

# 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 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. hot halo gas mass) 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 observational 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 star formation. This cycle of gas accretion, star formation, feedback,

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

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 approximate 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 hydrodynamical 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—an observable not widely available for Milky Way and lower mass halos to calibrate or validate SAMs. 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, inspired by the Henriques et al. (2013) SAM. This change allowed TNG to better match observed stellar mass functions and luminosity functions across redshift, showcasing the bidirectional synergy between SAMs and hydrodynamical simulations.

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 dominate the baryon cycle. By calibrating the SAM with measurements of galaxy- and halo-scale inflow and outflow rates for a subset of TNG galaxies extracted by Oren et al. (2025), 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 describes 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 TNG, 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 cubic volumes of roughly 50, 100, and 300 Mpc side length, 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 Santa Cruz 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.” 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 details.

	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	TNG100: $f_{\text{CGM}}^{\text{in}}$ ; $t_{\text{cool}}$ ; $\tau_*$ ; $\eta_{\text{ISM}}$ ; $\eta_{\text{CGM}}$ ; $\zeta^{\text{in/out}}_{\text{ISM}}$ ; $\zeta^{\text{in/out}}_{\text{halo}}$

**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, chemical enrichment, stellar feedback, metal circulation, and calibration strategies.

### 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 accretion  $\dot{m}_{\text{acc}}$  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<sub>2</sub> 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  $Z_{\text{cold}}$  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 Santa Cruz 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), Peeples 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<sub>2</sub> 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 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 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. Additionally, 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, the model represents an upper bound on the actual accretion rate.

### 3 COMPARING 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 (DMO) simulation run with the same initial conditions as the Full Physics (FP) TNG simulation. 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.

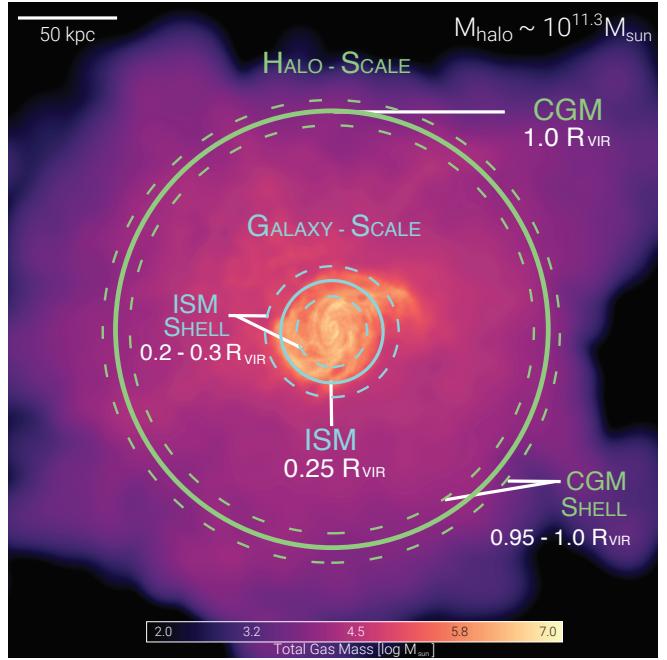
Although 99% of central galaxies have bijective matches, a direct comparison between the halo mass of the central galaxies between the two catalogs reveals slight offsets up to 20% for low mass halos (see Figure 2 in Gabrielpillai et al. 2022). These discrepancies are likely primarily influenced by baryonic 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) the galaxy-scale (ISM), and
- (ii) the halo-scale (CGM),

as illustrated in Figure 1. Properties at the “galaxy scale” describe the stellar mass ( $M_{\star}$ ), cold gas mass ( $M_{\text{cold}}$ ), their respective metallicities ( $Z_{\star}$  and  $Z_{\text{cold}}$ ), 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 ( $M_{\text{hot}}$ ), and the metallicity of the gas in the hot halo ( $Z_{\text{hot}}$ ). 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  $r_{\text{vir}}$



**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  $0.05$ – $0.15r_{\text{vir}}$  (galaxy scale) and  $0.95$ – $1.0r_{\text{vir}}$  (halo scale).

such that the mean density of dark matter haloes is  $\Delta_c$  times the critical density at a given redshift.

Although the SAM naturally segments and outputs all quantities defined above, mapping these quantities onto TNG requires careful consideration. For the galaxy’s cold gas, we compute  $M_{\text{cold}}$  and  $Z_{\text{cold}}$  in two complementary ways. For comparison with the SC SAM, we define  $M_{\text{cold}}$  as the sum of the combined neutral atomic hydrogen ( $\text{H I}$ ) and molecular hydrogen ( $\text{H}_2$ ) components to account for differences in treatment between the SC SAM and TNG’s star formation models. The SC SAM partitions gas into neutral, ionized, and molecular components following Gnedin & Kravtsov (2011). Since TNG does not explicitly resolve “cold” gas, we measure  $\text{H I+H}_2$  using post-processed catalogs from Diemer et al. (2018, 2019) and, as in Gabrielpillai et al. (2022), adopt the Gnedin & Kravtsov (2011) volumetric method to ensure consistency with the SC SAM’s gas-partitioning treatment. For comparison with the TNG SAM (described in Section 4), we instead measure  $M_{\text{cold}}$  and  $Z_{\text{cold}}$  using all gas particles within twice the stellar half-mass radius.

All remaining galaxy- and halo-scale properties are defined consistently when comparing the SC SAM and the TNG SAM to TNG. For stellar properties,  $M_{\star}$  and SFR are defined as the total mass of bound star particles and the total SFR of bound gas cells in the subhalo, respectively. The stellar metallicity  $Z_{\star}$  is computed from the metallicities of all bound star particles. For halo-scale properties like  $M_{\text{hot}}$  and  $Z_{\text{hot}}$ , we sum all gas in the host group minus the contribution from gas lying within twice the stellar half-mass radius of satellite group members. The corresponding field names used in both the SAM and TNG outputs can be found in the Appendix’s Table A2.

To enable reliable comparison between TNG and the SAM, we restrict the bulk of our analysis to well-resolved TNG halos. Follow-

ing 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 DMO, 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 ( $M_{\text{star}} \geq 1.4 \times 10^8 M_\odot$ ), and 100 gas 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 galaxy 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, ranging from  $M_{\text{vir}} = 10^{10} M_\odot$  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  $0.05 r_{\text{vir}}$  to  $2 r_{\text{vir}}$ . 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 6.2.3).

### 3.3 Baseline Comparison Between TNG and the SC SAM

With bijective matches between central halos in TNG and the SC SAM established in Section 3.1, we now assess how well the SC

SAM reproduces the galaxy–halo relationships found in the full-physics TNG model. [Gabrielpillai et al. \(2022\)](#) first performed this statistical comparison between the SC SAM and TNG at  $z = 0$  for a subset of key global properties, namely stellar mass, cold gas mass, star formation rate, and hot gas mass. Building upon their initial work, here we extend the comparison to cover a wider redshift range from  $z = 0 - 6$  and a broader set of global quantities, now including the metallicities of the stars, cold gas, and hot halo gas.

#### 3.3.1 Galaxy-Scale

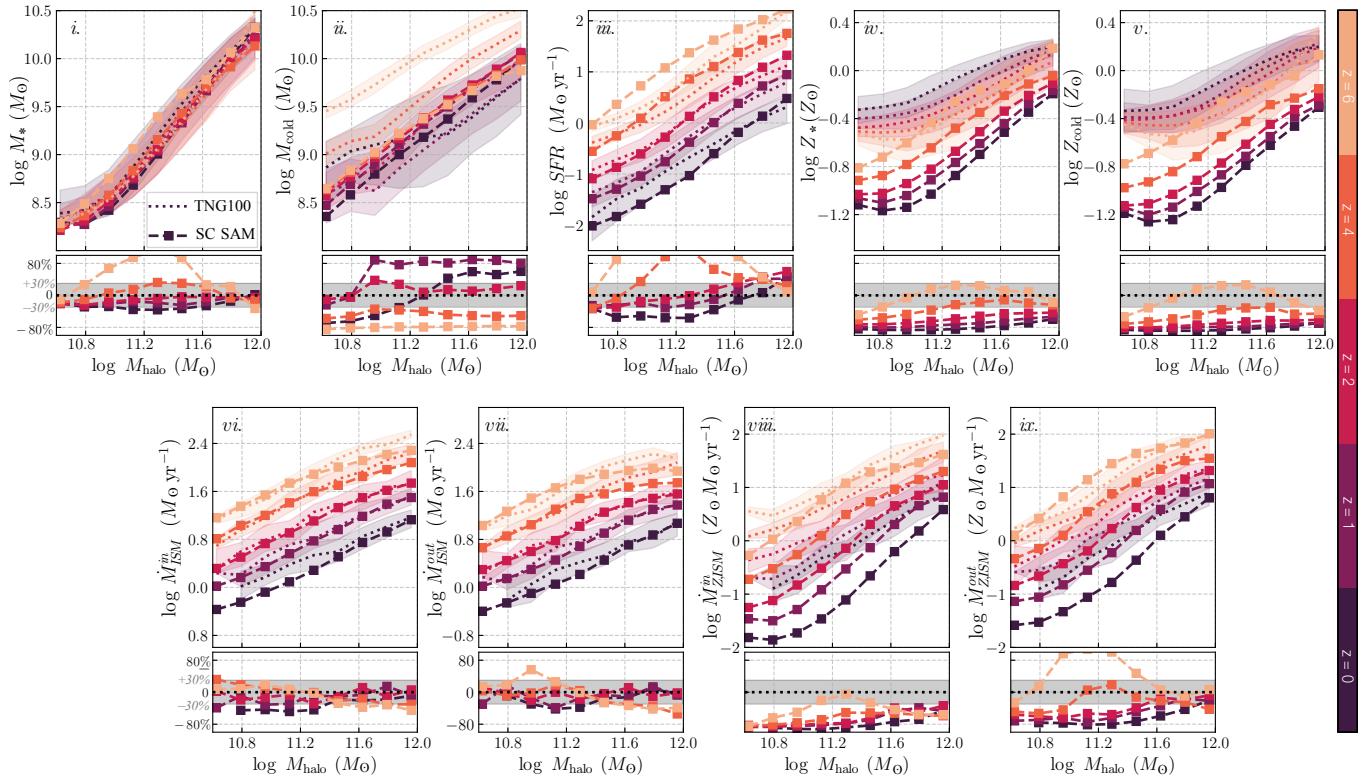
In the top row of Figure 2, we compare TNG100 (dotted lines) and the SC SAM (dashed squares) for the evolution of five galaxy-scale properties as a function of halo mass: stellar mass, cold gas mass, star formation rate, stellar metallicity and cold gas metallicity. Median trends are shown across multiple redshifts from  $z = 0$  to  $z = 6$ , with residual panels beneath each relation illustrating the fractional difference between the SC SAM and TNG.

Starting with the stellar mass - halo mass relation (panel *i*), the SC SAM and TNG predictions generally agree within  $\sim 30\%$  across the halo mass and redshift range shown. The SC SAM tends to predict up to  $\sim 40\%$  lower stellar masses than TNG at low redshift ( $z < 4$ ) and more than  $\sim 100\%$  higher masses at high redshift ( $z = 6$ ). For the cold gas mass (panel *ii*), the agreement between the SC SAM and TNG100 is considerably weaker. Across halo mass and redshift, the SC SAM typically differs from TNG by more than  $\pm 30\%$ , with deviations mostly spanning  $\pm 80\%$ . At low redshift ( $z < 4$ ), the SC SAM tends to underpredict  $M_{\text{cold}}$  by up to  $\sim 100\%$ , whereas at high redshift ( $z \geq 4$ ) it increasingly overpredicts the cold gas content, again by as much as  $\sim 100\%$ . A similar trend with redshift appears for the star formation rate (panel *iii*): the SC SAM predicts up to  $\sim 40\%$  lower rates of star formation than TNG at low redshift ( $z < 4$ ), and rates exceeding  $\sim 100\%$  at high redshift ( $z = 6$ ).

Although the SAM and TNG’s stellar and star formation rate predictions broadly track one another slope-wise across redshift, the divergence between the two models becomes much clearer for the metal populations. At  $z < 4$ , the SC SAM predicts stellar and gas-phase metallicities that are nearly  $100\%$  lower than those in TNG (panels *iv.–v.*). When  $z \sim 4$ , where the SC SAM yields begins to predict higher stellar masses, the two models agree better, but at  $z = 6$ , the SC SAM predicts higher metallicities than TNG, reflecting the SC SAM’s significantly larger stellar masses.

In the bottom row of Figure 2, we compare galaxy-scale baryon flow rates measured in the YO25 TNG subsample (dotted lines) with those predicted by the SC SAM (dashed squares). Considering first the inflow of gas from the CGM into the ISM ( $\dot{M}_{\text{ISM}}^{\text{in}}$ , panel *vi*), the SC SAM generally predicts rates within  $\sim 30\%$  of TNG, though at  $z < 4$  it tends to predict lower inflow rates, with discrepancies reaching up to  $\sim 60\%$ . For the outflow of gas from the ISM driven by stellar feedback, the SC SAM and TNG generally agree within  $30\%$ , with deviations reaching up to  $70\%$  at high redshift ( $z \geq 4$ ) most closely in higher-mass halos ( $\log M_{\text{halo}} > 11$ ).

In panels *viii.* and *ix.*, we compare the rates of metal inflow into the ISM  $\dot{M}_{\text{Z,ISM}}^{\text{in}}$  and metal outflow from the ISM  $\dot{M}_{\text{Z,ISM}}^{\text{out}}$  between the SC SAM and TNG. As expected from the differences in the global metal distributions, the underlying metal flow rates also differ: the SC SAM predicts rates mostly  $\sim 80\%$  lower than TNG across redshift. The agreement only improves at higher redshift ( $z \geq 4$ ) for the metal outflow rate, where the SAM’s predictions either agree with TNG’s within  $30\%$  or exceed them by more than  $100\%$ .



**Figure 2.** Comparison of global galaxy-scale properties (panels *i.*–*iv.*) and flow rates (panels *v.*–*ix.*) between TNG100 (dotted lines) and the Santa Cruz SAM (dashed squares). Each panel shows median trends as a function of halo mass, color-coded by a representative subset of redshifts ( $z = 0, 1, 2, 4, 6$ ): darker colors correspond to lower redshift and lighter colors to higher redshift, ranging from dark purple to yellow (see color bar to the right). This redshift color scheme is used consistently across all figures unless specified otherwise. Global-property medians are measured from the full resolved TNG100 sample, while flow-rate medians are drawn from the YO25 subsample. Shaded regions show the 16–84th percentiles for TNG100. Unless otherwise noted, lightly shaded regions in all subsequent figures represent the same TNG100 percentile range. Below each main panel, the residual panel shows the fractional difference of the SC SAM relative to TNG100, with a horizontal grey band highlighting  $\pm 30\%$  agreement. Positive residuals indicate that the SC SAM predicts larger values than TNG, while negative residuals indicate smaller values. The SC SAM mostly matches TNG’s predictions within  $\pm 30\%$  for the stellar mass ( $M_{\star}$ ), and the star formation rate (SFR) at low redshift, but deviates substantially outside this range for the cold gas mass ( $M_{\text{cold}}$ ) and the metallicities of the stars ( $Z_{\star}$ ) and cold gas ( $Z_{\text{cold}}$ ). Gas and metal inflow/outflow rates also vary, with metal flow differences often exceeding 80% across redshift.

### 3.3.2 Halo-Scale

In the top row of Figure 3, we compare the evolution of the total hot gas mass and its metallicity across halo mass and redshift. The SC SAM systematically predicts lower values of  $M_{\text{hot}}$  and  $Z_{\text{hot}}$  than TNG by up to 80% across most redshifts. The exception occurs in  $Z_{\text{hot}}$  when  $z = 6$ , where the SC SAM exceeds TNG’s predictions by up to  $\sim 100\%$ , likely reflecting the larger stellar masses the SAM predicts at high redshift (see Figure 2 panel *i.*).

Although the SC SAM predicts that galaxies have a more depleted CGM than TNG, the disagreement is noticeably smaller than reported in [Gabrielpillai et al. \(2022\)](#) (see Appendix Figure A1). This improvement stems from the updated cooling model adopted here: rather than assuming that gas cools into the ISM at the full accretion rate once  $r_{\text{cool}} > r_{\text{vir}}$ , the current SC SAM limits the cooling rate by the freefall time of the hot halo gas, motivated by the behavior of dwarf galaxies in hydrodynamical simulations ([Pandya et al. 2020](#)) (see Section 2.2.5).

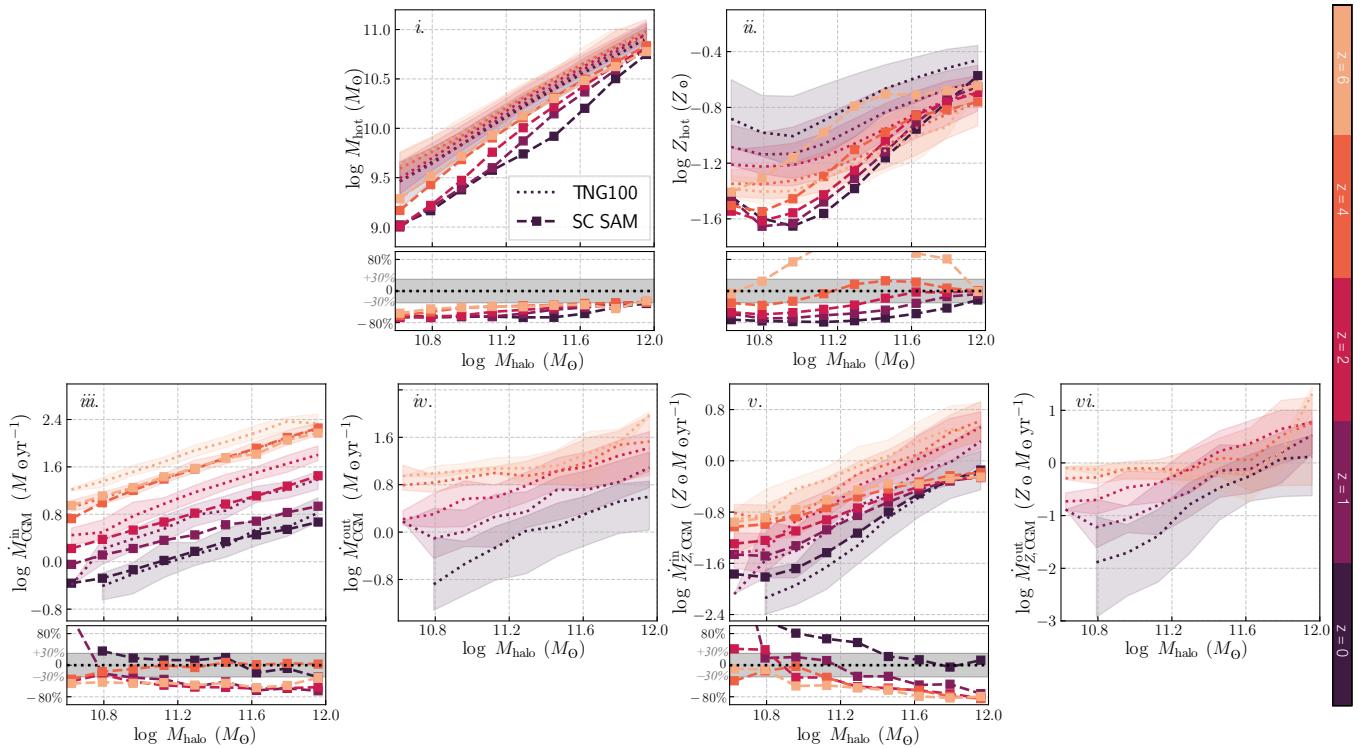
To understand how different baryon cycling drives the large discrepancies in halo-scale properties, the lower panels of Figure 3 compare the rates of gas and metal inflow and outflow. In panels *iii.* and *iv.*, the SC SAM reproduces TNG’s halo gas inflow rates  $\dot{M}_{\text{CGM}}^{\text{in}}$  to within  $\sim 30\%$  across most masses and redshifts, though it predicts lower values for them by up to  $\sim 80\%$  for higher-mass halos at  $z \sim 1, 2, 6$ . In contrast, the SC SAM includes no explicit channel for gas to leave the halo, whereas TNG predicts non-negligible halo

outflows — typically only a factor of  $\sim 2$ –3 times lower than the inflow rates.

A similar pattern holds for the metals. The SC SAM generally predicts either lower or higher metal inflow rates ( $\dot{M}_{\text{Z,CGM}}^{\text{in}}$ , panel *v.*) into the CGM at higher halo masses by up to  $\sim 100\%$ . Meanwhile, TNG predicts substantial metal outflows from the CGM ( $\dot{M}_{\text{Z,CGM}}^{\text{out}}$ , panel *vi.*), often comparable to (and in some cases larger than) the metal inflow rates.

The comparisons across the galaxy and halo scale show that although the SC SAM and TNG often arrive at broadly similar endpoints for properties that can be calibrated against observations, such as stellar mass, cold gas mass, and SFR trends, both models reach those outcomes through very different patterns of baryon cycling. At the scale of the galaxy, the SC SAM occasionally matches TNG’s stellar masses and star formation rates within tens of percent, yet differs more strongly in how material is exchanged within the ISM, with the largest discrepancies in ISM metal flows. These differences are further amplified at the scale of the halo, where the SC SAM lacks a channel for mass or metals to leave the CGM, whereas TNG continuously cycles both in and out, with the rate of outflowing gas/metals from the halo comparable to that of the inflows. As a result, the SC SAM ends up with a more depleted and less enriched CGM, even while producing global galaxy properties that look broadly similar to TNG’s.

Below, we present the modifications made to the SC-SAM to better



**Figure 3.** As in Figure 2, we compare median relations from TNG100 (dotted lines) and the SC SAM (dashed squares) over redshift. Here we show halo-scale properties: *i.* hot gas mass ( $M_{\text{hot}}$ ), *ii.* hot gas metallicity ( $Z_{\text{hot}}$ ), and the corresponding gas and metal flow rates (*iii.* - *vi.*). Across redshift, the SC SAM tends to significantly underpredict both  $M_{\text{hot}}$  and  $Z_{\text{hot}}$ , with differences reaching  $\sim 80\%$ . Gas and metal inflow rates (panels *iii.* and *v.*) vary widely, remaining within  $\sim 30\%$  of TNG at some redshifts, and with many deviations often exceeding  $\sim 80\%$  (e.g.,  $z = 1, 2, 6$  for the gas). Notably, the SC SAM does not include any explicit halo outflow channel for gas or metals (panels *iv* and *vi*), whereas in TNG the halo outflow rates are comparable in magnitude to the inflow rates.

emulate the underlying gas and metal cycling pathways that lead to TNG’s predictions.

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

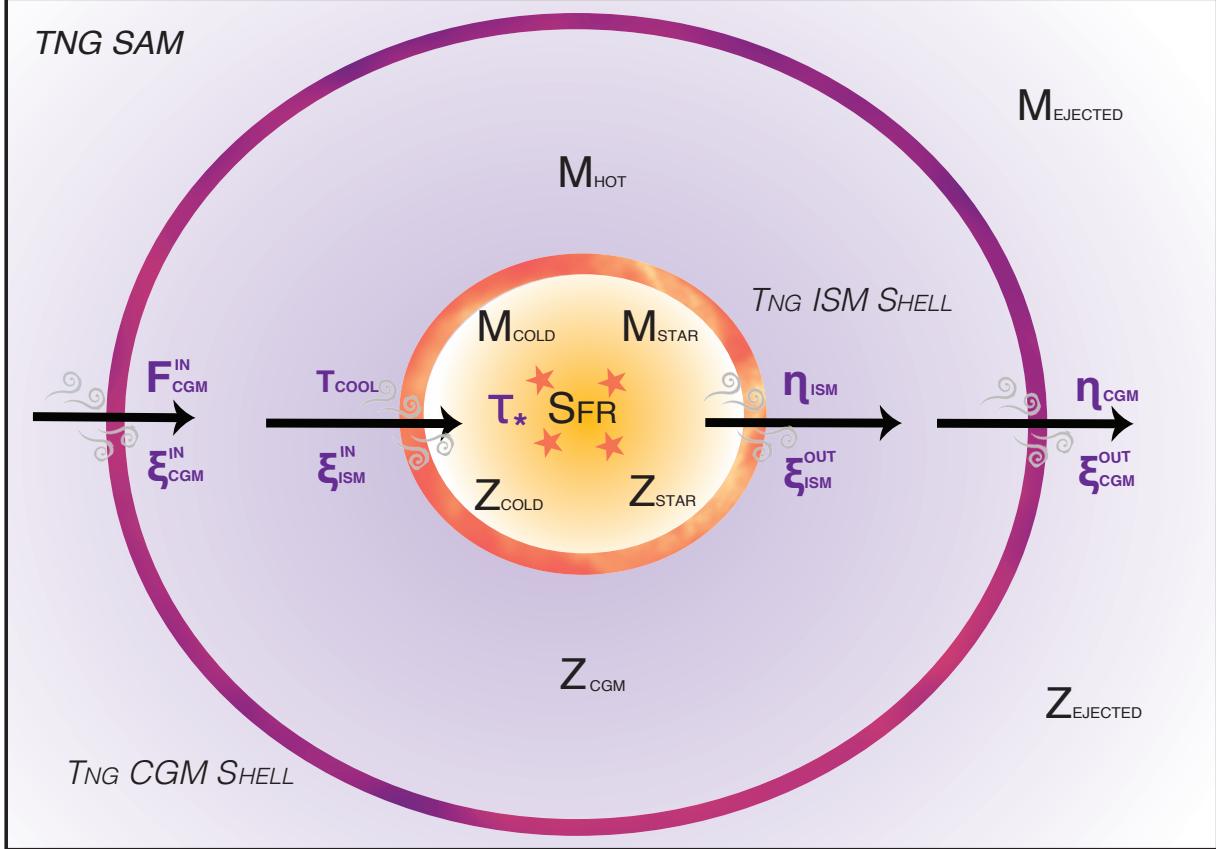
To translate the complex physics of a hydrodynamical simulation like TNG into the simplified framework of the SC SAM, we model how gas and metals circulate between the galaxy and its surrounding halo over time, as summarized in Figure 4. We achieve this using the YO25 subsample of  $\sim 100 - 1000$  TNG galaxies per redshift (see Section 3.2.1), which provides inflow and outflow rates of gas and metals for each galaxy. 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 5 as functions of halo mass and redshift, form the foundation for calibrating the SC SAM to capture the baryon cycling processes 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 5. 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 5) to model all halos with  $\log M_{\text{halo}} < 10.5$ .

Practically, this means that we use the medians at  $\log M_{\text{halo}} = 10.4$  and  $\log M_{\text{halo}} = 10.6$  as anchoring points to construct a smooth transition at  $\log M_{\text{halo}} \approx 10.5$ . While this procedure minimizes stark discontinuities, the transition region can still exhibit mild artifacts, which we discuss further in Section 6.2. Overall, 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 fits analytic functions to these scaling relations across halo mass and time to infer them at intermediate 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 fitting approach, providing a pragmatic approximation of TNG’s behavior with redshift. Each relation is fit within 30% accuracy, and provided 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.’



**Figure 4.** Illustration of the TNG SAM, calibrated to reproduce the baryon cycle of  $\sim 100 - 1000$  central galaxies per redshift in the TNG100 simulation. The diagram shows the main baryonic reservoirs—cold gas ( $M_{\text{cold}}$ ), stars ( $M_{\text{star}}$ ), hot halo gas ( $M_{\text{hot}}$ ), 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  $\eta_{\text{CGM}}$  control gas outflows). Metallicity enrichment is tracked for each reservoir as  $Z_{\text{cold}}$ ,  $Z_{\text{star}}$ ,  $Z_{\text{hot}}$ , and  $Z_{\text{ejected}}$ . Each parameter was calibrated using inflow and outflow rates measured at the ISM and CGM boundaries defined in Figure 1.

#### 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 into 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  $f_{\text{CGM}}^{\text{in}}$  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 ( $f_b = 0.1573$ ), 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. Additionally, since the SAM is run on the TNG DMO merger trees, whose halo masses differ from TNG FP's by up to 20% in low mass halos (see Sections 3.1 & 6.2.2), we rescale the analytic relation for  $f_{\text{CGM}}^{\text{in}}$  so that the resulting  $\dot{M}_{\text{in,CGM}}$  matches TNG within  $\pm 30\%$ .

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 5, 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.

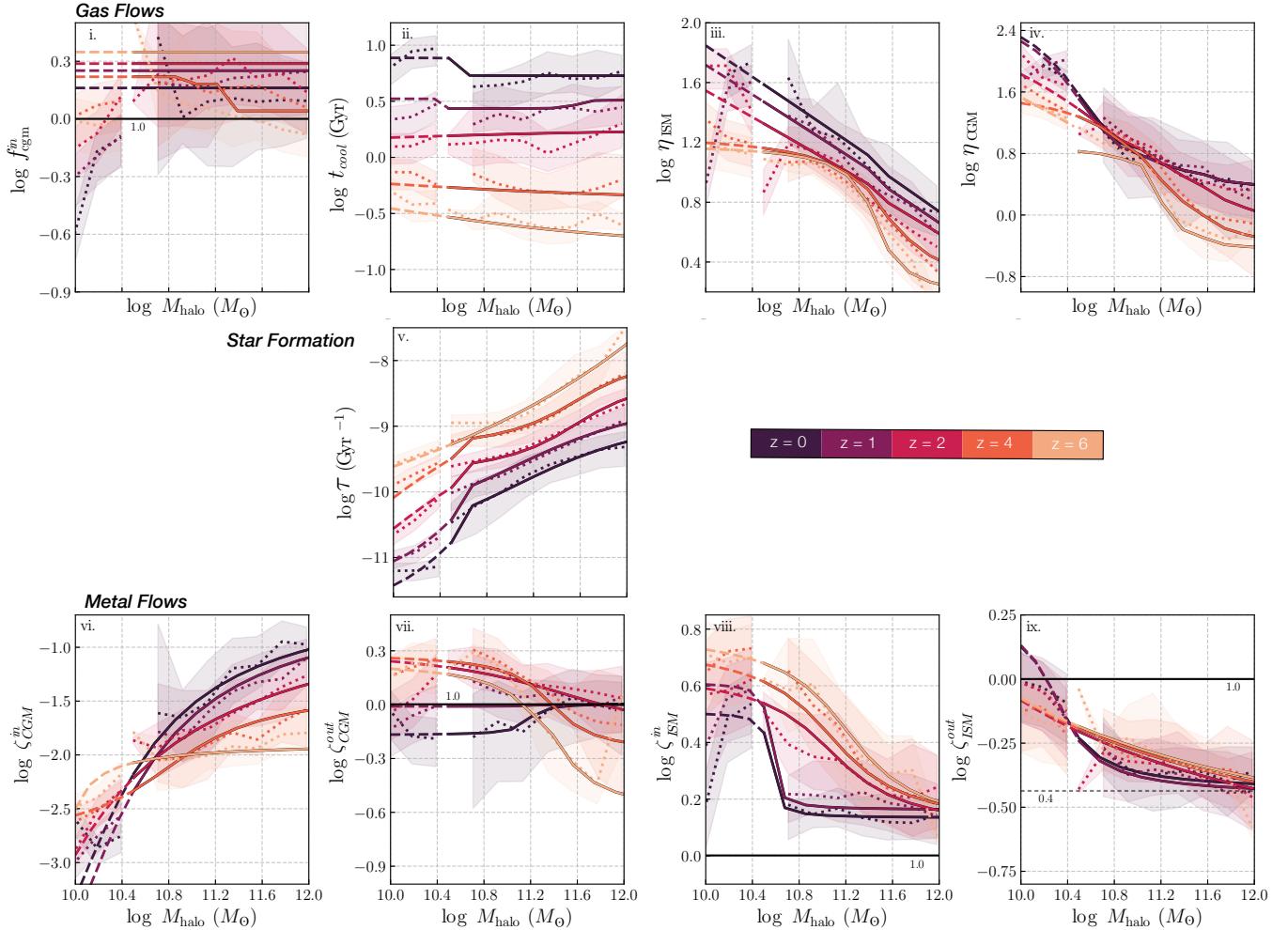
##### 4.1.2 CGM-ISM Cooling

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

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

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



**Figure 5.** Overview of the physical parameters measured from the subset of TNG100 galaxies from YO25, organized by relations regulating gas flows, star formation, and metal flows. Each panel shows the distribution of TNG100 values as a function of halo mass and over redshift as dotted lines colored by redshift. Solid lines show analytic fits based on the well-resolved sample (100–400 galaxies per redshift); dashed lines show fits based on the total sample (> 1000 galaxies per redshift). The parameters shown are: (i.) CGM inflow efficiency ( $f_{\text{CGM}}^{\text{in}}$ ), (ii.) Time for gas to cool from the CGM - ISM ( $t_{\text{cool}}$ ), (iii - iv.) ISM and CGM outflow mass-loading factors ( $\eta_{\text{ISM}}$ ,  $\eta_{\text{CGM}}$ ), (v.) star formation efficiency ( $\tau_*$ ), (vi - vii.) CGM/ISM metal inflow enrichment factors ( $\zeta_{\text{CGM}}^{\text{in}}$ ,  $\zeta_{\text{ISM}}^{\text{in}}$ ), and (viii - ix.) ISM/CGM outflow metal outflow enrichment factors ( $\zeta_{\text{ISM}}^{\text{out}}$ ,  $\zeta_{\text{CGM}}^{\text{out}}$ ). The analytic functions (accurate to within  $\sim 30\%$ ) used in the TNG SAM are provided in Table A1.

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 5 panel *ii.*, shows the resulting cooling times as a function of halo mass and redshift. Across all halo masses, the cooling time exceeds the dynamical time, indicating that the traditional “cold” versus “hot” accretion dichotomy does not describe cooling in TNG (discussed further in Section 6.1.2). In the new TNG SAM, we use the cooling times measured in TNG to determine the rate at which gas accretes onto the galaxy via Equation 8.

## 4.2 Star Formation

TNG models star formation using the subgrid two-phase ISM model of Springel & Hernquist (2003), where an effective equation of state is adopted for the ISM. In this model, dense, cold clouds are embedded in a hot tenuous medium, with the two phases in rough pressure equilibrium. For gas volume densities greater than a critical value  $\rho > \rho_{\text{th}}$ , star formation is assumed to take place at the rate:

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

where  $\rho_c$  is the density of gas in the cold phase.

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

choice is motivated by the fundamental mismatch in how galaxy sizes are defined and interpreted in the SAM versus in TNG. In the SC SAM, stars are assumed to reside in a rotationally supported exponential disk, and this disk scale length sets the gas surface density in the KS relation. In contrast, in TNG, the stellar component comprises dispersion-dominated spheroids, and the reported galaxy sizes (e.g. the stellar half mass radii) reflect the full three-dimensional stellar distribution rather than a rotationally supported disk. As a result, 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 5 panel  $v_*$ ), and calibrate the star formation rate with:

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

### 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, 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  $\dot{m}_{\text{ISM},\text{out}}$  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. In the TNG SAM, we instead apply  $\eta_w$  at the boundary of the ISM, which we denote as  $\eta_{\text{ISM}}$ . The left panel of Figure 6 shows that the value of  $\eta_{\text{ISM}}$  at the scale of the galaxy (solid lines) can differ significantly from the value at launch (dotted lines). At  $z > 0$ , the galaxy-scale outflow rates are systematically higher than the launch rates, indicating that winds entrain additional material as they propagate outward. At  $z = 0$ , the opposite trend is seen: the galaxy-scale outflow rates are lower than the launch rates, suggesting that winds stall, likely due to the higher density and pressure of the CGM. To capture this behavior in the SAM, we model the ratio between  $\eta_w$  at the launch scale and  $\eta_{\text{ISM}}$  at the galaxy scale in TNG, shown in the middle panel of Figure 6.

In the TNG SAM, the material ejected from the ISM is always deposited in the CGM. Unlike the fiducial SC SAM, which allows  $\eta_w$  to grow to extremely large values in low-mass halos (exceeding  $10^3$  at  $10^8 M_{\odot}$ ), the TNG-calibrated scaling relations yield values that remain near  $\sim 100$  in this regime, without the need for an explicit cap.

#### 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 a separate ejected reservoir. The ejected gas returns to 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  $\eta_{\text{CGM}}$ , defined as the ratio of the mass outflow rate  $\dot{m}_{\text{CGM},\text{out}}$  to the star formation rate  $\dot{m}_*$ . Complementing the existing  $\eta_{\text{ISM}}$  at the galaxy scale,  $\eta_{\text{CGM}}$  quantifies the efficiency with which stellar feedback ejects gas from the halo.

The TNG SAM model for  $\eta_{\text{CGM}}$  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:

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

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

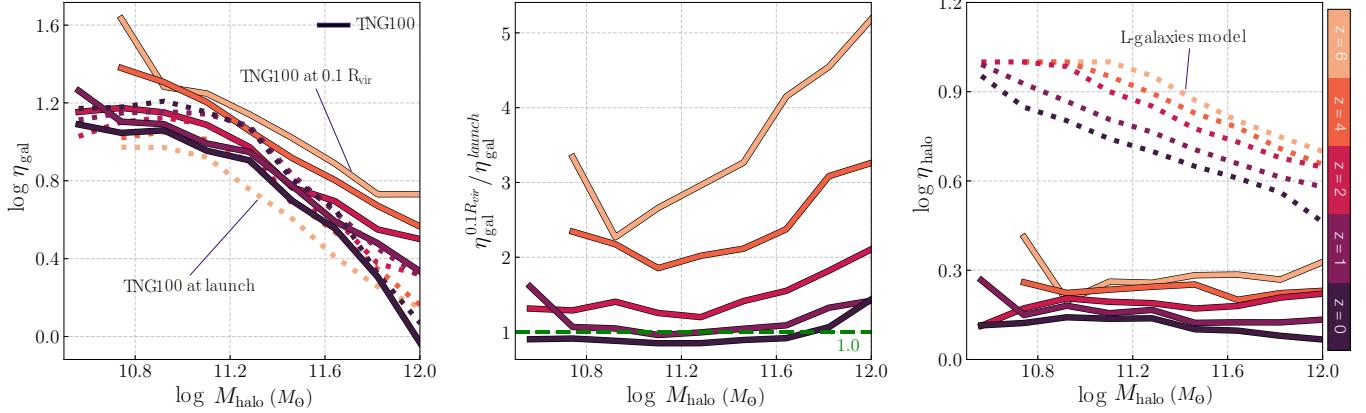
Assuming the quantity  $\Delta m_{\text{eject}}/\dot{m}_*$  is equivalent to our  $\eta_{\text{halo}} \equiv \dot{m}_{\text{CGM-IGM}}/\dot{m}_*$ , we derive the expression for  $\eta_{\text{CGM}}$  as:

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

where  $V_{\text{SN}} = 630 \text{ km/s}$  is constant and  $\epsilon_{\text{halo}}$  is a dimensionless efficiency parameter capped at unity and 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  $V_{\text{eject}} = 100 \text{ km/s}$ . However, as shown in the right panel of Figure 6, applying these values in TNG (dotted lines) results in values for  $\epsilon_{\text{halo}}$  significantly higher than what is measured in TNG (solid lines), overestimating the efficiency by up to a factor of 10. To account for this, the TNG SAM instead parameterizes  $\epsilon_{\text{halo}}$  based on its actual dependence on halo mass and redshift in TNG (solid lines).



**Figure 6.** Functions that govern the TNG SAM’s stellar feedback-driven outflows at the galaxy (left and middle 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  $\eta_{\text{CGM}}$ , 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 measured in TNG (solid lines) using Equation 17, which we used as inputs in the TNG SAM.

In the TNG SAM, gas ejected from the halo does not re-accrete. Instead, the  $\zeta_{\text{CGM}}^{\text{in}}$  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 stars, core-collapse supernovae, and Type Ia supernovae, 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, assumes  $\zeta_{\text{out}} = 1$  for both the ISM and CGM, so that outflowing gas carries metals in direct proportion to the metallicity of the originating reservoir. However, panels *vii.* and *ix.* in Figure 5 show that this assumption does not hold in TNG. At the scale of the halo, for  $z = 0$ ,  $\zeta_{\text{out,CGM}}$  often deviates from unity, mostly lying above 1 for halos with  $\log M_{\text{halo}} < 11.2$ . This indicates that winds leaving the halo are more metal-enriched than the mean metallicity of the hot halo gas. In contrast, for more massive halos, and across all halo masses at  $z = 0$ ,  $\zeta_{\text{out,CGM}}$  remains below unity, indicating that gas leaving the halo is less metal-enriched than the average metallicity of the hot halo.

At the scale of the galaxy,  $\zeta_{\text{out,ISM}}$  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), leading to under-enriched winds compared to the average metallicity of the ISM. However, as the winds reach the scale of the galaxy, they entrain additional metal mass, such that the fraction of metals outflowing from the ISM is mostly greater than 0.4, as shown by the grey dashed line in panel *ix.*

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

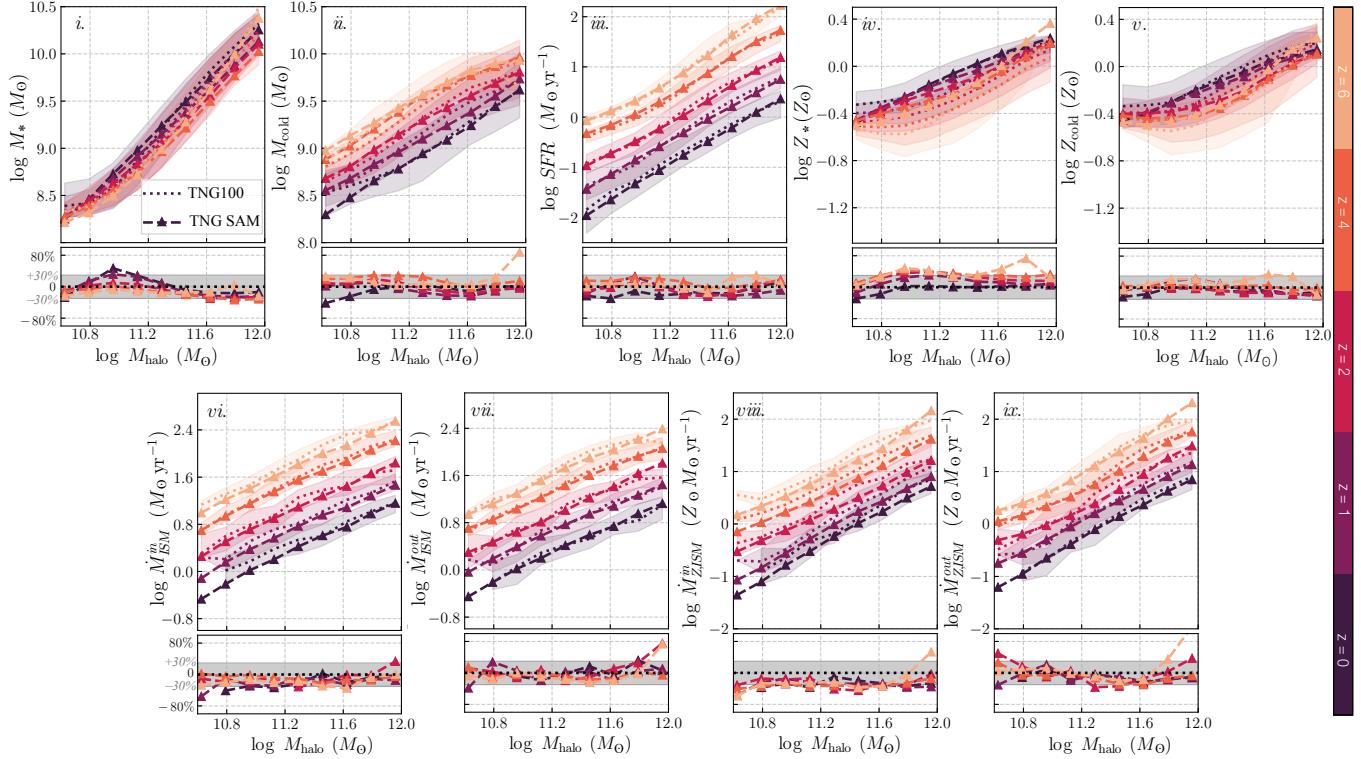
$$\zeta_{\text{ISM}}^{\text{in}} = \frac{\dot{M}_{Z,\text{ISM}}^{\text{in}}}{\dot{M}_{\text{ISM}}^{\text{in}} \cdot Z_{\text{CGM}}}, \quad (20)$$

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 5 panel *viii.* shows that the fraction of metals accreting into the ISM in TNG **consistently exceeds unity across redshift**. This indicates that the gas returning to the ISM is more metal-rich than the average CGM, suggesting efficient cooling of metal-enriched gas. At the scale of the halo, panel *vi.* shows a non-negligible inflow of metals into the CGM given the non-zero values of  $\zeta_{\text{CGM}}^{\text{in}}$ . This indicates that some fraction of the metals ejected by stellar feedback can escape the halo and subsequently be re-accreted, emphasizing the role of metal recycling across both galactic and halo scales.

To further improve the SAM’s agreement with TNG for the metal populations, we increased the stellar yield  $y$  from the fiducial SC SAM value of 1.2,  $Z_{\odot}$  to 1.5,  $Z_{\odot}$ . 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 $Z_{\odot}$ ,



**Figure 7.** Comparison of galaxy-scale properties (panels *i*–*v*) and flow rates (panels *vi*–*ix*) between TNG100 (dotted lines) and the newly calibrated TNG SAM (dashed triangles). As in Figure 2, medians are shown as a function of halo mass, and redshift is indicated by the same color scheme used throughout the paper. Overall, the TNG SAM reproduces both the baryon flow rates and global properties of TNG with substantially improved accuracy relative to the fiducial SC SAM, particularly for the metal population. Across most halo masses and redshifts, the TNG SAM matches TNG within  $\pm 30\%$  for stellar mass, cold gas mass, and star formation rate, metals, and both gas and metal inflow/outflow rates.

assuming  $Z_\odot = 0.02$ , corresponds to a metal production efficiency of  $\sim 3\%$  by mass. Although this is somewhat lower than TNG’s global value, the comparison is not truly one-to-one since TNG tracks delayed enrichment and mixing, 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 and observed in Section 3.3, 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 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.

### 5.1 The Galaxy Scale

In Figure 7, the *top row* (panels *i*–*v*) compares the stellar mass, cold gas mass, star formation rate, and the metallicities of stars and cold gas as a function of halo mass between TNG100 (dotted lines) and the newly calibrated TNG SAM (dashed triangles). The TNG SAM generally reproduces the stellar mass–halo mass relation (panel *i*) to within  $\sim 20$ – $30\%$  over most halo masses and redshifts. At intermediate halo masses ( $\log M_{\text{halo}} \sim 11$ , there is a modest excess at  $z = 0$ , where the TNG SAM predicts stellar masses up to  $\sim 40\%$  higher than TNG. This offset correlates with the slightly elevated cold gas mass and SFR at the same halo mass and redshift (both still within  $\sim 20$ – $30\%$  of TNG’s predictions), indicating that the SAM is

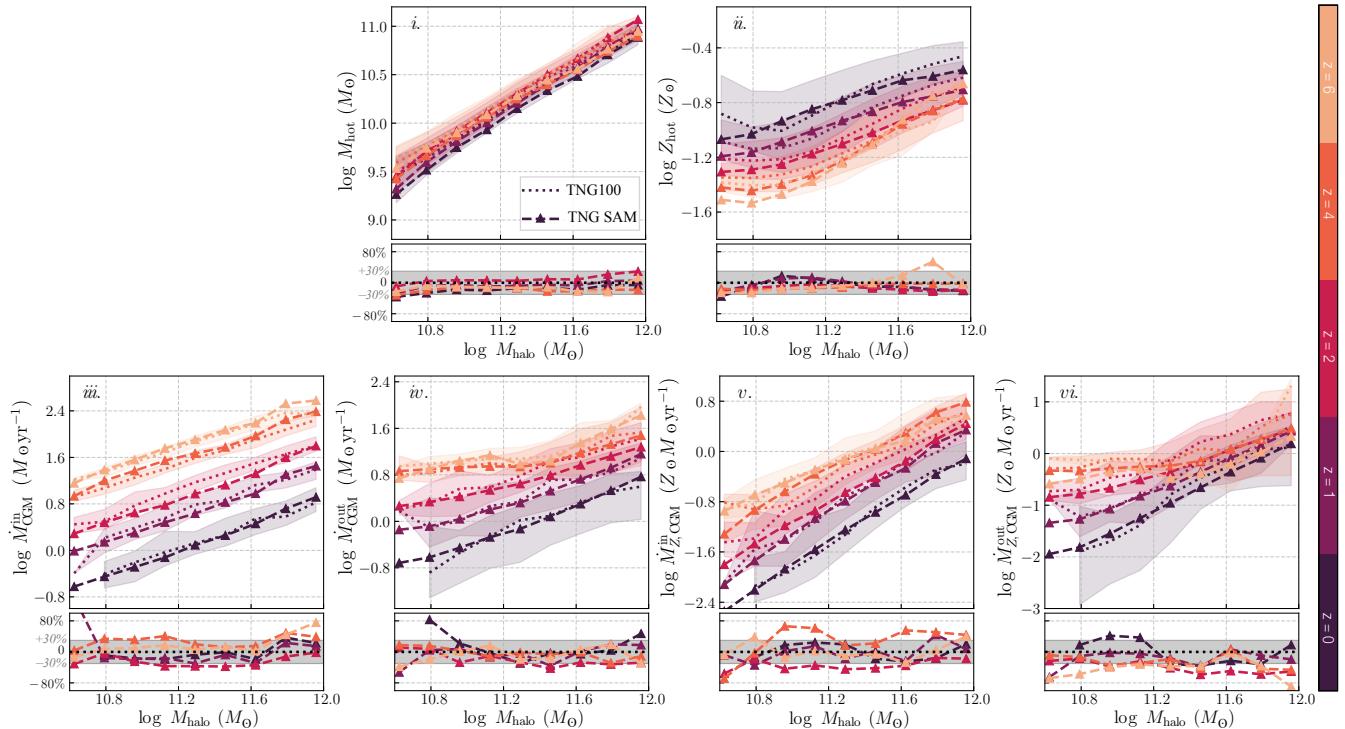
converting a somewhat larger available gas reservoir into stars over time.

The fact that the TNG SAM’s ISM inflow and outflow rates are now reproduced to within  $\sim 30\%$  across halo mass and redshift (panels *vi*–*vii*) is directly connected to the improvements in  $M_{\text{cold}}$  and SFR. Excluding behavior at the edges of the halo mass range (i.e.,  $\log M_{\text{halo}} \sim 10.6$  or  $\log M_{\text{halo}} \sim 12.0$ ), the SAM balances gas inflow into the ISM and gas removal via outflows in a manner consistent with TNG, allowing the cold gas reservoir to remain well regulated over time.

Among the global properties, the metals provide the most direct evidence that reproducing a hydrodynamical simulation’s baryon cycle leads to improved agreement with its predictions. Panels *iv*–*v* show that the TNG SAM matches TNG’s stellar and cold gas metallicities within  $\sim 30\%$  across halo mass and redshift, with deviations up to  $\sim 80\%$  limited to the stellar metallicity at  $z = 6$  in high mass halos. This stands in stark contrast to the SC SAM, which predicted ISM metallicities a factor of 2 times lower. The improvement stems from the fact that the TNG SAM now reproduces metal inflow and outflow rates to within  $\sim 30\%$  of TNG (panels *viii*. and *ix*), with larger deviations confined to the edges of the halo mass range, whereas the SC SAM differed by factors of  $\sim 2$ .

### 5.2 The Halo Scale

In Figure 8, the *top row* (panels *i*–*ii*) compares the total hot gas mass and hot gas metallicity between TNG100 (dotted lines) and the TNG SAM (dashed triangles) as a function of halo mass and redshift. The TNG SAM reproduces both quantities to within  $\sim 30\%$  across most halo masses and redshifts, with small deviations up to



**Figure 8.** Comparison of halo-scale quantities (panels *i*–*ii*) and the corresponding gas and metal flow rates (panels *iii*–*vi*) between TNG100 (dotted lines) and the newly calibrated TNG SAM (dashed triangles). The TNG SAM reproduces the hot gas mass and hot gas metallicity within  $\pm 30\%$  across most halo masses and redshifts. Gas and metal inflow rates are similarly well-matched, typically within  $\pm 30\%$  of TNG. A major improvement over the SC SAM is that the TNG SAM now captures the halo outflow channels with comparable accuracy, reproducing both gas and metal outflows to mostly within  $\sim 30\%$ .

~ 80% mostly near the low and high ends of the mass distribution. This marks a substantial improvement over the SC SAM (Figure 3), which predicted CGM gas masses and metallicities  $\sim 80\%$  lower than TNG.

Panels *iii*–*vi*. show the corresponding baryon flows into and out of the halo. Gas inflow rates (*iii*.) are typically reproduced to within  $\sim 30\%$  across mass and redshift, with some larger deviations (up to  $\sim 100\%$ ) appearing at the edges of the resolved mass range around  $z = 1$  and  $z = 6$ . Metal inflow rates (*v*.) show similar overall agreement, although differences of up to  $\sim 80\%$  appear at  $z = 4$ , mostly confined to the edges of the mass range shown. These wider deviations, and the broader spread around the 30% band, likely arise because  $\dot{M}_{\text{CGM}}^{\text{in}}$  itself already exhibits significant scatter. In the SAM, the halo inflow rate is obtained by finite differencing halo masses from the DMO simulation, which differ from those in the hydrodynamical TNG run. Discussed further in Section 6.2.2, these halo mass differences reach  $\sim 20\%$  for low mass halos and increase toward higher redshift, directly impacting accretion rate estimates. The scatter in  $\dot{M}_{\text{CGM}}^{\text{in}}$  is then propagated and amplified when computing the metal inflow rate, since  $\dot{M}_{\text{Z,CGM}}^{\text{in}}$  is determined via the enrichment factor  $\zeta_{\text{CGM}}^{\text{in}}$ . Despite this, the TNG SAM still captures the overall scaling of CGM gas and metal inflows with halo mass and redshift.

Another stark improvement over the SC SAM appears in the halo outflow channels. The TNG SAM predicts both gas and metal outflow rates to mostly within  $\sim 30\%$  of TNG; by contrast, the SC SAM contained no explicit mechanism for material to leave the halo (Figure 3). The TNG SAM’s gas outflow rates show only mild deviations (occasionally up to  $\sim 80\%$  near the lower mass edge of the halo mass range) from TNG. The metal outflow rates show a slightly wider spread, with deviations reaching up to  $\sim 80\%$  at  $z = 6$ . This is not unexpected, since  $Z_{\text{hot}}$  also deviates more strongly from TNG

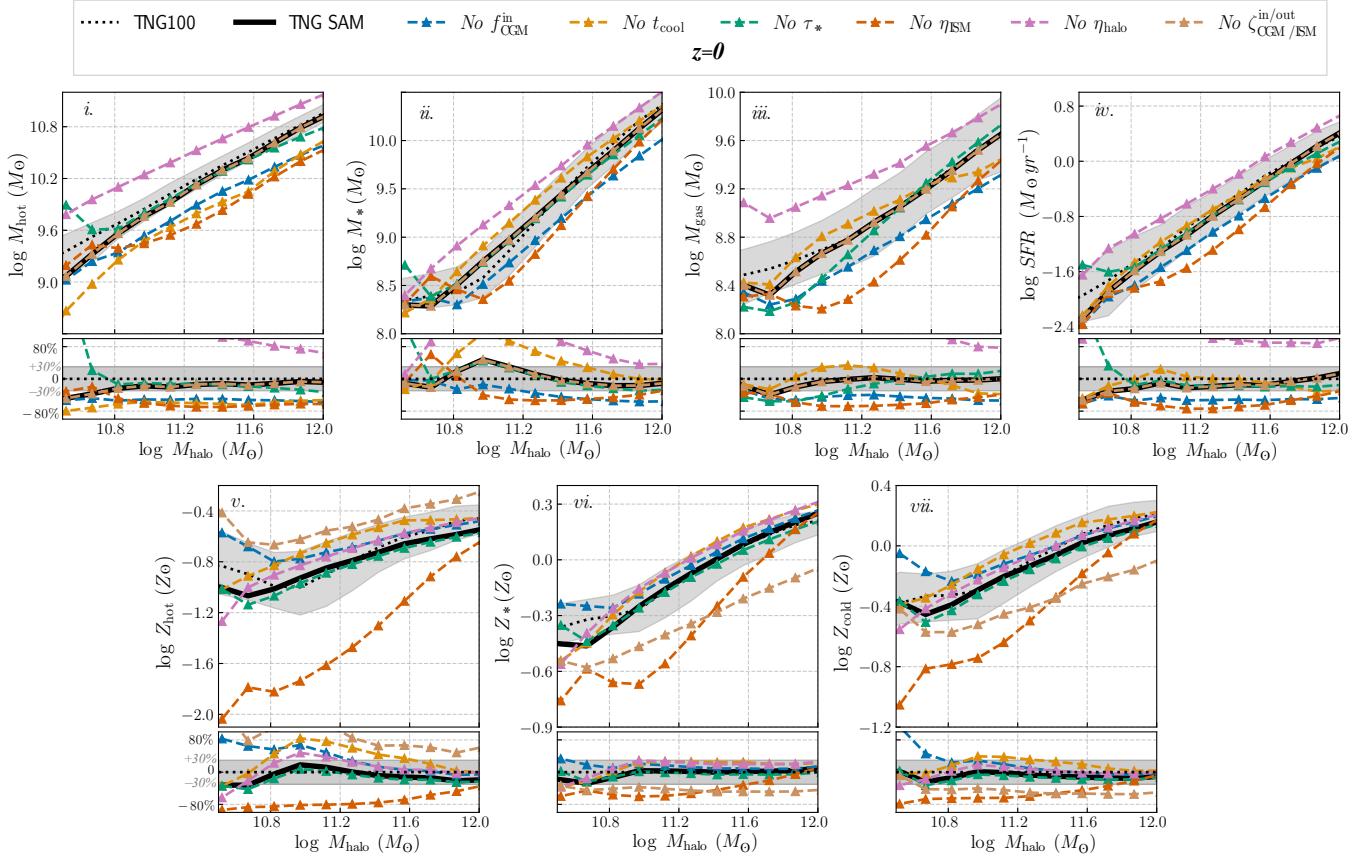
at  $z = 6$ , and directly sets the mass of metals carried by halo-scale outflows.

Although the TNG SAM reproduces halo inflows and outflows less accurately than galaxy-scale flows, capturing the correct qualitative and quantitative trends still has important physical consequences. In the SC SAM, the absence of halo-scale outflows, combined with reduced inflow rates and lower ISM metallicities, produced a significantly under-enriched, less massive hot halo. In contrast, the TNG SAM allows material to both enter and exit the halo at rates comparable to TNG, and also regulates the amount of hot gas that cools into the ISM (panel *vi*. in Figure 7). Together, these effects produce a CGM that better retains both gas and metal mass, bringing the halo-scale properties into substantially closer agreement with TNG.

## 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  $\sim 30\%$  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.



**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.

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

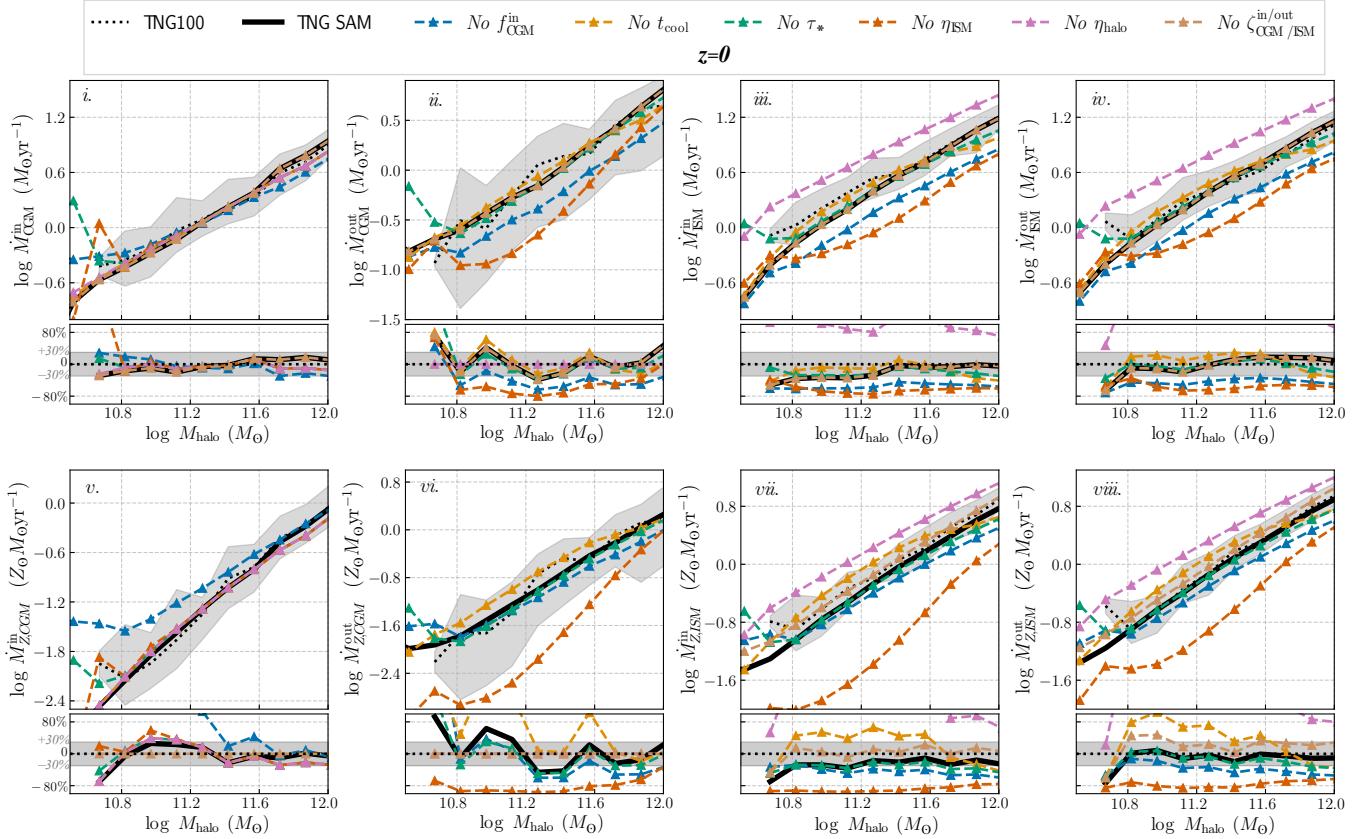
The efficiency of gas accretion into halos, regulated by  $f_{\text{CGM}}^{\text{in}}$ , plays a significant role in the TNG SAM’s ability to replicate TNG’s baryon cycle. As shown in Figure 5,  $f_{\text{CGM}}^{\text{in}}$  often exceeds unity, particularly at lower redshifts, indicating that gas inflow frequently meets or surpasses the expected baryon-to-dark matter ratio. In contrast,  $f_{\text{CGM}}^{\text{in}}$  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  $f_{\text{CGM}}^{\text{in}}$  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  $f_{\text{CGM}}^{\text{in}}$  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 shown in YO25. At higher masses, AGN feedback increasingly contributes, such that the population is split roughly evenly between SN- and AGN-dominated systems. Despite this transition,  $f_{\text{CGM}}^{\text{in}}$  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 recycles back into the halo over time.

In the TNG SAM, the function  $f_{\text{CGM}}^{\text{in}}$  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  $f_{\text{CGM}}^{\text{in}}$ , 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  $\sim 50\%$  (Figure 9, panel *i*), suggests that missing merger-driven inflows alone likely does not account for the full deficit.



**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 reduced CGM inflow rate limits the buildup of hot gas, which in turn suppresses cooling into the ISM, lowering both the mass of the cold gas reservoir and the star formation rate (Figure 9, panels *iii.* & *iv.*). This also propagates through the baryon cycle: reduced hot-halo mass leads to diminished cold-gas accretion (Figure 10 panel *iii.*), which produces less star formation and thus lower stellar mass growth. Interestingly, the suppressed inflow of hot gas into the ISM brings the predicted stellar masses closer to those of TNG (Figure 9, panels *ii.*), despite the fact that the SAM fundamentally relies on an incorrect estimate of how much cold gas is available to form stars in TNG. This gives further credence to the idea that the TNG SAM’s slight overprediction of  $M_*$  at  $\log M_{\text{halo}} \sim 11$  is likely due to definitional differences in how cold gas is measured in TNG (which is ultimately used to set the star formation efficiency). We discuss this further in Section 6.1.3.

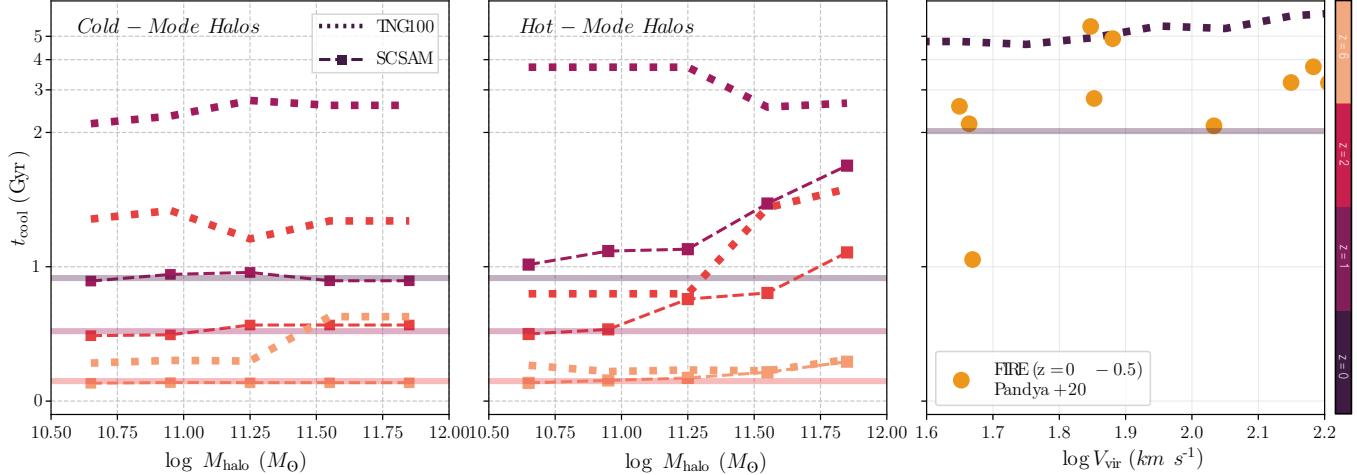
Despite the effectiveness of  $f_{\text{CGM}}^{\text{in}}$  in the SAM, uncertainties remain regarding the precise nature of gas recycling in TNG. TNG’s mesh-based approach makes it difficult to track whether accreted gas is pristine or recycled, unlike particle-based simulations like EAGLE, where gas 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. 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  $f_{\text{CGM}}^{\text{in}}$  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 hydrodynamical 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



**Figure 11.** Cooling times ( $t_{\text{cool}}$ ) for TNG (dotted lines) and the SC SAM (dashed squares). The left and middle panels show the time evolution of  $t_{\text{cool}}$  for cold-mode ( $r_{\text{cool}} > r_{\text{vir}}$ ) and hot-mode ( $r_{\text{cool}} < r_{\text{vir}}$ ) halos, respectively, at  $z = 1, 2, 6$ . The solid lines represent the dynamical time for halos with mass  $10^{11} M_{\odot}$ . 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, and individual FIRE halos (orange circles,  $z = 0 - 0.5$ ; Figure 4.22 in Pandya 2021) at  $z = 0$ . Despite differing methodologies, both TNG and FIRE predict  $t_{\text{cool}}$  values generally longer than the dynamical time (solid line).

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 rates into the ISM 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 hydrodynamical simulations besides TNG. The right panel of Figure 11 compares cooling times in TNG and FIRE, where each yellow circle 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 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 sig-

nificantly improves the SAM’s predictions of halo-scale properties compared to the traditional cooling model, as shown in panel *i*. 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 *iii*), this reflects the lower available hot gas mass rather than the true long-term cooling behavior. The excessive gas delivery into the ISM in the traditional model leads to an overly massive cold gas reservoir, which fuels higher star formation, yielding stellar masses that are systematically larger than those in TNG. Because more stars are formed, the metal populations across the ISM and halo also increase.

While the TNG SAM’s 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.

#### 6.1.3 Cooling and Star Formation Efficiency Regulate 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 successfully reproduces global ISM properties very well. The cold gas mass, stellar mass and star formation rate all closely follow TNG’s trends within 30%, with the cold gas mass and star formation rate performing better than the stellar mass. Panels *ii–iv.* of Figure 9 show that the fiducial SAM’s Bigiel et al. (2008)-inspired Kennicutt–Schmidt star formation recipe leads to good agreement with TNG’s star formation rate, cold gas mass and stellar mass at higher halo masses ( $\log M_{\text{halo}} > 11$ ). At lower halo masses, however, the agreement in the cold gas mass and star formation rate degrades, with deviations reaching up to  $\sim 80\%$ . 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 globally applied 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). Thus, improving disk size models is a necessary prerequisite for reliably applying surface-density-based star-formation laws in SAMs.

Given the current uncertainties in predicting disk sizes in SAMs, tuning the star formation efficiency provides a more robust and flexible way to regulate star formation in the TNG SAM. However, this does not mean that the good agreement between the TNG SAM and TNG is purely a consequence of calibrating  $\tau_*$ . The TNG SAM succeeds because it gets both components of the baryon cycle broadly right: (i) realistic cooling times deliver the correct amount of gas to the ISM (Section 6.1.2), and (ii)  $\tau_*$  regulates how efficiently that gas forms stars. When cooling proceeds too efficiently (as in the traditional model), the extra cold gas fed to the ISM leads to more star formation (Figure 9 panel *ii.*) even though  $\tau_*$  is tuned.

#### 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 metallicity-dependent wind model into the TNG SAM significantly improved agreement with TNG’s outflow rates. Without this modification, the SC SAM’s prescription underestimates ISM outflow rates up to  $\sim 80\%$  across the halo masses explored (Figure 10, panel *iv.*). Because this gas is immediately ejected rather than cycled through the CGM, the SC SAM fails to build up a more massive CGM reservoir (Figure 9, panel *i.*). As a result, the SC SAM predicts lower inflow rates into the ISM (Figure 10, panel *iii.*), suppressed cold gas masses, reduced star formation,

and ultimately lower stellar masses across the resolved mass range (Figure 9, panel *ii–iv.*). In contrast, the TNG SAM’s scaling relations matches the ISM outflow rate within 30%, yielding values that closely track those found in TNG.

Although not obvious in Figure 9 panel *ii.*, 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 in practice the TNG-calibrated scaling relations keep  $\eta_{\text{ISM}}$  near  $\sim 100$  in low-mass halos, rather than driving it to the high levels applied in the SC SAM.

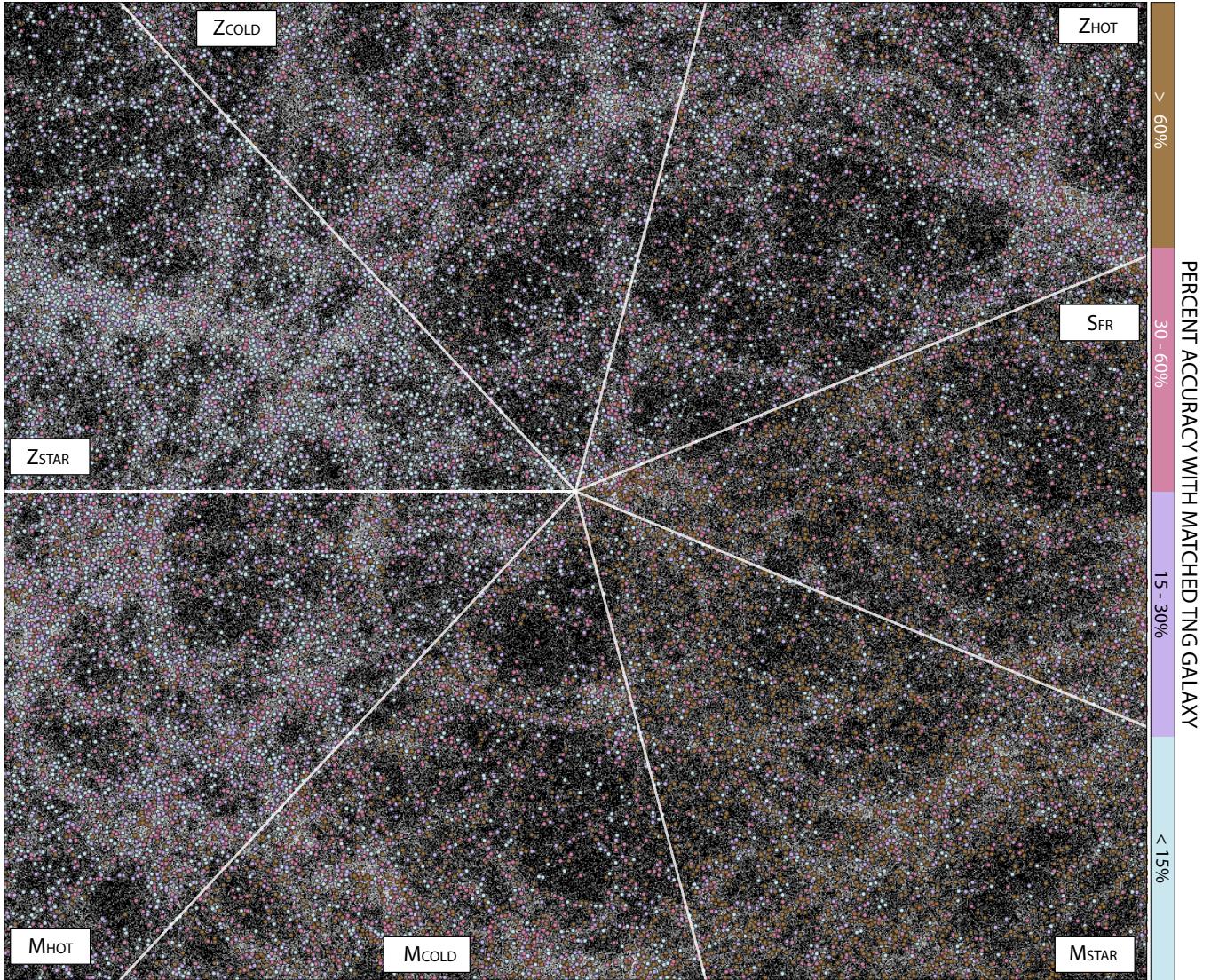
A critical update to the TNG SAM’s feedback model was the explicit inclusion of halo-scale outflows. Traditional SAMs, including the SC SAM (e.g., Croton et al. 2006, 2016; Lacey et al. 2016; Hirschmann et al. 2016), typically assume that gas expelled from the ISM is transferred directly to an ejected reservoir/IGM without first passing through the hot 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.

When halo-scale outflows are neglected, the hot halo grows far too massive and enriched:  $M_{\text{hot}}$  and  $Z_{\text{hot}}$  exceed TNG by factors of 2–3 and 40%, respectively (Figure 9, panels *i* & *v*). This excess mass then cools too efficiently into the ISM, driving cold gas masses, SFRs, and stellar masses to overshoot TNG’s predictions by factors of 2–3 (panels *ii–iv.* in Figure 9).

#### 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 the metallicities found in TNG100. If we instead retain the SC SAM’s original assumptions, namely that the proportion of metals flowing into and out of the ISM is unity (i.e.,  $\zeta^{\text{in}/\text{out}}_{\text{ISM}} = 1$ ), and that there is no explicit prescription for the proportionality of metals entering or leaving the CGM (i.e.,  $\zeta^{\text{in}/\text{out}}_{\text{CGM}} = 0$ ), the TNG SAM underpredicts the cold gas and stellar metallicities by up to  $\sim 80\%$ , while overpredicting the hot gas metallicity by more than  $\sim 80\%$ , as shown in Figure 9 panels *v–vii.* The excess in  $Z_{\text{hot}}$  arises because metals expelled from the ISM into the CGM are retained in the halo due to the absence of a halo-scale metal outflow channel, leading to a buildup of metals in the hot halo. We note that this comparison is not strictly one-to-one with the SC SAM. In the SC SAM, metals entering the CGM are implicitly tied to the return fraction of gas ejected from the halo, whereas setting  $\zeta^{\text{in}/\text{out}}_{\text{CGM}} = 0$  in the TNG SAM removes the halo-scale metal inflow and outflow channels without reinstating the associated reaccretion pathway.

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



**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  $10.6 < \log M_{\text{halo}} < 12$  are shown in white. The map is divided into seven sectors corresponding to a specific galaxy property: 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 sector. Blue and purple points represent agreement within 30%, while pink and brown points highlight discrepancies greater than 30%. While the majority of galaxies agree within 30% for the metal populations and mass of the hot gas, notable variance is observed for galaxy-scale properties like  $M_{\text{star}}$  and  $M_{\text{cold}}$ , illustrating the challenges of calibrating to median trends alone.

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 future SAMs will be critical for accurately modeling how metals cycle through galaxies and their halos.

## 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 TNG’s medians, Figure 12 shows that the spread in-

creases significantly when comparing individual galaxies. The TNG SAM still demonstrates generally good agreement for several key properties, namely  $M_{\text{hot}}$  and all metal populations, in which the majority of all matched galaxies agree within 30% of TNG. The agreement is less consistent for  $M_{\text{star}}$ ,  $M_{\text{cold}}$  and the star formation rate. Here, the TNG SAM predicts each property within 30% accuracy for only 30-40% of galaxies.

The disagreement among the gas properties at the galaxy scale likely stems from mismatches in how scatter arises in the star formation rate. In TNG, galaxy-to-galaxy variation in SFR and cold-gas mass reflects both local physical processes and gas cycling in the ISM. In the TNG SAM, the global star-formation efficiency regulates the mean gas consumption rate successfully, but does not reproduce the full spread in individual systems. In addition, the cold gas mass used to form stars in the SAM corresponds to gas within twice the stellar half mass radius, whereas in TNG star formation can proceed from gas that was, or is, located outside this aperture. As a result, the SAM may assign slightly too much or too little cold gas to star formation in individual galaxies, leading to deviations in scatter even when median trends agree well.

Despite this limitation, the results are encouraging. Nearly half of the 20,000 matched galaxies agree with TNG’s predictions within 30%, even though the SAM was primarily calibrated using medians derived from a much smaller resolved sample of just  $\sim 400$  galaxies. This level of agreement is particularly promising given the computational simplicity 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 [Gabrielpilai et al. \(2022\)](#), the halo masses in the TNG100 DMO and TNG100 FP 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  $f_{\text{CGM}}^{\text{in}}$ , 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  $f_{\text{CGM}}^{\text{in}}$  in the SAM to match the hydrodynamical simulation was not feasible. Instead, we inferred the value of  $f_{\text{CGM}}^{\text{in}}$  by aligning the rate of gas entering the CGM with TNG, which performed remarkably well (see Section 6.1.1).

### 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 ( $\log M_{\text{halo}} > 10.6$ ); 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 TNG SAM’s overall 30% agreement with TNG.

### 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 maintains agreement with TNG at the galaxy scale to within  $\sim 40\%$  even when calibrated using only 100 randomly selected galaxies per redshift. Across redshift, aggregate galaxy- and halo-scale properties are generally reproduced at this level, although the cold gas mass shows the weakest agreement, with deviations reaching  $\sim 80\%$  in some regimes (see Figures A4 and A5). The baryon flow rates are less tightly matched: gas and metal inflow and outflow rates are typically reproduced to within  $\sim 40\%$ , while the rates of gas and metals leaving the halo show the largest discrepancies, with divergence up to  $\sim 80\%$ . These results suggest that a larger sample size would likely improve the accuracy currently achieved by the TNG SAM. They also show that reproducing the correct evolutionary behavior of gas and metal cycling, rather than matching instantaneous flow rates as closely as possible, is sufficient to achieve good agreement with TNG’s global properties.

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 sam-

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

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

ple—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  $10.6 < \log M_{\text{halo}} < 12$ , 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 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  $f_{\text{CGM}}^{\text{in}}$ , except  $f_{\text{CGM}}^{\text{in}}$  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  $M_{\text{star}}$  as all stars bound to the subhalo, and  $M_{\text{cold}}$  within twice the stellar half-mass radius. Other works have adopted different choices. For example, N12 also define  $M_{\text{star}}$  in their hydrodynamical comparison as the total stellar mass bound to the subhalo, but they define  $M_{\text{cold}}$  as the mass of all gas particles eligible to form stars and  $M_{\text{hot}}$  as all remaining gas, without clearly stating whether gas associated with satellite galaxies is also excluded. Such defi-

nitional choices directly affect derived properties, particularly  $M_{\text{hot}}$ , since what counts as “hot” depends directly on how “cold” is defined. This raises an important question: to what extent do reported levels of agreement in earlier works depend on their specific definitions, and how might their conclusions 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 *ii.* showing that the stellar mass is most sensitive to  $\eta_{\text{CGM}}$  and  $\eta_{\text{ISM}}$ .

[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. [MS22](#) likewise found that the ISM mass is more sensitive to halo-scale flows, whereas the stellar mass is more sensitive to galaxy-scale flows. Similarly the TNG SAM shows that  $\eta_{\text{CGM}}$  significantly impacts both  $M_{\text{cold}}$  and  $M_{\text{star}}$ , but with  $M_{\text{cold}}$  more impacted than  $M_{\text{star}}$ .

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 simulation. The TNG SAM bridges the detailed physical processes captured in hydrodynamical simulations with the computational efficiency of SAMs, offering a powerful tool to study the 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, cold 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 balance between gas recycling and preventative feedback varies across hydrodynamical simulations, requiring flexible (re-)accretion models in SAMs.* In TNG, the gas inflow efficiency ( $f_{\text{CGM}}^{\text{in}}$ ) at the scale of the halo often exceeds unity, indicating substantial gas recycling, where previously ejected gas re-enters the halo, plays a dominant role. This behavior contrasts with simulations like FIRE and EAGLE, which show lower  $f_{\text{CGM}}^{\text{in}}$  values due to stronger preventative feedback, defined here as processes that suppress gas accretion into halos. The majority of SAMs rely on fixed or static gas return fractions and therefore do not capture the dynamic recycling efficiencies seen in hydrodynamical simulations. For future SAMs, incorporating flexible (re-)accretion models that vary with halo mass and redshift such as the TNG SAM’s  $f_{\text{CGM}}^{\text{in}}$  will be critical for accurately predicting the baryon content of galaxies (Section 6.1.1, Figure 9).

(ii) *The classic 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 hydrodynamical 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).

(iii) *Directly calibrating the cooling time and star-formation efficiency yields more accurate global ISM predictions.* In the TNG SAM, the star formation rate is regulated by a star-formation efficiency calibrated from the full TNG100 sample, while gas condensation into the ISM is governed by cooling times matched to the TNG subsample. These two calibrations work jointly: realistic cooling times deliver the correct amount of gas to the ISM, and the calibrated star-formation efficiency regulates how that gas forms stars, producing stellar masses, cold gas masses, and star formation rates to mostly within  $\sim 30\%$  of TNG’s predictions across time. (Section 6.1.3, Figure 9 panels *ii.–iv.*) Attempts to apply a global Kennicutt–Schmidt law instead lead to large discrepancies because the fiducial SAM does not compute disk sizes in a manner comparable to TNG. 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.

(iv) *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, cold gas, and hot gas mass predictions. (Section 6.1.4, Figures 9 panels *i.–iv.* and 10 panels *i. & iv.*).

(v) *Accurately modeling metal-enriched and metal-depleted flows*

between the ISM and CGM is essential for reproducing galaxy- and halo-scale metallicities. The TNG SAM's use of metallicity-weighted enrichment factors that capture how metal-enhanced or metal-depleted gas flows are relative to their source reservoirs significantly improves its ability to match TNG's predictions for the metallicity of the stars, cold gas and hot gas. Explicitly modeling these metal enrichment factors is therefore critical for accurately tracking the redistribution of metals between the ISM, CGM, and IGM. (Section 6.1.5, Figures 9 panels *v* - *vii*. and 10 panels *v-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)

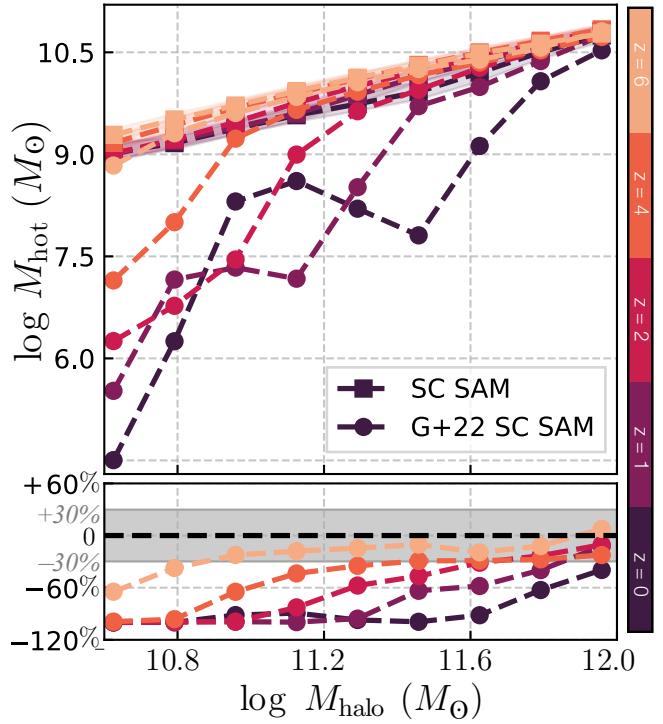
## DATA AVAILABILITY

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

## 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, J 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
- Diemer B., et al., 2019, *MNRAS*, 487, 1529
- 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
- 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, J 10.48550/arXiv.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  
Rodríguez-Gómez 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  
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



**Figure A1.** Comparison of the Santa Cruz SAM used in this work (dashed squares) to the Santa Cruz SAM employed in [Gabrielpillai et al. 2022](#) (dashed circles) for the hot halo gas mass ( $M_{\text{hot}}$ ). The original SC SAM cooling model depleted the CGM by up to  $\sim 120\%$  relative to the updated model presented here. While the revised cooling prescription produces a substantially less depleted hot halo, the updated SC SAM still underpredicts  $M_{\text{hot}}$  relative to TNG100 by up to  $\sim 80\%$  (see Figure 3).

## APPENDIX A: BUILDING BLOCKS OF THE TNG SAM

### A1 Comparison to the published SC SAM

[Pandya et al. \(2020\)](#) showed that the SC SAM substantially underpredicts the mass of the hot gas relative to the FIRE simulations, and [Gabrielpillai et al. \(2022\)](#) (G+22) showed that the same discrepancy appears when the SC SAM is 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. The updated cooling model suggested by [Pandya et al. \(2020\)](#) significantly improved the prediction of the hot halo gas mass, as shown in Figure A1. However, despite this major improvement at the halo-scale, the impact on galaxy-scale gas properties and overall metal content remained modest, with deviations remaining within 0.5 dex relative to the SC SAM presented in [Gabrielpillai et al. \(2022\)](#).

### A2 Analytic Scaling relations

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

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

All coefficients used in Equation A1 are described in Table A1. All galaxy- and halo-scale properties entering this calibration are

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	
...	$\eta_{\text{CGM}}$		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.

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	GroupMassType part 0 ['SubhaloGrNr'] - $\sum_i^n$ SubhaloMassInRadType part 0	Gas mass of the circumgalactic halo
<i>Galaxy</i>	$M_*$	GalpropMstar	SubhaloMassType part 4	Stellar mass
	$M_{\text{cold}}$	GalpropMcold	SubhaloMassInRadType part 0	Gas mass of the interstellar medium
	SFR	GalpropSfr	SubhaloSFR	Star formation rate
	$Z_*$	GalpropZstar	SubhaloStarMetallicity	Stellar metallicity
	$Z_{\text{cold}}$	GalpropZcold	SubhaloGasMetallicity	Cold gas metallicity
	$Z_{\text{hot}}$	HalopropZhot	GroupGasMetallicity['SubhaloGrNr'] - $\left( \sum_i^n \right)$ SubhaloGasMetallicity . SubhaloMassInRadType part 0 )	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.

defined consistently between the SAM and TNG. A complete mapping between the physical quantities and their corresponding field names in the SAM and TNG outputs is provided in Table A2.

### A3 TNG SAM Results at Intermediate Redshifts

Figures 7 and 8 show the results of the TNG SAM at select redshifts to avoid visual clutter. For completeness, Figures A2 and A3 show the remaining intermediate redshifts. Across the full redshift range explored, the TNG SAM generally reproduces TNG’s predictions to within  $\pm 30\%$ . The agreement degrades primarily at  $z = 8$ , where the number of well-resolved galaxies in TNG is substantially lower ( $\sim 10$ ).

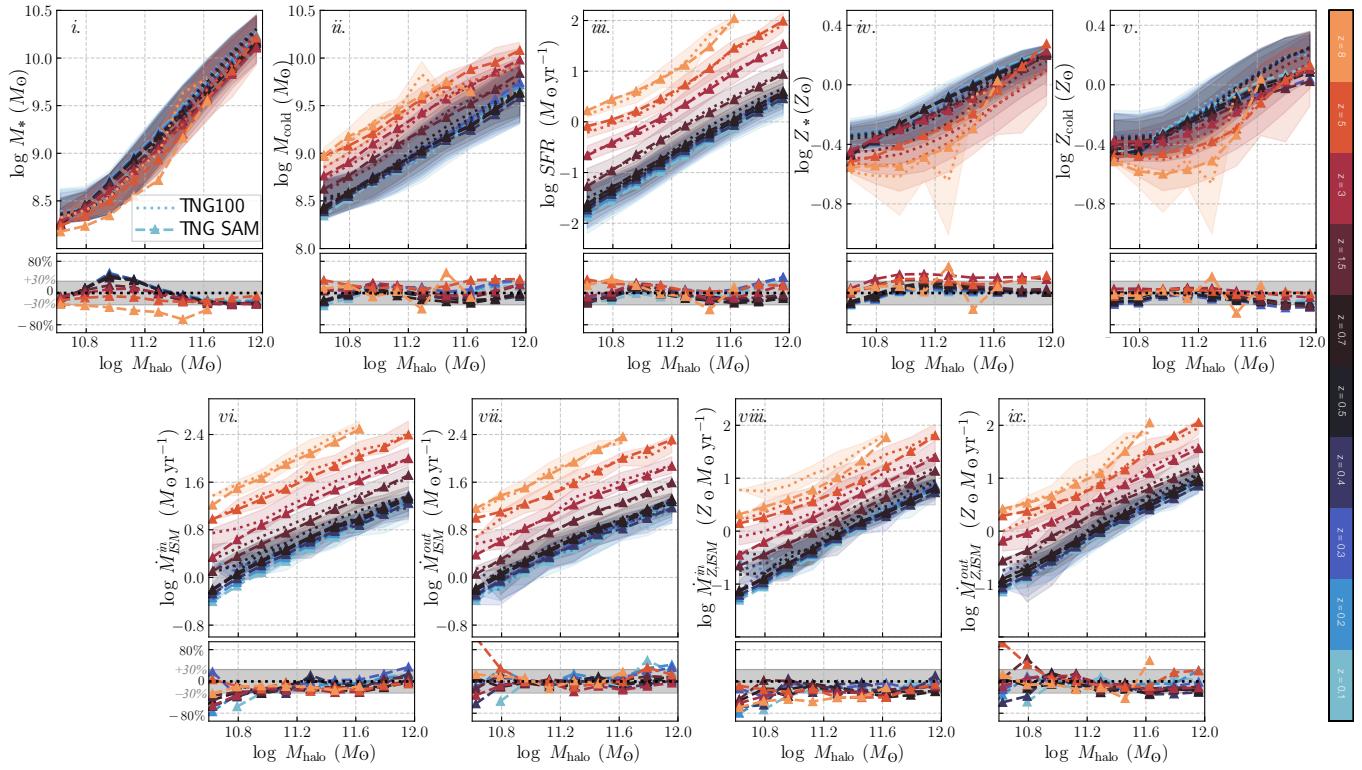
### A4 How Few Galaxies Are Enough? Calibration with 100 Galaxies

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  $10^{10} - 10^{12} M_{\odot}$ . Despite the substantially reduced calibration set, the TNG SAM matches TNG’s predictions at

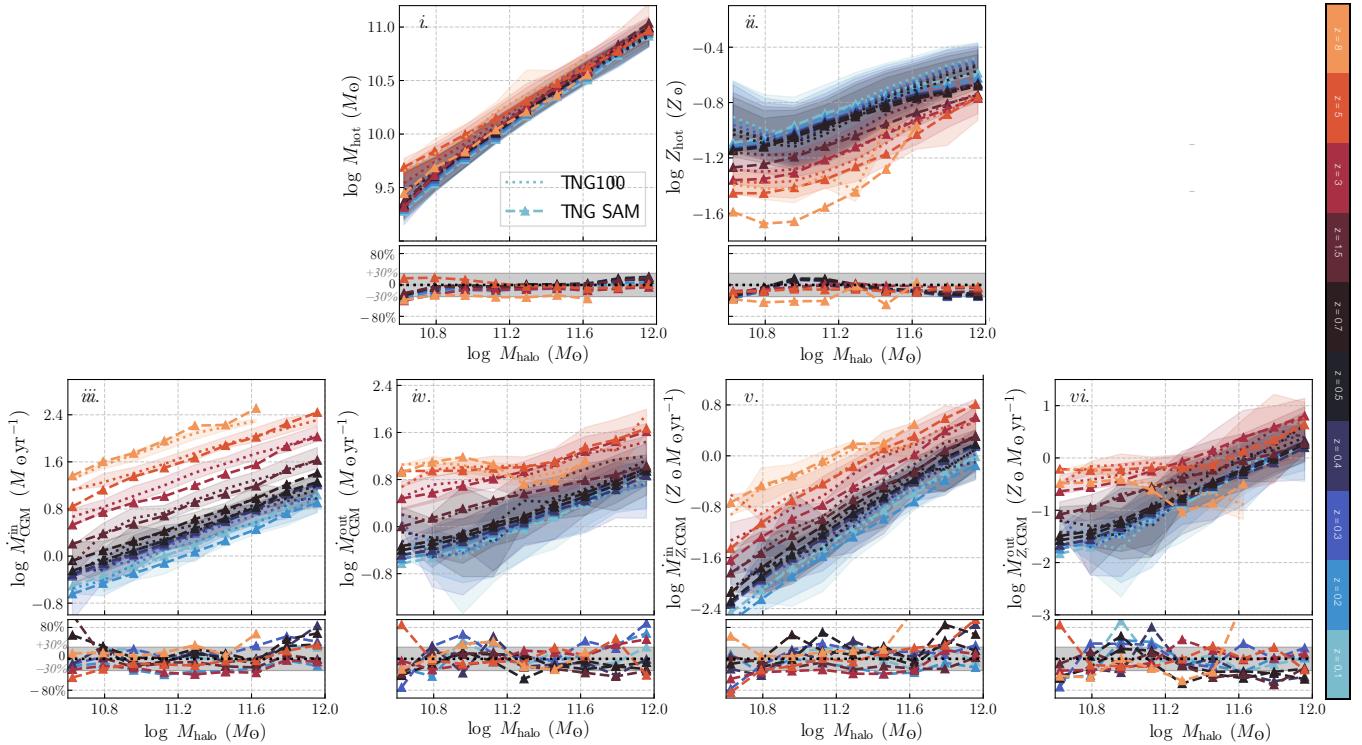
the galaxy-scale mostly within  $\sim 40\%$  across redshift. Figures A4 and A5 show that aggregate quantities such as stellar mass, star formation rate, and halo-scale gas content generally remain within this range, although the cold gas mass shows the weakest agreement, with deviations reaching up to  $\sim 80\%$  in some regimes.

Figures A4 and A5 also show that the baryon flow rates are more sensitive to the reduced calibration set. Gas and metal inflow and outflow rates are typically recovered to within  $\sim 40\%$ , while the largest discrepancies—up to  $\sim 80\%$ —occur for gas and metals leaving the halo. Nevertheless, the overall redshift and halo-mass evolution of these flows is captured reasonably well.

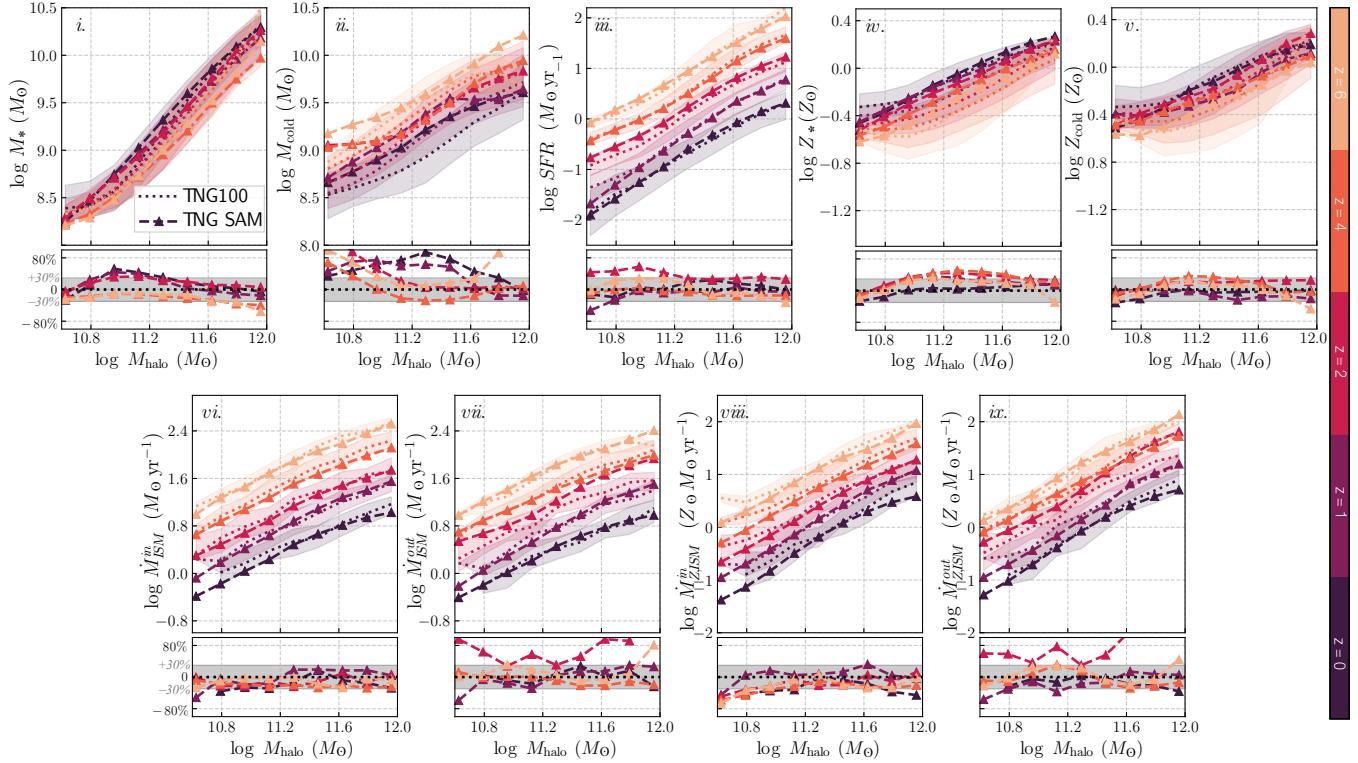
This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.



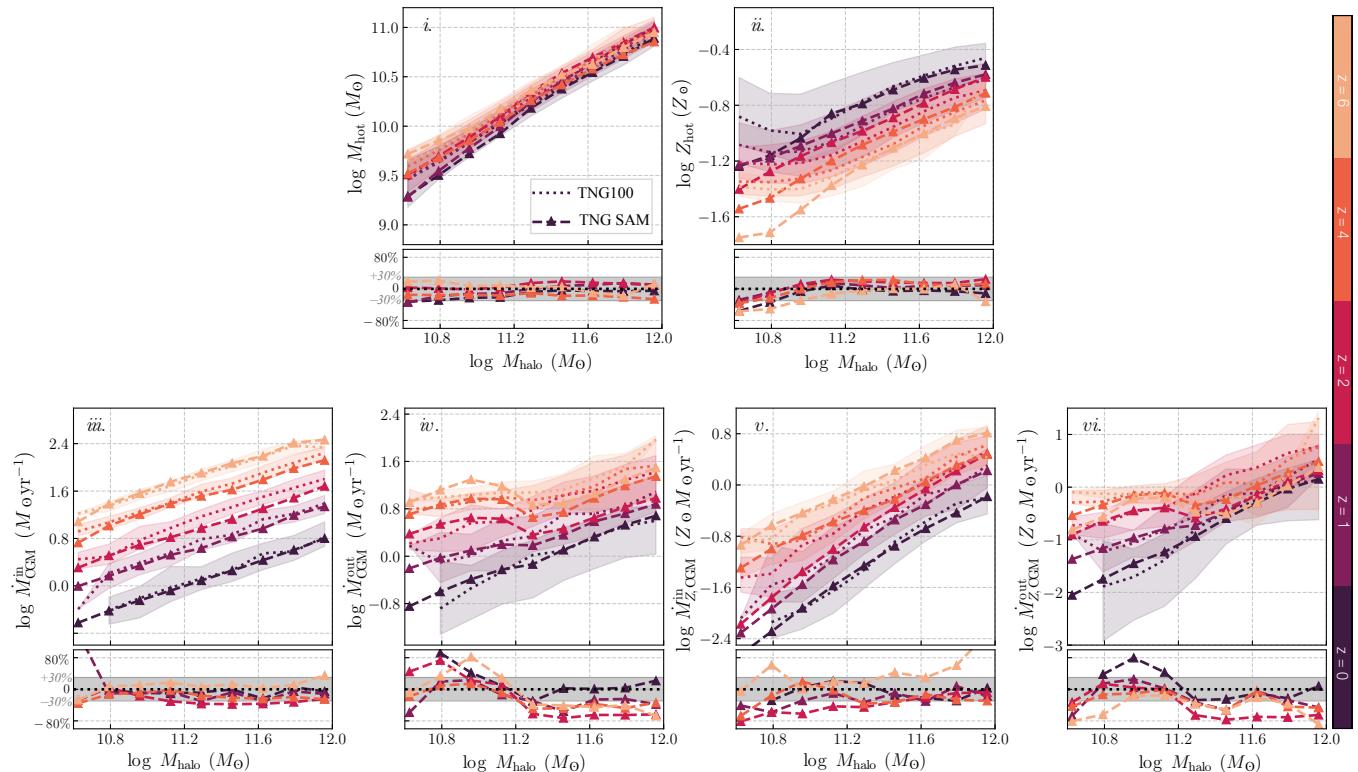
**Figure A2.** Extension of Figure 7 showing the evolution of galaxy-scale properties (panels *i*–*iv*) and baryon flow rates (panels *v*–*ix*) for the redshifts not displayed in Figure 7. Low-redshift relations are mostly shown in blues ( $z < 0.5$ ), high-redshift relations in reds ( $1.5 \leq z \leq 8$ ), and intermediate redshifts in darker colors ( $0.5 \leq z \leq 0.7$ ). Across most halo masses and redshifts, the TNG SAM continues to reproduce TNG100's global galaxy properties and flow rates to within  $\sim 30\%$ , with larger deviations appearing primarily at the highest redshift ( $z = 8$ ) where the number of well-resolved systems is limited.



**Figure A3.** Extension of Figure 8 showing halo-scale quantities (panels *i*–*ii*) and the corresponding gas and metal flow rates (panels *iii*–*vi*) for the redshifts not displayed in Figure 8. Median relations are shown using the same redshift color-coding as in Figure A2. The TNG SAM maintains agreement with TNG100 at the  $\sim 30\%$  level for most halo masses and redshifts, including for halo-scale gas and metal inflows and outflows, with discrepancies increasing toward the higher redshift.



**Figure A4.** Galaxy-scale properties and baryon flow rates for the TNG SAM calibrated using only  $\sim 100$  randomly selected galaxies per redshift. The figure follows the same layout, quantities, and redshift color-coding as Figure 7, but for the reduced calibration sample. Despite the limited number of galaxies used for calibration, the TNG SAM continues to reproduce aggregate galaxy-scale properties to mostly within  $\sim 40\%$  accuracy across redshift. The cold gas mass shows the weakest agreement, with deviations reaching up to  $\sim 80\%$  in some regimes.



**Figure A5.** Halo-scale quantities and baryon flow rates for the TNG SAM calibrated using only  $\sim 100$  randomly selected galaxies per redshift, shown using the same layout and redshift color scheme as Figure 8. While aggregate halo properties remain mostly consistent with TNG100 within  $\sim 40\%$ , halo-scale flow rates are more sensitive to the reduced calibration set. The largest discrepancies occur for gas and metal outflows from the halo, which can differ by up to  $\sim 80\%$ .