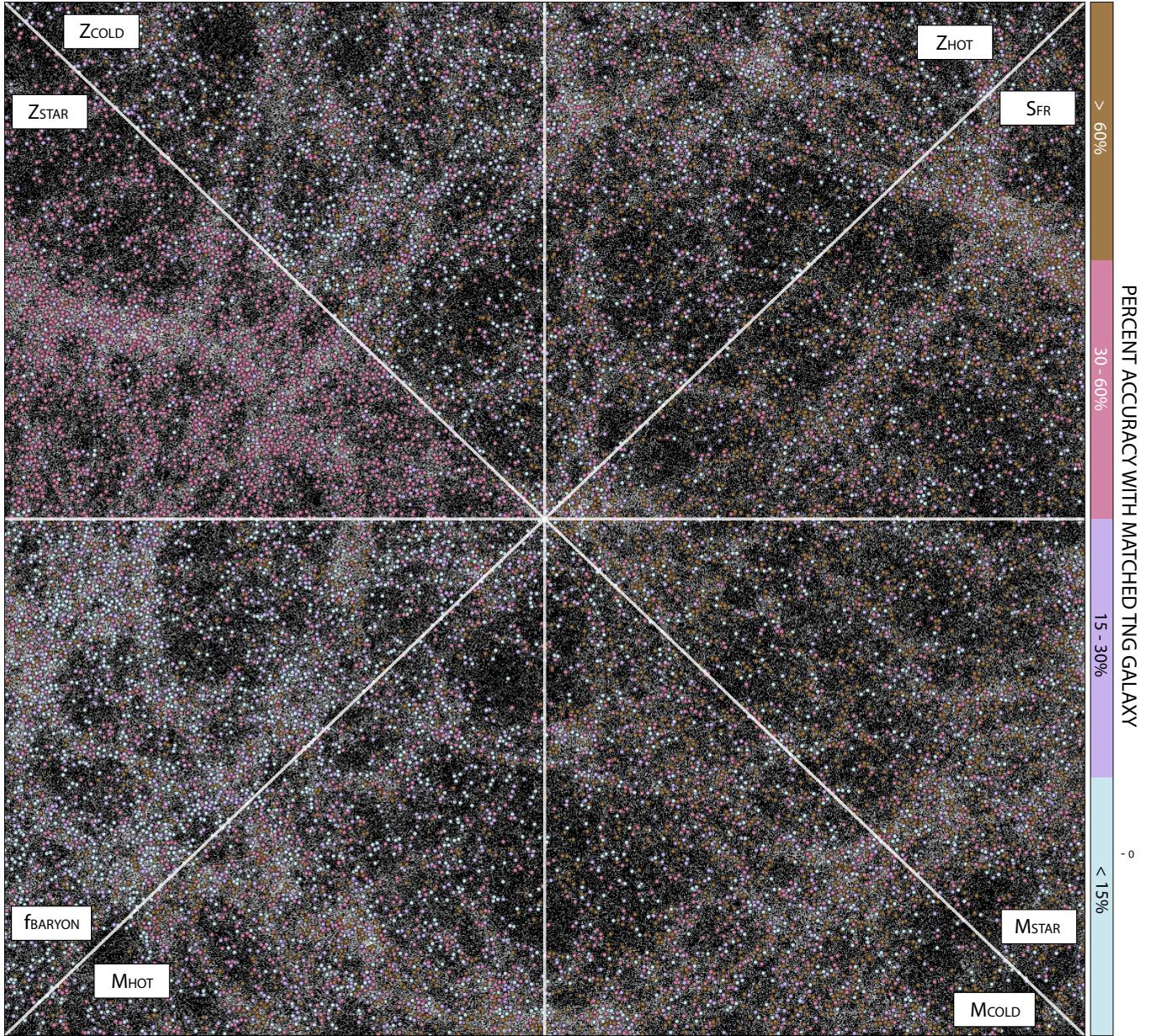
Although the TNG SAM’s median galaxy population is mostly within 30% of the TNG median, Figure 12 shows that the spread increases significantly when comparing individual galaxies. The TNG SAM still demonstrates generally good agreement for several key properties, although this agreement is primarily observed for properties at the halo scale. For instance, the baryon fraction is accurately predicted within 30% for 66% of galaxies, highlighting the SAM’s strong performance in capturing the overall gas content of halos. Similarly, the hot halo mass is generally well-reproduced, with 54% of galaxies falling within 30%.

The agreement is less consistent for galaxy-scale properties. The TNG SAM predicts stellar masses within 30% accuracy for only 40% of galaxies. Cold gas masses show similar performance, with 40% of galaxies meeting the same threshold. Predictions for star formation rates exhibit the weakest agreement, with only 31% of galaxies being accurate to within 30%.

The disagreement at the galaxy scale largely stems from the SAM’s inability to correctly model the variation in galaxy sizes in TNG. Fixing galaxy sizes to 0.25  $r_{\text{vir}}$  proved effective for reproducing median trends but clearly fails to capture the variability in galaxy sizes, which is critical for modeling the spread in galaxies’ star formation rates and stellar and cold gas masses. We tested introducing scatter in disk sizes using the angular momentum model from the fiducial SAM (results not shown), but this only worsened predictions for individual galaxies. While the angular momentum model does introduce galaxy size variability, it does not generate the correct variation found in TNG ([Karmakar et al. 2023](#)). These results further emphasize the need to develop more sophisticated approaches to modeling galaxy sizes and their dependence on halo properties or halo formation histories. However, this may prove challenging, as the sizes in TNG do not show simple correlations with the most obvious halo properties ([Somerville et al. 2025](#)).

Lastly, the TNG SAM’s metallicity predictions show varying levels of agreement, consistent with the trends observed for the median predictions in Section 5. The cold gas metallicity demonstrates relatively strong performance, with 57% of galaxies within 30% of TNG values. The hot gas metallicity performs slightly worse, with 42% of galaxies meeting this threshold. Stellar metallicity shows the weakest performance, with only 22% of galaxies within 30%, though most ( $\sim 76\%$ ) are within 60%.

Despite these limitations, the results are encouraging. Nearly half of the 20,000 matched galaxies agree with TNG’s predictions within 30%, even though the SAM was calibrated using medians derived from a much smaller sample of just  $\sim 400$  galaxies. This level of agreement is particularly promising given the computational simplic-



**Figure 12.** Spatial distribution of matched halos between TNG100 and the TNG SAM at  $z = 0$ , plotted in physical space (x vs. y position). All TNG100 halos with  $10.5 < \log M_{\text{halo}} < 12$  are shown in white. The map is divided into eight octants corresponding to a specific galaxy property: baryon fraction, hot gas mass, cold gas mass, stellar mass, star formation rate, and metallicities of hot gas, cold gas, and stars. Each matched galaxy is colored by the percent difference between the TNG SAM and TNG100 predictions for the given property in each octant. Blue and purple points represent agreement within 30%, while pink and brown points highlight discrepancies greater than 30%. While the overall agreement is strong across most properties, notable scatter is observed for galaxy-scale properties like  $M_{\star}$  and  $M_{\text{ISM}}$ , illustrating the challenges of calibrating to median trends alone.

ity of the SAM compared to the detailed hydrodynamical simulation. Extending the work presented here will provide a foundation for developing the next generation of flexible, computationally efficient, yet physically detailed SAMs needed to model galaxy formation across wider cosmological volumes.

#### 6.2.2 TNG100 DMO vs Hydrodynamical Simulations

A significant limitation to our approach arises from running the TNG SAM on the TNG100-1-DMO simulation while comparing results to the hydrodynamical TNG100-1 simulations. As noted in Section 3.2, the halo masses in the TNG100-1-DM and TNG100-1 simulations differ by up to 20% for low halo masses at  $z = 0$ . This discrepancy

stems from the inclusion of baryonic processes in the hydrodynamical simulation, and varies across redshift.

This variation has significant implications for the TNG SAM, particularly in how the model calculates the rate of gas accretion. The SAM derives the rate of gas accretion by finite differencing the halo masses reported by the DMO simulation. The mismatch between halo masses in the N-body and hydrodynamical simulations affects how well we can model parameters like  $f_{\text{in,CGM}}$ , which governs the growth of the hot halo, which in turn regulates the build up of cold gas and stars in the ISM. Due to this discrepancy, directly tuning  $f_{\text{in,CGM}}$  in the SAM to match the hydrodynamical simulation was not feasible. Instead, we inferred the value of  $f_{\text{in,CGM}}$  by aligning the rate of gas entering the CGM with TNG. This approach performed

reasonably well below  $z = 2$  but still led to a slight overprediction of the baryon fraction in the SAM at  $z = 1$  by nearly 30% (see Figure 8). The discrepancy worsened with redshift, peaking at  $z = 6$ , where the difference in dark matter accretion rates between the SAM and TNG grew to nearly 30%. This caused the SAM to underpredict gas inflows into the CGM, disrupting the modeled baryon cycle at high redshifts.

Furthermore, the above discrepancies in virial mass result in  $\sim 20\%$  differences in the virial radius, directly affecting the SAM’s ability to consistently define the galaxy radius as  $0.25r_{\text{vir}}$  to align with TNG. As discussed in Section 6.1.3, accurate disk sizes are crucial for determining surface densities used in the Kennicutt-Schmidt relation, which directly influence predictions for star formation rates, stellar masses, and cold gas masses, introducing inherent uncertainties when disk sizes are incorrectly estimated.

### 6.2.3 Resolution

The TNG SAM’s ability to reproduce TNG100’s results is further constrained by the halo mass and temporal resolution of the calibration sample. We consider halos well-resolved if they contain more than 100 star particles ( $\log M_{\text{halo}} > 10.5$ ); however, the baryon flow measurements used to calibrate TNG SAM are extracted from radial shells  $0.1r_{\text{vir}}$  thick within each halo. For lower-mass halos, particularly those around  $10^{10} M_{\odot}$ , these shells often contain far fewer resolution elements, producing noisy estimates and contributing to discrepancies in the SAM’s predictions at the low-mass end, where deviations are most pronounced. Moreover, the  $10^{10} M_{\odot}$  halos modeled at  $z = 0$  evolved from progenitors with halo masses of  $\sim 10^8 M_{\odot}$  at  $z = 6$ , which contain poorly resolved stellar and gas components.

This mass resolution limit precluded us from tracking the baryon cycle’s time evolution in a select group of galaxies selected at  $z = 0$ . Instead, we had to rely on aggregate measurements at each redshift, which introduces additional uncertainties for low-mass systems. To account for unresolved galaxies below the well-resolved mass range, we extrapolated their behavior using baryon flow trends observed for halos in the range  $10 < \log M_{\text{halo}} < 10.5$ . While this approach introduces uncertainty—since it is unclear whether these trends hold for  $\log M_{\text{halo}} < 10$ —it plays a critical role in achieving the TNG SAM’s overall 30% agreement with TNG for higher-mass halos. This was particularly important for  $M_{\text{star}}$ , which exceeded TNG’s distribution by 40% in halos with  $\log M_{\text{halo}} < 11$ , if  $\eta_{\text{ISM}}$  was too low for  $\log M_{\text{halo}} < 10.5$  halos. However, this extrapolation also introduces unquantifiable uncertainty, raising questions about the robustness of the 30% agreement achieved.

### 6.2.4 Sample Size

The limited size of the calibration sample used to calibrate the TNG SAM naturally raises two questions: (1) could the TNG SAM’s accuracy improve with a larger calibration set, and (2) how many galaxies are required to maintain reliable predictions? Although the model already reaches  $\sim 30\%$  accuracy using only  $\sim 400$  resolved TNG galaxies, expanding the calibration sample is an obvious path toward reducing noise, particularly at the low-mass end where the measurements are most uncertain. Since the extraction of baryon flow quantities is a one-time task and not computationally prohibitive, scaling to larger samples is entirely feasible and a worthwhile direction for future work, which we discuss further in Section 6.2.5.

Interestingly, we find that the TNG SAM can maintain accuracy for galaxy-scale properties within  $\sim 40\%$  even when calibrated to only

100 galaxies per redshift, randomly sampling the  $10^{10} - 10^{12} M_{\odot}$  halo mass range. The reduced sample size does impact the accuracy of the halo-scale properties quite a bit, however, ranging from 20–80% accuracy (see Figures ?? & ??). These results suggest that a larger sample size could certainly improve the accuracy currently achieved by the TNG SAM, but also that the quality of the calibration sample is more critical than its size.

These results are particularly promising for future efforts, such as those by the SMAUG<sup>1</sup> and Learning The Universe<sup>2</sup> collaborations, to develop next-generation SAMs and hydrodynamical simulations using detailed physical insights from zoom-in simulations and observations. They demonstrate that the quality of the calibration sample—its ability to represent key physical processes—can be as impactful as computationally expensive increases in sample size. This is also promising for multi-wavelength observational surveys, which are often constrained by the difficulty in securing enough observing time across numerous telescope time allocation committees. For example, large surveys like CALIFA comprise around 600 galaxies, while highly detailed surveys such as PHANGS feature  $\sim 90$  galaxies. The TNG SAM’s success with just 100 well-sampled galaxies per snapshot suggests that small but carefully selected observational datasets could also help to refine feedback prescriptions in numerical simulations. While observational data inevitably include greater uncertainties than simulations, high-quality, representative datasets can still aid in improving our understanding of the baryon cycle and galaxy evolution across cosmic time.

### 6.2.5 Future Directions

The TNG SAM is successful in reproducing TNG100’s baryon cycle and scaling relations over the halo mass range  $10.5 < \log M_{\text{halo}} < 12$ , but it still has several limitations, as discussed above. Although TNG100 was the more appropriate choice for this study given its use in the calibration of the fiducial TNG galaxy formation model and the known resolution-dependent discrepancies in TNG (Pillepich et al. 2018a), expanding the analysis to galaxies in the TNG50 simulations, which resolves halos down to  $\sim 10^8 M_{\odot}$ , would provide critical insight into the baryon cycle in lower-mass systems and offer a more comprehensive test of the TNG SAM’s ability to reproduce galaxy properties across a wider range of stellar feedback dominated halos. Applying the same approach to TNG300 would also allow us to investigate the lack of convergence across the TNG suite, potentially uncovering which aspects of the baryon cycle are most sensitive to differences in resolution and box size.

The TNG SAM’s focus on replicating the physical processes of galaxy evolution, rather than just the final outcomes, also opens up exciting possibilities for future refinements. Whereas traditional SAMs track mass flows, as we have in this work, Pandya et al. (2023); Carr et al. (2023); Voit et al. (2024b) and Voit et al. (2024a) advocate for explicitly tracking energy flows as well. Incorporating an energy budget for the CGM that accounts for both energy sources (e.g., stellar and AGN feedback) and sinks (e.g., radiative cooling and turbulence dissipation) would provide a more nuanced understanding of how different processes interact to shape the CGM and its influence on galaxy evolution.

Integrating such an energy-based CGM model into the TNG SAM

<sup>1</sup> <https://www.simonsfoundation.org/flatiron/center-for-computational-astrophysics/galaxy-formation/smaug/>

<sup>2</sup> <https://learning-the-universe.org/>

would make the treatment of feedback more self-consistent, linking the energy injected by stellar winds and AGN directly to the thermal and kinetic state of the CGM. This could improve how the model handles gas cooling, star formation, and other key processes. An energy-based CGM model also allows for calibration against both observations and simulations: observational constraints from JWST, MUSE, and ALMA increasingly constrain feedback at ISM scales, while hydrodynamical simulations like TNG provide information about the CGM’s larger-scale energy budget and dynamics. Combining these two calibration sources would produce a SAM that is anchored in both empirical constraints and the physical insight available from hydrodynamical simulations.

### 6.3 Comparison to Similar Work

Previous attempts to align the predictions of SAMs with hydrodynamical simulations have varied in methodology, scope, and success. Here, we compare our approach and results with those of [Stringer et al. \(2010\)](#), [Neistein et al. \(2012\)](#) and [Mitchell & Schaye \(2022\)](#), highlighting key differences and similarities, and the relative accuracy of each approach.

Early efforts to compare the results from SAMs and hydrodynamical simulations (e.g., [Helly et al. \(2003\)](#), [Benson et al. \(2001\)](#), and [Yoshida et al. \(2002\)](#)) found that gas cooling rates from the hot halo into the ISM could be brought into good agreement when SAMs were modestly adjusted. For example, [Helly et al. 2003](#) showed that reducing the core radius of the gas density profile and increasing cooling times in low-mass halos allowed semi-analytic models to match SPH simulations within 25–50% across halo masses. [Stringer et al. \(2010\)](#) expanded this approach by attempting to reproduce the entire formation history of a single disk galaxy within the GASOLINE hydrodynamical simulation using a modified version of the GALFORM SAM. To achieve broad agreement with the simulation data, they adjusted parameters related to gas cooling, star formation, and feedback, demonstrating the potential of SAMs to reproduce the detailed evolution of individual galaxies. However, the focus on a single object left uncertainties about the general applicability of their findings.

[Neistein et al. \(2012\)](#) (hereafter N12) expanded on this to a wider range of galaxies by extracting efficiencies describing accretion, cooling, star formation, and feedback from the OverWhelmingly Large Simulations (OWLS) and applying them within the SAM presented in [Neistein & Weinmann \(2010\)](#) to reproduce OWLS’s results. By tuning these efficiencies as functions of halo mass and redshift, they reproduced various global galaxy and halo properties up to  $z = 3$  within  $\sim 0.1 - 0.2$  dex. In terms of approach and scope, N12 is most similar to our work here with the TNG SAM.

However, there are both similarities and significant differences in our implementations. Similar to the TNG SAM, N12 also utilizes a set of efficiencies to describe the rate at which gas enters the halo, cools from the hot halo to the galaxy, forms stars, and is ejected by stellar feedback. For instance, the halo gas accretion efficiency parameter  $f_a = \dot{M}_{\text{CGM}}^{\text{in}} / \dot{M}_{\text{halo}}$  in N12 is similar to the TNG SAM’s  $f_{\text{in,CGM}}$ , except  $f_{\text{in,CGM}}$  adjusts for the baryon fraction and includes both first-time infall and gas recycling. Despite this nuance, both models similarly use their respective parameters to match the gas accretion rate found in their hydrodynamical simulations.

Overall, N12 used less detailed efficiencies to calibrate their SAM against OWLS, reflecting their primary focus on reproducing the hot gas mass, cold gas mass, stellar mass, and total galaxy mass. Their approach yielded very good agreement for these quantities, reporting a standard deviation ranging from 0.1 - 0.2 dex for each

mass reservoir. However, SFR deviations were larger, with a standard deviation of 0.5 dex. The TNG SAM, despite aiming to reproduce a wider range of global galaxy and halo properties and extending to higher redshifts, achieves comparable success within 0.1-0.3 dex for almost all galaxy and halo scale properties explored, including the SFR.

Here, we note a broader issue that affects most comparisons between SAMs and hydrodynamical simulations: key galaxy- and halo-scale properties are not always defined in a consistent way. In the TNG SAM, we define all galaxy-scale quantities within twice the stellar half-mass radius. Other works have adopted different choices. For example, N12 define the stellar mass in their hydrodynamical simulation as the total mass of all star particles in the subhalo, the cold gas mass as the mass of all gas particles eligible to form stars, and the hot gas mass as all gas particles that do not belong to  $M_{\text{ISM}}$  (though it is unclear if they also remove the contribution from satellite galaxies within the halo). Such differences directly affect derived properties, particularly  $M_{\text{CGM}}$ , since what counts as “hot” depends directly on how “cold” is defined (see Figure A2 for examples).

In the TNG SAM, we aim for consistency by applying a single definition—twice the stellar half-mass radius—for all galaxy-scale measurements. However, this raises an important question: to what extent do the reported accuracies in earlier works, including N12, depend on their specific definitions, and how would their level of agreement change under a definition consistent with ours (or vice versa)?

Despite these uncertainties, even comparisons with inconsistent definitions can provide valuable insights into the baryon cycle. Recently, [Mitchell & Schaye \(2022\)](#) (hereafter MS22) expanded on the N12 framework to emulate the results of the EAGLE simulations and investigate the parameters most critical to reproducing the stellar mass–halo mass (SHM) relation using a gas regulator model, which is a simplified version of a SAM. They found that EAGLE’s SHM relation is primarily shaped by gas ejection via outflows for halos with  $M_{\text{vir}} < 10^{12} M_{\odot}$ , with halo-scale preventative feedback and recycling of ejected gas playing secondary roles. They also found that the redshift evolution of the SHM relation is most sensitive to the efficiencies of first-time gas accretion and ejection by outflows, and is less sensitive to the efficiency of wind recycling, and of gas consumption by star formation. In the TNG SAM, we find similar trends, with Figure 9 panel *v*. showing that the stellar mass is most sensitive to  $\eta_{\text{halo}}$ , and then  $f_{\text{in,CGM}}$ .

MS22 also investigated how star formation and gas flows affect the relationship between halo mass and the masses of both the ISM and CGM. They found that the CGM mass is most sensitive to variations in gas inflows and outflows, particularly those occurring at the halo scale. This finding is echoed in our analysis of the TNG SAM, where we also observed a strong dependence of the CGM mass in TNG on halo-scale gas flows. However, while MS22 found that ISM mass is more sensitive to halo-scale flows than stellar mass, the TNG SAM shows that  $\eta_{\text{halo}}$  has a comparable impact on both  $M_{\text{ISM}}$  and  $M_{\text{star}}$ . MS22 attributed the stronger dependence of ISM mass on halo-scale gas flows to its direct connection with recent accretion, star formation, and outflow activity, noting that in EAGLE, galaxy-scale outflows dominate at high redshifts, while halo-scale outflows are more efficient at low redshifts. In contrast, the TNG SAM directly couples ISM mass to stellar mass via the Kennicutt-Schmidt relation, meaning that any variation in outflow rates—whether from the ISM or CGM—ultimately affects the stellar mass by regulating the gas available for star formation.

A significant difference between the TNG SAM and the above studies is its explicit tracking of metals across different mass reservoirs,

including stars, cold gas, and hot gas, as well as the flows between them. It remains unclear whether the other approaches discussed above could replicate metal content as effectively as the TNG SAM. However, given our success using the metal enrichment factors, it is reasonable to assume that incorporating similar prescriptions in those models could yield comparable improvements.

## 7 SUMMARY AND CONCLUSIONS

In this paper, we introduced the TNG SAM, a modified version of the Santa Cruz semi-analytic model, designed to replicate the complex baryon cycle of galaxies in the IllustrisTNG cosmological hydrodynamical simulations. This model bridges the detailed physical processes captured in hydrodynamical simulations with the computational efficiency of SAMs, offering a powerful tool for studying galaxy evolution. By incorporating TNG’s galaxy formation physics into a simplified framework, the TNG SAM allows for deeper insights into the interplay of baryonic processes that shape galaxies.

Focusing on stellar feedback-dominated systems, we aimed to reproduce the baryon cycle in low- to intermediate-mass dwarf galaxies ( $M_{\text{halo}} \sim 10^{10} - 10^{11} M_{\odot}$ ) and Milky Way-mass galaxies ( $M_{\text{halo}} \sim 10^{12} M_{\odot}$ ) in TNG100. Using measurements of gas flows from a subset of  $\sim 400$  central galaxies as a proxy for the larger TNG100 sample, we updated the SC SAM’s physical prescriptions for halo gas accretion, cooling, stellar feedback, and metal circulation. These updates, implemented as a function of halo mass and redshift, led to more accurate predictions of galaxy-scale properties such as stellar mass, gas content, and metallicity, as well as halo-scale properties like the overall baryon content.

Several key insights about the baryon cycle in TNG emerged from this work, with important implications to keep in mind when building the next generation of SAMs:

(i) *The classical hot mode/cold mode cooling model used in nearly all SAMs does not provide a good description of cooling and halo gas accretion in TNG.* In TNG and FIRE, cooling times often exceed dynamical timescales, showing that the simple cold-mode vs. hot-mode framework commonly used in SAMs does not adequately represent the cooling process in state-of-the-art cosmological hydrodynamic simulations. The TNG SAM’s revised cooling model, calibrated to the cooling times in TNG, results in improved predictions of gas accretion onto galaxies and the mass of the hot halo. This suggests that SAMs need to move beyond the traditional “cold mode” vs. “hot mode” dichotomy and instead adopt models that better reflect the full range of gas cooling timescales seen in hydrodynamical simulations (Section 6.1.2, Figure 11).

(ii) *Directly calibrating the star-formation efficiency yields the most accurate global ISM predictions.* In the TNG SAM, the star formation rate is tied to the star-formation efficiency measured from the full TNG100 sample across halo mass and redshift. This reproduces TNG’s median stellar masses, cold gas masses, and star formation rates to within  $\sim 30\%$  across time. Attempts to apply a global Kennicutt–Schmidt law instead lead to large discrepancies because the fiducial SAM’s predicted galaxy sizes do not follow TNG’s size evolution (Section 6.1.3, Figure 9 panels iii.–v.). Until disk sizes can be modeled more reliably, calibrating against the star formation efficiency remains the most robust approach for matching TNG’s global ISM relations.

(iii) *Modeling outflows at the galaxy- and halo-scale is essential for accurately predicting the gas content in galaxies.* While galaxy-scale outflows are a well-known driver of galaxy evolution, the TNG

SAM highlights that halo-scale outflows are equally critical for predicting the distribution of baryons. Incorporating a two-step outflow model—where gas is first ejected from the ISM into the CGM, and then from the CGM into the IGM—provides a more realistic depiction of baryon cycling than traditional SAMs, which typically eject gas directly into an external reservoir, and improves the accuracy of stellar mass, hot gas mass and baryon fraction predictions. (Section 6.1.4, Figures 9 panels i. & ii. and 10 panels ii. & vi.).

(iv) *Accounting for metal-enriched and metal-depleted flows between the ISM and CGM leads to more realistic predictions of galaxy and CGM metallicities.* The TNG SAM’s use of metallicity-weighted mass-loading factors significantly improves its ability to match the metallicity evolution found in TNG. This approach captures the fact that outflows and inflows can be systematically metal-enhanced or metal-depleted relative to their source reservoirs—a feature observed in TNG and other high-resolution simulations. Explicitly modeling these metal enrichment factors is critical for accurately tracking the redistribution of metals between the ISM, CGM, and IGM. (Section 6.1.5, Figures 9 panels vi. - viii. and 10 panels iii., iv., vii. & viii.).

## ACKNOWLEDGEMENTS

OO acknowledges support from the National Science Foundation Graduate Research Fellowship.

*Software:* ASTROPY Robitaille et al. (2013); Collaboration et al. (2018, 2022), IPYTHON (Pérez & Granger 2007), MATPLOTLIB (Caswell et al. 2022), NUMPY (van der Walt et al. 2011), SCIPY (Virtanen et al. 2020)

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author

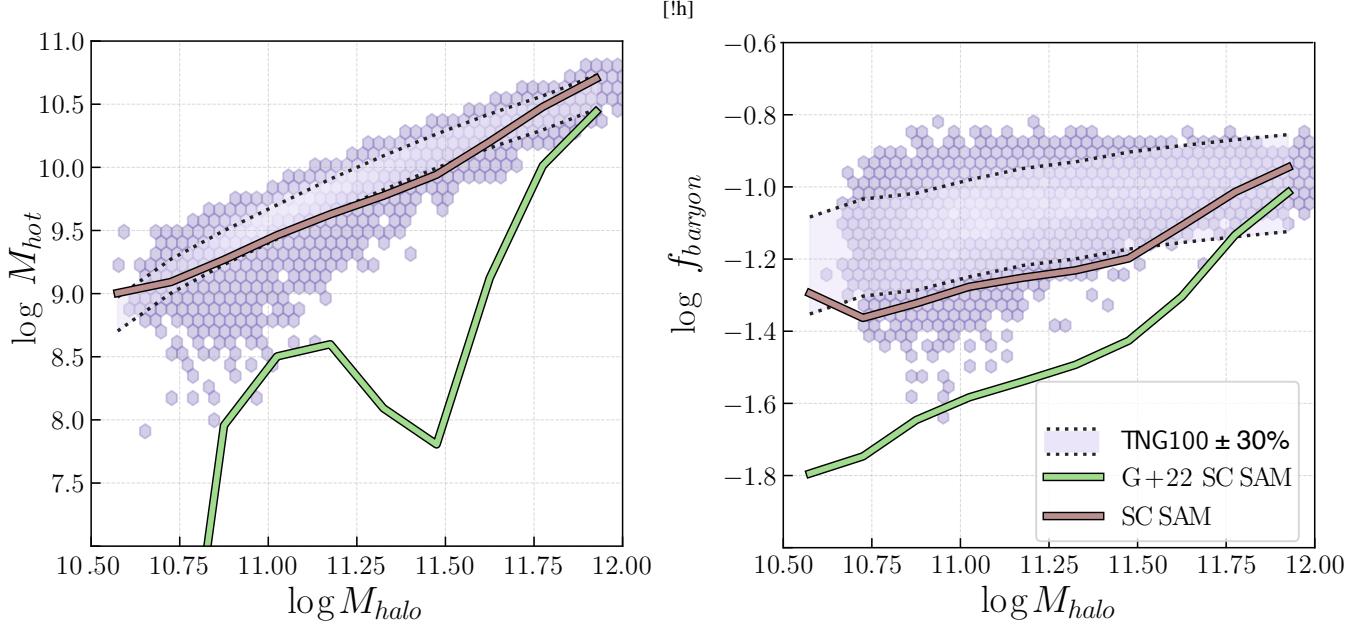
## APPENDIX A: BUILDING BLOCKS OF THE TNG SAM

### A1 Comparison to the published SC SAM

Pandya et al. (2020) showed that the CGM mass in the SC SAM was grossly underpredicted relative to the FIRE simulations, and Gabrielpillai et al. (2022) showed that the same discrepancy appears compared with TNG, as shown in Figure A1. This motivated the change to the cooling model described in Section 2.2.5, which is implemented in the SC SAM used throughout this work. This change had the most substantial impact on the halo-scale properties. The updated cooling model suggested by Pandya et al. (2020) already significantly improved the prediction of the hot halo gas mass, as shown in Figure A1. Naturally, this correction also led to a substantial improvement in the SAM’s prediction of the baryon fraction, illustrated in the lower panel of Figure A1. Although these changes drastically affected the halo-scale gas properties, the impact on the ISM-scale gas properties and overall metal content was minimal, with deviations from the Santa Cruz SAM published in Gabrielpillai et al. (2022) (G+22) reaching a maximum of 0.5 dex.

### A2 Analytic Scaling relations

We alternatively calibrated the SC SAM using analytic scaling relations as a function of halo mass and redshift using the following sophisticated functional form:



**Figure A1.** Comparison of the Santa Cruz SAM used in this work (solid brown line) to TNG100 (purple shaded region) and the SC SAM from [Gabrielpillai et al. 2022](#) (green line) for hot halo gas mass ( $M_{\text{CGM}}$ , top) and baryon fraction ( $f_{\text{baryon}}$ , bottom) as a function of halo mass. The modified cooling model described in Section 2.2.5 substantially improves agreement with TNG100 for halo scale properties.

Model Update	Parameter	Description	Coefficients (a, b, c, d, $\alpha_0$ , $\alpha_z$ )	Comments
Gas Cooling	$f_{\text{in,CGM}}$		1.15, 0.0088, 10.2189, 0.0454, -3.5025, -1.0791	
...	$t_{\text{cool}}$		1.15, 9.4401, 11.5681, 359.5587, -0.0275, -7.4258	
Star Formation	$R_{\text{disk,TNG}}/R_{\text{disk,SAM}}$		1.0, 0.0009, 9.4550, -0.0028, 9.2402, 4.2137	
Stellar Feedback	$\eta_{\text{launch}}/\eta_{\text{ISM}}$		1.2, 6.7252, 11.6654, 4.6427, -0.1362, -3.7849	
...	$\eta_{\text{halo}}$		0.85, -1.6941, 10.8106, 1.7008, -0.7913, -3.1102	
Metal Circulation	$\zeta_{\text{in,halo}}$		1.0, -5.0310, 11.3533, 7.8483, 0.1622, -4.3837	
...	$\zeta_{\text{out,halo}}$		1.0, 4.0473, 11.3201, -0.0693, -0.1187, -2.9203	
...	$\zeta_{\text{in,ISM}}$		1.0, 1.1155, 10.0801, -1.4657, 0.0161, -2.9189	
...	$\zeta_{\text{out,ISM}}$		1.0, 48.4092, -62.9615, 5.9549, 0.0501, -18.1282	$z \leq 1$
...	...		1.0, 7.6844, 13.7629, 1.4179, -4.2491, -0.8384	$z > 1$

**Table A1.** Parameters calibrated to match TNG and the coefficients used in Equation A1. (1) Parameter; (2) SAM Label; (3) TNG Label; (4) comment.

$$\log(f(M_{\text{halo}}, z, b, c, d, \alpha_0, \alpha_z)) = \\ -a(\arctan(b \cdot [\log(M_{\text{halo}}) - c]) + d) \cdot [\alpha_0 + e^{\alpha_z}(1+z)] \quad (\text{A1})$$

Using these scaling relations also provided agreements with TNG within the same range as the linearly interpolated values. All coefficients used in Equation A1 are described in Table A1.

### A3 The Importance of Consistent Radial Definitions

To facilitate meaningful comparisons between the SAM and TNG100, we adopted the definitions listed in Table A2. While these definitions do not yield perfect one-to-one correspondence—owing to inherent structural and methodological differences between SAMs

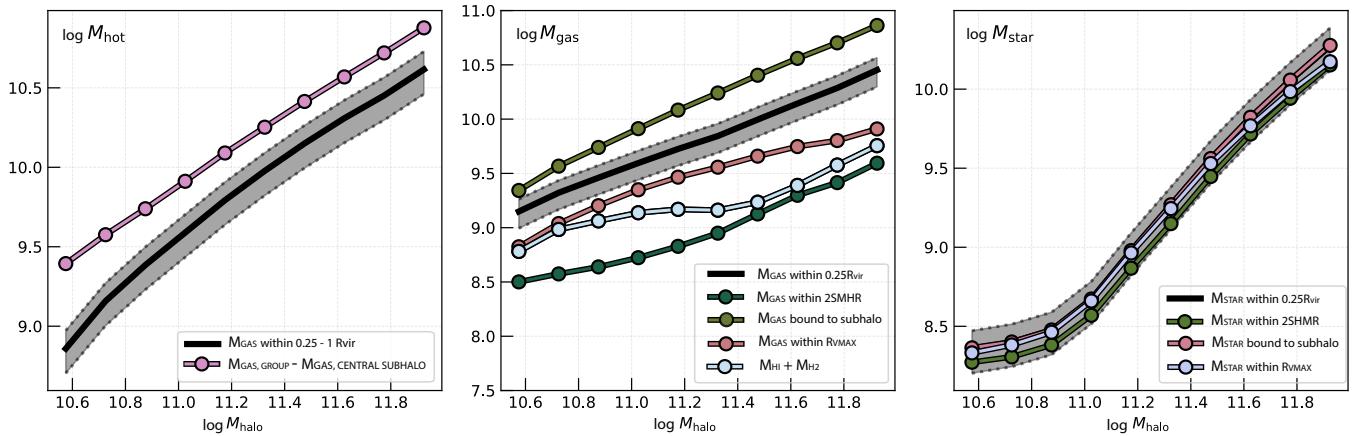
and hydrodynamical simulations—they help reduce systematic discrepancies and clarify the origin of any remaining mismatches.

While developing the TNG SAM, we found that the definitions of key mass reservoirs—such as  $M_{\text{CGM}}$ ,  $M_{\text{ISM}}$ , and  $M_{\star}$ —vary significantly in the literature, particularly when comparing the outputs of SAMs, hydrodynamical simulations, and observations. Figure A2 shows that even within a single simulation, the choice of spatial boundaries can lead to systematic differences of up to 0.5 dex in the quantities analyzed.

For instance,  $M_{\text{CGM}}$  may be defined as all gas between  $0.25$ – $1.0 r_{\text{vir}}$  (black line), or alternatively as the total gas bound to the halo minus the gas in the central subhalo (purple line)—the latter yields higher values of  $M_{\text{CGM}}$  by up to 0.4 dex across halo mass. Similarly,  $M_{\text{ISM}}$  can be measured within  $0.25 r_{\text{vir}}$ , within twice the stellar half-mass radius ( $2R_{\text{SHMR}}$ ), or by summing  $M_{\text{HI}} + M_{\text{H}_2}$ —with the latter two

Scale	Parameter	SAM Label	TNG Label	Description
<i>Halo</i>	$M_{\text{halo}}$	GalpropMvir	Group_M_TopHat_200['SubhaloGrNr']	Total virial mass
	$M_{\text{DM}}$	GalpropMvir ( $1 - f_{\text{baryon}}$ )	SubhaloMassType Part 1 (DM)	Total dark matter mass
	$M_{\text{CGM}}$	HalopropMhot	$\sum_i^n m_{\text{gas}}^i$ for $0.25r_{\text{vir}} < r < r_{\text{vir}}$	Gas mass of the circumgalactic halo
<i>Galaxy</i>	$M_*$	GalpropMstar	$\sum_i^n m_{\text{star}}^i$ for $r < 0.25r_{\text{vir}}$	Stellar mass
	$M_{\text{ISM}}$	GalpropMcold	$\sum_i^n m_{\text{gas}}^i$ for $r < 0.25r_{\text{vir}}$	Gas mass of the interstellar medium
	SFR	GalpropSfr	$\sum_i^n \dot{m}_{\text{star}}^i$ for $r < 0.25r_{\text{vir}}$	Star formation rate
	$< Z_* >$	GalpropZstar	$\frac{1}{n} \sum_i^n Z_{\text{star}}^i$ for $r < 0.25r_{\text{vir}}$	Average stellar metallicity
	$< Z_{\text{ISM}} >$	GalpropZcold	$\frac{1}{n} \sum_i^n Z_{\text{gas}}^i$ for $r < 0.25r_{\text{vir}}$	Average cold gas metallicity
	$< Z_{\text{CGM}} >$	HalopropZhhot	$\frac{1}{n} \sum_i^n Z_{\text{gas}}^i$ for $0.25r_{\text{vir}} < r < r_{\text{vir}}$	Average hot halo gas metallicity

**Table A2.** Correspondence between the parameters studied in this paper, their fields in the SAM, and their corresponding fields in the IllustrisTNG simulations. A description of each parameter is also provided.



**Figure A2.** Comparison of  $M_{\text{CGM}}$ ,  $M_{\text{ISM}}$ , and  $M_{\text{star}}$  using different definitions in TNG at  $z = 0$ . The black solid line shows the definitions adopted in this paper ( $M_{\text{CGM}}$  within  $0.25 - 1 r_{\text{vir}}$ ,  $M_{\text{ISM}}$  within  $0.25 r_{\text{vir}}$ , and  $M_{\text{star}}$  within  $0.25 r_{\text{vir}}$ ), with shaded regions indicating deviations within 30%. Colored lines represent alternate definitions commonly used in the literature, showing significant variations for  $M_{\text{CGM}}$  and  $M_{\text{ISM}}$  (up to 50%) and smaller deviations for  $M_{\text{star}}$  (within 30%). [add galaxy size comparisons](#)

being common choices in the literature. Each of these definitions yields notably different trends with halo mass.  $M_{\text{star}}$  is more stable across the options tested (e.g., within  $0.25r_{\text{vir}}$ ,  $2R_{\text{SHMR}}$ ,  $R_{\text{Vmax}}$ , or total subhalo mass), but still shows differences of up to 0.2 dex at the low and high mass ends.

These variations demonstrate that inconsistent definitions can introduce substantial biases, especially for gas reservoirs, where the inferred mass strongly depends on aperture choice or particle binding criteria. This issue is particularly relevant when comparing SAMs, hydrodynamical simulations, and observational datasets, all of which often adopt different conventions for measuring the same physical quantities. Without careful alignment, such comparisons risk showing artificial agreement or disagreement—not due to real physical differences, but as an artifact of inconsistent methodology. To avoid this ambiguity, we adopt consistent radial definitions throughout this work (see Section 4.1.1) and explicitly define all quantities used in our comparisons.

#### A4 How Few Galaxies Are Enough? Calibration with 100 Galaxies

To evaluate how sensitive the TNG SAM’s performance is to the size of the calibration sample, we constructed a version of the model using only  $\sim 100$  randomly selected galaxies per redshift, spanning the halo mass range  $10^{10} - 10^{12} M_{\odot}$ . This reduced-sample version allows us to explore how few galaxies are needed to maintain reliable predictions without incurring the significant computational cost of extracting flow measurements for thousands of objects.

Even with this significantly smaller calibration set, we find the TNG SAM continues to reproduce most galaxy-scale quantities within  $\sim 40\%$  of TNG values. At  $z = 0$ , for example, gas inflow and outflow rates into the ISM are recovered to within 14% and 12%, respectively, and remain within 25% at higher redshifts. Global properties such as stellar mass, cold gas mass, and SFR are typically accurate to within 20–30%, though stellar mass deviates by as much as 66% at  $z = 6$ .

In contrast, halo-scale predictions show a more substantial decline in accuracy with redshift. At  $z = 6$ , halo-scale inflow and outflow rates deviate by 55% and 41%, respectively, and metal flow rates show similar scatter. The hot halo gas mass is recovered to within 50–90%

depending on redshift, while baryon fraction deviations reach up to 55% at  $z = 0$ .

## REFERENCES

- Anderson L. D., Bania T. M., Balser D. S., Cunningham V., Wenger T. V., Johnstone B. M., Armentrout W. P., 2014, *ApJS*, 212, 1
- Arrigoni M., Trager S. C., Somerville R. S., Gibson B. K., 2010, *MNRAS*, 402, 173
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2012, *ApJ*, 762, 109
- Benson A. J., 2012, *New Astronomy*, 17, 175
- Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001, *MNRAS*, 320, 261
- Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, *MNRAS*, 436, 697
- Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, 136, 2846
- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *APJ*, 379, 440
- Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, *A&A*, 564, A66
- Bower R. G., 1991, *MNRAS*, 248, 332
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Calette A. R., Avila-Reese V., Rodríguez-Puebla A., Hernández-Toledo H., Papastergis E., 2018, *Revista Mexicana de Astronomía y Astrofísica*, 54, 443
- Carr C., Bryan G. L., Fielding D. B., Pandya V., Somerville R. S., 2023, *ApJ*, 949, 21
- Caswell T. A., et al., 2022, [10.5281/zenodo.6982547](https://doi.org/10.5281/zenodo.6982547)
- Cole S., 1991, *APJ*, 367, 45
- Collaboration T. A., et al., 2018, *AJ*, 156, 123
- Collaboration T. A., et al., 2022, *ApJ*, 935, 167
- Collacchioni F., Cora S. A., Lagos C. D. P., Vega-Martínez C. A., 2018, *MNRAS*, 481, 954
- Croton D. J., et al., 2006, *MNRAS*, 365, 11
- Croton D. J., et al., 2016, *ApJS*, 222, 22
- Côté B., Silvia D. W., O'Shea B. W., Smith B., Wise J. H., 2018, *ApJ*, 859, 67
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- Diemer B., et al., 2018, *ApJS*, 238, 33
- Donnari M., et al., 2019, *MNRAS*, 485, 4817
- Gabrielpillai A., Somerville R. S., Genel S., Rodriguez-Gomez V., Pandya V., Yung L. Y. A., Hernquist L., 2022, *MNRAS*, 517, 6091
- Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, 362, 41
- Genel S., et al., 2018, *MNRAS*, 474, 3976
- Gnedin N. Y., Kravtsov A. V., 2011, *ApJ*, 728, 88
- Guo Q., et al., 2011, *MNRAS*, 413, 101
- Guo Q., et al., 2016, *MNRAS*, 461, 3457
- Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., Pearce F. R., 2003, *MNRAS*, 338, 913
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, *MNRAS*, 431, 3373
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663
- Hirschmann M., Naab T., Somerville R. S., Burkert A., Oser L., 2012, *MNRAS*, 419, 3200
- Hirschmann M., De Lucia G., Fontanot F., 2016, *MNRAS*, 461, 1760
- Hopkins P. F., et al., 2018, *MNRAS*, 480, 800
- Karmakar T., Genel S., Somerville R. S., 2023, *MNRAS*, 520, 1630
- Kennicutt Jr. R. C., 1998, *ApJ*, 498, 541
- Keres D., Yun M. S., Young J. S., 2003, *ApJ*, 582, 659
- Kobayashi C., Springel V., White S. D. M., 2007, *MNRAS*, 376, 1465
- Lacey C., Cole S., 1993, *MNRAS*, 262, 627
- Lacey C., Silk J., 1991, *APJ*, 381, 14
- Lacey C. G., et al., 2016, *MNRAS*, 462, 3854
- Lagos C. d. P., Lacey C. G., Baugh C. M., 2013, *MNRAS*, 436, 1787
- Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, *MNRAS*, 481, 3573
- Lu Y., Kereš D., Katz N., Mo H. J., Fardal M., Weinberg M. D., 2011, *MNRAS*, 416, 660
- Marinacci F., et al., 2018, *MNRAS*, 480, 5113
- McConnell N. J., Ma C.-P., 2013, *ApJ*, 764, 184
- Mitchell P. D., Schaye J., 2022, *MNRAS*, 511, 2948
- Mitchell P. D., et al., 2018, *MNRAS*, 474, 492
- Mitchell P. D., Schaye J., Bower R. G., Crain R. A., 2020, *MNRAS*, 494, 3971
- Monaco P., Fontanot F., Taffoni G., 2007, *MNRAS*, 375, 1189
- Monaco P., Benson A. J., De Lucia G., Fontanot F., Borgani S., Boylan-Kolchin M., 2014, *MNRAS*, 441, 2058
- Naiman J. P., et al., 2018, *MNRAS*, 477, 1206
- Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, *MNRAS*, 421, 3127
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Neistein E., Weinmann S. M., 2010, *MNRAS*, 405, 2717
- Neistein E., Khochfar S., Dalla Vecchia C., Schaye J., 2012, *MNRAS*, 421, 3579
- Nelson D., et al., 2018, *MNRAS*, 475, 624
- Nelson D., et al., 2019, *MNRAS*, 490, 3234
- Obreschkow D., Croton D., Lucia G. D., Khochfar S., Rawlings S., 2009, *ApJ*, 698, 1467
- Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
- Oren Y., Pandya V., Somerville R. S., Genel S., Omoruyi O., Sternberg A., 2025, [10.48550/arXiv.2510.23343](https://arxiv.org/abs/2510.23343)
- Pakmor R., Springel V., 2013, *MNRAS*, 432, 176
- Pakmor R., Bauer A., Springel V., 2011, *MNRAS*, 418, 1392
- Pandya V., 2021, PhD thesis, UC Santa Cruz, <https://escholarship.org/uc/item/9xc1v7c9>
- Pandya V., et al., 2020, *ApJ*, 905, 4
- Pandya V., et al., 2021, *MNRAS*, 508, 2979
- Pandya V., et al., 2023, *ApJ*, 956, 118
- Peeples M. S., Shankar F., 2011, *MNRAS*, 417, 2962
- Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, *ApJ*, 786, 54
- Pillepich A., et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A., et al., 2018b, *MNRAS*, 475, 648
- Pillepich A., et al., 2019, *MNRAS*, 490, 3196
- Planck Collaboration et al., 2016, *A&A*, 594, A13
- Popping G., Somerville R. S., Trager S. C., 2014, *MNRAS*, 442, 2398
- Porter L. A., Somerville R. S., Primack J. R., Johansson P. H., 2014, *MNRAS*, 444, 942
- Pérez F., Granger B. E., 2007, *Computing in Science & Engineering*, 9, 21
- Robitaille T. P., et al., 2013, *A&A*, 558, A33
- Rodriguez-Gomez V., et al., 2015, *MNRAS*, 449, 49
- Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, *MNRAS*, 470, 651
- Saro A., De Lucia G., Borgani S., Dolag K., 2010, *MNRAS*, 406, 729
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Somerville R. S., Primack J. R., 1999, *MNRAS*, 310, 1087
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008a, *MNRAS*, 391, 481
- Somerville R. S., et al., 2008b, *ApJ*, 672, 776
- Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, *MNRAS*, 423, 1992
- Somerville R. S., Popping G., Trager S. C., 2015, *MNRAS*, 453, 4337
- Somerville R. S., Gabrielpillai A., Hadzhiyska B., Genel S., 2025, [10.48550/arXiv.2502.03679](https://arxiv.org/abs/2502.03679)
- 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
- Stringer M. J., Brooks A. M., Benson A. J., Governato F., 2010, *MNRAS*, 407, 632
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253

- Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L.,  
 2014, [MNRAS](#), 438, 1985
- Torrey P., et al., 2019, [MNRAS](#)
- Virtanen P., et al., 2020, [Nat Methods](#), 17, 261
- Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L.,  
 2013, [MNRAS](#), 436, 3031
- Voit G. M., Pandya V., Fielding D. B., Bryan G. L., Carr C., Donahue M.,  
 Oppenheimer B. D., Somerville R. S., 2024a, [ApJ](#), 976, 150
- Voit G. M., Carr C., Fielding D. B., Pandya V., Bryan G. L., Donahue M.,  
 Oppenheimer B. D., Somerville R. S., 2024b, [ApJ](#), 976, 151
- Weinberger R., et al., 2017, [MNRAS](#), 465, 3291
- White S. D. M., Frenk C. S., 1991, [ApJ](#), 379, 52
- White S. D. M., Rees M. J., 1978, [MNRAS](#), 183, 341
- Wise J. H., Turk M. J., Norman M. L., Abel T., 2011, [ApJ](#), 745, 50
- Wright R. J., Lagos C. d. P., Power C., Mitchell P. D., 2020, [MNRAS](#), 498,  
 1668
- Wright R. J., Somerville R. S., Lagos C. d. P., Schaller M., Davé R., Anglés-  
 Alcázar D., Genel S., 2024, [MNRAS](#), 532, 3417
- Yates R. M., Kauffmann G., Guo Q., 2012, [MNRAS](#), 422, 215
- Yoshida N., Stoehr F., Springel V., White S. D. M., 2002, [MNRAS](#), 335, 762
- Yung L. Y. A., et al., 2023, [MNRAS](#), 519, 1578
- Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz  
 M. J., 2013, [ApJL](#), 771, L19
- van der Walt S., Colbert S. C., Varoquaux G., 2011, [Computing in Science &  
 Engineering](#), 13, 22

This paper has been typeset from a [TeX/LaTeX](#) file prepared by the author.