The CAMELS Project: Expanding the Galaxy Formation Model Space with New ASTRID and 28-parameter TNG and SIMBA Suites

Yueying Ni[1,2], Shy Genel[3,4], Daniel Anglés-Alcázar[3,5], Francisco Villaescusa-Navarro[3,6], Yongseok Jo[3],
Simeon Bird[7], Tiziana Di Matteo[2,8], Rupert Croft[2,8], Nianyi Chen[2], Natalí S. M. de Santi[3,9], Matthew Gebhardt[5],
Helen Shao[6], Shivam Pandey[10,11], Lars Hernquist[1], and Romeel Dave[12]

[1] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA; yueying.ni@cfa.harvard.edu
[2] McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[3] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA
[4] Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA
[5] Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269, USA
[6] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544 USA
[7] Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
[8] NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[9] Instituto de Física, Universidade de São Paulo, R. do Matão 1371, 05508-900, São Paulo, Brasil
[10] Department of Physics, Columbia University, New York, NY 10027, USA
[11] Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA
[12] Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
Received 2023 April 10; revised 2023 September 21; accepted 2023 October 9; published 2023 December 14

Abstract

We present CAMELS-ASTRID, the third suite of hydrodynamical simulations in the Cosmology and Astrophysics with MachinE Learning (CAMELS) project, along with new simulation sets that extend the model parameter space based on the previous frameworks of CAMELS-TNG and CAMELS-SIMBA, to provide broader training sets and testing grounds for machine-learning algorithms designed for cosmological studies. CAMELS-ASTRID employs the galaxy formation model following the ASTRID simulation and contains 2124 hydrodynamic simulation runs that vary three cosmological parameters (Ω_m, σ_8, Ω_b) and four parameters controlling stellar and active galactic nucleus (AGN) feedback. Compared to the existing TNG and SIMBA simulation suites in CAMELS, the fiducial model of ASTRID features the mildest AGN feedback and predicts the least baryonic effect on the matter power spectrum. The training set of ASTRID covers a broader variation in the galaxy populations and the baryonic impact on the matter power spectrum compared to its TNG and SIMBA counterparts, which can make machine-learning models trained on the ASTRID suite exhibit better extrapolation performance when tested on other hydrodynamic simulation sets. We also introduce extension simulation sets in CAMELS that widely explore 28 parameters in the TNG and SIMBA models, demonstrating the enormity of the overall galaxy formation model parameter space and the complex nonlinear interplay between cosmology and astrophysical processes. With the new simulation suites, we show that building robust machine-learning models favors training and testing on the largest possible diversity of galaxy formation models. We also demonstrate that it is possible to train accurate neural networks to infer cosmological parameters using the high-dimensional TNG-SB28 simulation set.

Unified Astronomy Thesaurus concepts: Large-scale structure of the universe (902); Hydrodynamical simulations (767)

1. Introduction

Traditional methods used to extract information from cosmological surveys typically rely on studying the properties of the Universe (e.g., galaxy distribution, neutral hydrogen distribution) on sufficiently large scales beyond a few megaparsecs (see, e.g., Mandelbaum et al. 2013; Abbott et al. 2018), so that uncertainties from astrophysical processes such as halo–galaxy connection, feedback from supernovae (SNe), and supermassive black holes (SMBHs) remain small and under control. However, the study of the distribution on small, nonlinear scales of a few megaparsecs potentially holds the strongest statistical constraining power on cosmology owing to the high signal-to-noise ratio measurements. Recent studies have shown that extracting information embedded in (mildly) nonlinear scales can significantly tighten the constraints on the cosmological parameters (e.g., Seljak et al. 2017; Hahn et al. 2020; Banerjee & Abel 2021).

In order to extract the maximum available information on nonlinear scales from cosmological surveys, we need to overcome two important obstacles: (1) the optimal estimator that can capture the statistical properties of the considered tracer (e.g., galaxies) is unknown, and (2) astrophysical processes alter the properties and spatial distribution of tracers in an unknown manner. Recent advances in machine learning (ML) have enabled new possibilities to tackle this problem. On one hand, neural networks can be used to build surrogate models for summary statistics such as the power spectrum, allowing predictions to be made for different cosmologies (Villaescusa-Navarro et al. 2020). On the other hand, we can attempt to marginalize uncertainties from baryonic effects by training networks based on dark matter, gas, and stellar fields from a variety of simulations (Villaescusa-Navarro et al. 2021a, 2021c). Thus, it seems that a powerful way to approach this

problem will be to train models based on a large variety of simulations that explore different galaxy formation models spanning the possible space of cosmological and astrophysical parameters.

This is one of the main ideas behind the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project (Villaescusa-Navarro et al. 2021a). CAMELS contains thousands of N-body and state-of-the-art hydrodynamic simulations that cover a range of cosmological and astrophysical parameters. Its data have been used for a large variety of tasks, from weighing the Milky Way (Villanueva-Domingo et al. 2021) to checking the robustness of halo finders (Vallés-Pérez et al. 2022). The hydrodynamic simulation suites in CAMELS contain two distinct subgrid physics models based on IllustrisTNG (henceforth TNG; Weinberger et al. 2017; Pillepich et al. 2018) and SIMBA (Davé et al. 2019), run with the AREPO (Springel 2010) and GIZMO (Hopkins 2015) codes, respectively. This allows us to test the robustness of developed ML models, i.e., to investigate whether a model trained on one simulation suite (produced by a particular subgrid physics model) will still perform well when tested on the other. We note that it is essential to distinguish between the precision and accuracy of the ML model. Models that are precise but not accurate are not useful. CAMELS aims to build accurate ML models by providing many different simulations that can be used to test the model's robustness (or accuracy).

Some works have shown that certain ML models are not robust to variations in subgrid physics (see, e.g., Villaescusa-Navarro et al. 2021a; Villanueva-Domingo & Villaescusa-Navarro 2022; Delgado et al. 2023), while others have found the opposite (Villaescusa-Navarro et al. 2021c; Shao et al. 2022; Villanueva-Domingo et al. 2022; de Santi et al. 2023; Shao et al. 2023; Wadekar et al. 2023). Several natural questions arise given this situation. Will the ML models that are robust across the two subgrid physics implementations also be accurate when tested on a third different subgrid physics model? Can we build more robust ML models when training on simulations from two different subgrid models combined? Should we build ML models based on extensions of the original galaxy formation models to larger parameter spaces that capture more of their innate flexibility?

Those essential questions motivate us to expand the CAMELS simulation sets by incorporating new suites of hydrodynamic simulations with distinct subgrid physics models, as well as to extend the previous simulation model parameter space to higher dimensions to explore other cosmological and astrophysical parameters that can play crucial roles in various observational properties. Ideally, ML models built to marginalize over astrophysics uncertainties should be tested on simulations from as many galaxy formation models (with well-explored parameter spaces) as possible to gain confidence that those ML models will perform well when tested on real data.

In this paper, we introduce the third hydrodynamic simulation suite in CAMELS: CAMELS-ASTRID, carried out with the galaxy formation models of the ASTRID simulation and distinct from the models of TNG and SIMBA. The ASTRID simulation is a recently developed large-volume, high-resolution cosmological hydrodynamic simulation. The production run of ASTRID evolves a $(250 \ h^{-1} \ \mathrm{Mpc})^3$ cosmic volume with 2×5500^3 particles (Bird et al. 2022; Ni et al. 2022), currently run to $z = 1.3$. ASTRID is descended from the simulation and model code used to run the

BlueTides (Feng et al. 2016) and MassiveBlack-II simulations (Khandai et al. 2015) and incorporates a full array of modern subgrid models for stellar and black hole (BH) feedback. ASTRID uses the simulation code MP-Gadget, an extremely scalable thread-based variant of Gadget that incorporates the hierarchical gravitational algorithm from Gadget-4 (Springel et al. 2021) and scales to 5×10^5 cores.

We also introduce the CAMELS extension simulation sets of TNG, SIMBA, and ASTRID, which aim to further broaden the study of cosmological and astrophysical parameters in their respective galaxy formation models, in addition to the six fiducial parameters varied in the core CAMELS simulation sets. This incorporates the new TNG-SB28 set that contains 1024 simulations varying 28 parameters in the TNG model, accompanied by the TNG-1P-28 set that varies one parameter at a time, and an analogous SIMBA-1P-28 set based on the SIMBA model. It also incorporates the ASTRID-SBOb set that contains 1024 simulations varying Ω_b in addition to the six fiducial parameters, focusing on disentangling the effect of Ω_b with Ω_m for cosmological studies based on baryonic observables.

This paper is organized as follows. In Section 2, we first summarize the astrophysical subgrid models employed in CAMELS-ASTRID and then give an overview of the new CAMELS-ASTRID suite and the CAMELS parameter extension simulation sets. Section 3 presents an illustrative comparison between the fiducial model of the ASTRID simulation and that of the TNG and SIMBA simulations. In Section 4, we diagnose the effect of the six fiducial cosmological and astrophysical parameters varied in CAMELS-ASTRID and compare them to TNG and SIMBA. We give a detailed review of the 28 parameters that are varied in the new TNG- and SIMBA-based simulations in the appendices. Section 5 provides an overview of some cosmological and astrophysical properties covered by the five large simulation sets in CAMELS: the three Latin Hypercube (LH) training sets of ASTRID, TNG, and SIMBA and the two parameter extension sets of ASTRID-SBOb and TNG-SB28. In Section 6, we present some test cases and examples of ML applications enabled by introducing the large hydrodynamic simulation suites of CAMELS-ASTRID and TNG-SB28. We finally summarize this work in Section 7.

2. Simulations

In this section, we give a brief overview of the new simulation sets brought by the ASTRID astrophysical model. We also introduce additional simulation sets in CAMELS that aim to explore an extended parameter space in the TNG and SIMBA models. This section is organized as follows. Section 2.1 introduces the astrophysical models and the varied parameters in CAMELS-ASTRID. Section 2.2 gives an overview of the main simulation sets of CAMELS-ASTRID. Section 2.3 gives a brief introduction of the CAMELS extension simulation sets for extended cosmological and astrophysical parameter studies.

2.1. ASTRID Models in CAMELS

The ASTRID galaxy formation model is described in Bird et al. (2022) and Ni et al. (2022). ASTRID utilizes a new version of the MP-Gadget simulation code, a massively scalable version of the cosmological structure formation code Gadget-3 (Springel 2005), to solve the gravity (with an N-body tree-particle-mesh approach, TreePM), hydrodynamics (with a

smoothed particle hydrodynamics method), and astrophysical processes with a series of subgrid models. An earlier version of MP-Gadget was used to run the BlueTides simulation (Feng et al. 2016), and the most recent version was used for ASTRID.

Radiative cooling and photoionization heating include primordial radiative cooling (Katz et al. 1996), metal line cooling with the gas and stellar metallicities traced following Vogelsberger et al. (2014), a spatially uniform ionizing background from Faucher-Giguère (2020), and hydrogen self-shielding following Rahmati et al. (2013).

Star formation is implemented based on the multiphase star formation model in Springel & Hernquist (2003a) and accounts for the effects of molecular hydrogen based on H_2 fraction calculated from the metallicity and local column density (Krumholz & Gnedin 2011). ASTRID tracks metal enrichment from AGB stars, Type II SNe, and Type Ia SNe, following nine individual elements (H, He, C, N, O, Ne, Mg, Si, Fe).

2.1.1. SN Wind Models and A_{SN1}, A_{SN2}

Galactic winds driven by stellar feedback are implemented kinetically via temporarily hydrodynamically decoupled particles. Winds are sourced by newly formed star particles, which randomly pick gas particles from within their SPH smoothing length to become wind particles. Once a particle is in the wind, it is hydrodynamically decoupled for a minimum of 60 Myr, or $20\,\mathrm{kpc}/v_w$, and is recoupled once its density drops by a factor of 10. Particles in the wind do not experience or produce pressure forces, but they do receive the mass return, cool, and contribute to density estimates.

In the CAMELS-ASTRID suite, we use two parameters A_{SN1} and A_{SN2} to vary the strength of the SN wind feedback. In particular, A_{SN1} modulates the total energy injection rate per unit star formation, and A_{SN2} controls the speed of the SN wind. In the fiducial ASTRID model, the prescribed wind speed v_w is proportional to the local one-dimensional dark matter velocity dispersion $v_{w,\mathrm{fid}} = \kappa_\omega \sigma_{DM}$ with $\kappa_\omega = 3.7$ (following the Illustris model; see Vogelsberger et al. 2013). In the ASTRID suite, we have

$$v_w = A_{SN2} \times v_{w,\mathrm{fid}}. \tag{1}$$

The SN feedback model in ASTRID is purely energy driven. Therefore, the asymptotic mass-loading factor of the SN wind scales with the wind speed by $\eta_w \propto v_w^{-2}$. In the fiducial ASTRID model, $\eta_{w,\mathrm{fid}} = (v_w/\sigma_{0,\mathrm{fid}})^{-2}$ with $\sigma_{0,\mathrm{fid}} = 353\ \mathrm{km\ s^{-1}}$ (Bird et al. 2022).

With A_{SN1} modulating the power of the SN feedback energy, the SN wind mass-loading factor in the ASTRID suite is

$$\eta_w \equiv \frac{\dot{M}_w}{\dot{M}_{SFR}} = A_{SN1} \times \eta_{w,\mathrm{fid}} = A_{SN1} \times \left(\frac{\sigma_{0,\mathrm{fid}}}{v_w}\right)^2. \tag{2}$$

We note that A_{SN1} in CAMELS-ASTRID controls the total energy injection rate (power) per unit star formation, similar to the A_{SN1} parameter applied in the CAMELS-TNG suite.

2.1.2. Black Hole Models and A_{AGN1}, A_{AGN2}

The SMBH models in ASTRID are described in Ni et al. (2022). They are built on the earlier models applied in BlueTides (Feng et al. 2016) and have added new features related to dynamical friction to improve the BH dynamics and mergers, as well as applied a power-law seeding prescription

with BH seed mass M_{sd} between $3 \times 10^4\,h^{-1}\,M_\odot$ and $3 \times 10^5\,h^{-1}\,M_\odot$.

However, due to the low resolution of CAMELS, we did not adopt the BH seeding and dynamical friction model as applied in the ASTRID production run. Instead, we adopted the BH seeding and dynamics prescriptions applied in BlueTides. BHs with initial mass $M_{sd} = 5 \times 10^5\,h^{-1}\,M_\odot$ are seeded in halos with $M_h = 5 \times 10^{10}\,h^{-1}\,M_\odot$ by means of the on-the-fly friends-of-friends (FOF) halo finding. BH particles are repositioned to the location of the local potential minimum at each active time step, and two BHs located within $2\epsilon_g$ (where ϵ_g is the gravitational softening length) of each other are instantaneously merged.

The gas accretion rate onto the BH is estimated via the Bondi–Hoyle–Lyttleton–like prescription (Di Matteo et al. 2005). We allow for short periods of super-Eddington accretion in the simulation but limit the accretion rate to 2 times the Eddington accretion rate. The BH radiates with a bolometric luminosity L_{Bol} proportional to the accretion rate $\dot{M}_{BH}$, with a mass-to-light conversion efficiency $\eta = 0.1$ in an accretion disk according to Shakura & Sunyaev (1973).

SMBH feedback follows a two-mode approach including the thermal and kinetic feedback mode delineated by the Eddington ratio of the instantaneous BH accretion rate. The Eddington threshold χ_{thr} is capped at $\chi_{thr,max} = 0.05$ and is also a function of the BH mass such that the kinetic mode is turned on only for massive BHs with $M_{BH} \gtrsim 5 \times 10^8\,h^{-1}\,M_\odot$. In CAMELS-ASTRID, we use A_{AGN1} and A_{AGN2} to modulate the efficiency of the kinetic and thermal feedback separately, where A_{AGN1} has the same physical meaning as that in the CAMELS-TNG suite. In the high-accretion mode, the AGN feedback is deposited thermally with

$$\Delta\dot{E}_{high} = A_{AGN2} \times \epsilon_{f,th}\epsilon_r \dot{M}_{BH}c^2 \quad (\lambda_{Edd} > \chi_{thr}). \tag{3}$$

In the fiducial run, we have a mass-to-light conversion efficiency $\epsilon_r = 0.1$ and $\epsilon_{f,th} = 0.05$, assuming that 5% of the radiation energy is thermally injected into the surrounding gas within a feedback sphere with a radius two times that of the SPH kernel.

The formalism of the AGN kinetic feedback in the low-accretion mode largely follows Weinberger et al. (2017) with different choices of parameter values. The AGN kinetic feedback energy is deposited as

$$\Delta\dot{E}_{low} = A_{AGN1} \times \epsilon_{f,kin}\dot{M}_{BH}c^2 \quad (\lambda_{Edd} < \chi_{thr}), \tag{4}$$

where $\epsilon_{f,kin}$ scales with the BH local gas density and has a maximum value of $\epsilon_{f,kin,max} = 0.05$. The energy is accumulated over time and released in a bursty way once the accumulated kinetic feedback energy exceeds the threshold $E_{inj,min} = f_{re}\frac{1}{2}\sigma_{DM}^2 m_{enc}$. Here σ_{DM}^2 is the one-dimensional dark matter velocity dispersion around the BH, m_{enc} is the gas mass in the feedback sphere, and $f_{re} = 5$ for the fiducial run. The released kinetic energy kicks each gas particle in the feedback kernel in a random direction with a prescribed momentum weighted by the SPH kernel.

Compared to TNG (Weinberger et al. 2017), ASTRID turns on the AGN kinetic feedback based on a more stringent criterion (with a lower χ_{thr} and a higher $M_{BH,pivot}$) and also adopts a lower upper limit for the feedback efficiency $\epsilon_{f,kin,max}$. Therefore, we note that the AGN kinetic feedback in the

ASTRID suite is milder compared to TNG, as will be discussed in the next sections.

2.1.3. Summary of the CAMELS-ASTRID Model

Table 1 briefly summarizes the physical meaning of the four astrophysical feedback parameters. A_{SN1}, A_{SN2}, and A_{AGN1} in ASTRID modulate the energy of SN feedback per unit star formation, the wind speed, and AGN kinetic feedback energy per unit BH accretion (similarly to those parameters in the TNG suite). A_{AGN2} modulates the AGN thermal feedback energy per unit BH accretion, as the thermal mode is the dominant channel of AGN feedback in the ASTRID simulation.

We note again that the CAMELS-ASTRID suite has applied some adaptations in astrophysical models compared to the production run of ASTRID, mainly due to the limited volume and resolution of the CAMELS simulations. In particular, simulations in CAMELS-ASTRID do not adopt the BH power-law seeding and dynamical friction model as applied in the ASTRID production run (Ni et al. 2022), as those models require higher mass resolution. Moreover, with the limited volume of 25 h^{-1} Mpc, CAMELS-ASTRID does not model patchy hydrogen reionization, helium reionization, or the effect of massive neutrinos as in the ASTRID production run.

2.2. CAMELS-ASTRID Suite

CAMELS-ASTRID is a third hydrodynamic simulation suite in CAMELS (Villaescusa-Navarro et al. 2021a) brought by the astrophysical models of the ASTRID simulation. The core CAMELS-ASTRID suite shares a similar design to the simulation sets from the "TNG" and "SIMBA" suites, containing 1092 simulation runs in four simulation sets (CV, 1P, LH, EX) as will be briefly discussed below. In Table 2, we summarize the core simulation sets of the CAMELS-ASTRID suite (appending to Table 2 of Villaescusa-Navarro et al. 2021a) and describe them individually below. We introduce the extended simulation sets of CAMELS (based on ASTRID, TNG, and SIMBA) in the next subsection.

(1) The CV set (CV for cosmic variance) contains 27 simulations with different realizations (random seed) of initial conditions and fiducial cosmological and astrophysical parameters fixed at $\Omega_m = 0.3$, $\sigma_8 = 0.8$, $A_{SN1} = A_{SN2} = A_{AGN1} = A_{AGN2} = 1$. Note that the CV set shares the same set of the initial random seeds across ASTRID, TNG, and SIMBA, allowing cross-comparison between the three simulation suites of the same realizations with their respective fiducial models.

(2) The 1P set (1P for one-parameter) contains 61 simulations sharing the same realization (random seed) of the initial conditions and varying the cosmological and astrophysical parameters one at a time. The varied ranges of the parameters are $\Omega_m \in [0.1, 0.5]$, $\sigma_8 \in [0.6, 1.0]$, $A_{SN1} \in [0.25, 4.0]$, $A_{SN2} \in [0.5, 2.0]$, $A_{AGN1} \in [0.25, 4.0]$, and $A_{AGN2} \in [0.25, 4.0]$. The spacing is linear for Ω_m and σ_8 and logarithmic for the four astrophysical parameters.

We note that the different range of A_{AGN2}, which is varied between 0.25 and 4.00 for ASTRID, and between 0.5 and 2.0 for TNG and SIMBA, is motivated by their corresponding physical meaning. Unlike A_{AGN2} in the TNG and SIMBA suites that modulates the gas ejection speed in the jet/kinetic mode feedback, the function of A_{AGN2} in ASTRID resembles that of A_{AGN1} and controls the AGN feedback efficiency in

another channel (thermal feedback); therefore, it shares the same variation range as A_{AGN1}.

(3) The LH set contains 1000 simulations with the value of the cosmological and astrophysical parameters (Ω_m, σ_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}) arranged in an LH with the same parameter range described for the 1P set. The initial random seed is also different for each simulation.

(4) The EX set (EX for extreme) contains four simulations with the same initial random seed and fiducial cosmological parameters $\Omega_m = 0.3$ and $\sigma_8 = 0.8$. One simulation has the fiducial astrophysical parameters, one has extreme SN feedback with $A_{SN1} = 100$, one has extreme kinetic AGN feedback with $A_{AGN1} = 100$, and the last one has no SN or AGN kinetic feedback with $A_{SN1} = A_{AGN1} = 0$.

We note again that the CV, 1P, and EX sets of ASTRID share the same realization of the initial conditions as their counterparts used in the TNG and SIMBA suites (e.g., the three simulation suites share the same initial conditions for CV0), while the LH and SBOb sets (see Section 2.3) use a separate set of random seeds to generate the initial conditions.

As in the CAMELS-TNG and CAMELS-SIMBA suites, each simulation in the CAMELS-ASTRID suite evolves a periodic box of comoving volume equal to $(25 \, h^{-1} \, \text{Mpc})^3$ from $z = 127$ to $z = 0$, with 256^3 dark matter particles and 256^3 gas particles in the initial conditions. In the core simulation sets described above, each simulation shares the cosmological parameters of $h = 0.6711$, $n_s = 0.9624$, $M_\nu = 0.0 \, \text{eV}$, $\omega = -1$, $\Omega_K = 0$ and $\Omega_b = 0.049$.

The simulations in the CAMELS-ASTRID suite each contain 91 snapshots from $z = 15$ to $z = 0$, covering the time stamps in the TNG and SIMBA suites but with higher time resolution for merger tree generation. For each snapshot, dark matter halos and subhalos/galaxies are identified using the SUBFIND (Springel et al. 2001), Rockstar (Behroozi et al. 2013a), and AMIGA (Knollmann & Knebe 2009) halo finders. Moreover, we also have merger trees generated from Rockstar (Behroozi et al. 2013b) and SubLink (Srisawat et al. 2013) available for each simulation.

2.3. CAMELS Extension Simulation Sets

The core CAMELS hydrodynamic simulation sets have six cosmological and astrophysical parameters varied, covering the key parameter space of interest with 1000 samples in the LH sets of ASTRID, TNG, and SIMBA suites. However, there are still many unexplored parameters in cosmology and astrophysical models that can have a significant impact on many observational properties and can have nonnegligible consequences for cosmological studies. Sampling the higher-dimensional parameter space with hydrodynamic simulations is rather expensive. Here we introduce the recently developed extension simulation sets in CAMELS that serve as the first stepping stone for extended parameter studies. Table 3 gives a summary of the extension simulation sets based on the ASTRID, TNG, and SIMBA models.

TNG Extension: We introduce here the TNG-SB28 set, which is an analog of the original TNG-LH set except where not 6 but 28 parameters of the model are varied simultaneously. They represent a semicomplete account of the free parameters in the TNG model, and they include 5 cosmological parameters, 2 parameters concerning star formation and the interstellar medium (ISM), 2 parameters related to stellar population modeling, 10 parameters controlling galactic wind

Table 1
Summary of the Physical Meaning of the Four Astrophysical Parameters (A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}) in the ASTRID, TNG, and SIMBA Suites

Simulation	A_{SN1}	A_{SN2}	A_{AGN1}	A_{AGN2}
ASTRID	Galactic winds: energy per unit SFR [0.25–4.00]	Galactic winds: wind speed [0.50–2.00]	Kinetic mode BH feedback: energy per unit BH accretion [0.25–4.00]	Thermal mode BH feedback: energy per unit BH accretion [0.25–4.00]
TNG	Galactic winds: energy per unit SFR [0.25–4.00]	Galactic winds: wind speed [0.50–2.00]	Kinetic mode BH feedback: energy per unit BH accretion [0.25–4.00]	Kinetic mode BH feedback: ejection speed / burstiness [0.50–2.00]
SIMBA	Galactic winds: mass loading [0.25–4.00]	Galactic winds: wind speed [0.50–2.00]	QSO & jet mode BH feedback: momentum flux [0.25–4.00]	Jet mode BH feedback: jet speed [0.50–2.00]

Note. The fiducial parameter value in each simulation is normalized to $A_{SN1} = A_{SN2} = A_{AGN1} = A_{AGN2} = 1$. The variation of each parameter is also shown in each cell. We note that the range of A_{AGN2} in the ASTRID suites is different from that in TNG and SIMBA, as A_{AGN2} in the ASTRID suite represents energy flux, similar to A_{AGN1}.

Table 2
Summary of the Core ASTRID Simulation Sets, which Share the Same Design as CAMELS-TNG and CAMELS-SIMBA, as Described in Villaescusa-Navarro et al. (2021a, See Their Table 2)

Name	Hydrodynamic Method	Code	Simulation Set	Number	Varying Parameters
ASTRID	Pressure–Entropy SPH	MP-Gadget	CV	27	S
			1P	61	Ω_m, σ_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}
			LH	1000	Ω_m, σ_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}, S
			EX	4	A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}

Note. A_{SN1}, A_{SN2}, A_{AGN1}, and A_{AGN2} represent the values of subgrid physics parameters controlling stellar and AGN feedback. S is the random seed of the initial conditions of a simulation. The LH set is a Latin hypercube where the values of $\{\Omega_m, \sigma_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}, S\}$ are varied simultaneously. Note that the parameters in the LH are different between ASTRID, TNG, and SIMBA. Simulations in the 1P set have the same initial random seed and vary only one of the parameters at a time. Simulations in the CV set have fixed values of the cosmological and astrophysical parameters (at fiducial values) but different initial random seeds. The EX set has the simulations with the same cosmological parameter and random seeds while exploring the extreme values in the four astrophysical parameters. See text for further details.

Table 3
Summary of the Additional Simulation Sets in CAMELS (Based on ASTRID, TNG, and SIMBA) for Extended Parameter Studies

Name	Simulation Set	Number	Varying Parameters
ASTRID Extension	ASTRID-SBOb	1024	$\{\Omega_m, \sigma_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}\} + \Omega_b$, S
	ASTRID-1P-Ob	8	Ω_b
TNG Extension	TNG-SB28	1024	$\{\Omega_m, \sigma_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}\}$ + 22 additional parameters, S
	TNG-1P-28	88	22 additional parameters (details in Appendix A)
SIMBA Extension	SIMBA-1P-28	88	22 additional parameters (details in Appendix A)

Note. The ASTRID-SBOb set is the extension of the ASTRID-LH set that also has Ω_b varied in addition to the original varied six parameters, where the seven parameters are sampled in a Sobol sequence. TNG-SB28 set is the extension of the TNG-LH set that explores the variation of 22 additional cosmological and astrophysical parameters, with individual parameter variation studied in the TNG-1P-28 set. The SIMBA-1P-28 simulation set presents individual parameter variations in the SIMBA model for 22 additional cosmological and astrophysical parameters. See text for further details.

feedback, 3 parameters governing the growth of SMBHs, and 6 parameters describing AGN feedback. We provide concise yet well-defined descriptions of each of the 28 parameters, including their range of variation, in Appendix A.1.

Another, minor difference between TNG-SB28 and the original TNG-LH is that the 28-dimensional parameter space is sampled using a Sobol sequence rather than an LH. Both methods provide a quasi-uniform (or low-discrepancy), yet irregular, sampling of the space, but compared to the LH, the Sobol sequence is deterministic, is faster to compute, has a lower discrepancy, and more easily allows the potential addition of newer, future samples.

In addition to the semi-uniform sampling of the 28-dimensional parameter space, we introduce another new simulation set, TNG-1P-28, which is composed of simulations that vary one parameter at a time, iterating over each of the 22 new parameters beyond the original six of the TNG-LH set. In TNG-1P-28, there are four simulations around the fiducial one for each parameter, two where the value of the parameter is lower than the fiducial value and two where it is higher.[13] The minimum and maximum values of each parameter are the same between TNG-1P-28 and TNG-SB28. The spacing of values for each parameter in TNG-1P-28 is uniform either in linear or in logarithmic space, and that distinction translates directly to whether the sampling of values in the Sobol sequence of TNG-SB28 is performed in linear or logarithmic space.

The range of variation of each parameter was chosen based on a combination of two (sometimes competing) considerations: (i) a physical intuition into its realistic range of values, with an attempt to err on the side of a large range in order to avoid edge effects; and (ii) an empirical examination of its effects on the simulation results, with a (very) rough goal of having the variations of different parameters resulting in effects of similar magnitudes. Appendix A presents the effects of the various parameters on a select number of physical quantities from the simulation results based on the TNG-1P-28 set. It is important to note that for any given physical quantity some model parameters are significantly more influential than others. We have verified, however, that each of the 28 parameters has a significant impact on at least some physical quantities that can be considered as basic to any cosmological galaxy formation model.

ASTRID Extension: Apart from the fiducial hydrodynamic simulation sets, the ASTRID suite also contains an additional SBOb set that adds one more degree of freedom by also varying Ω_b in addition to the six parameters in the ASTRID-LH set. ASTRID-SBOb contains 1024 simulations with the value of the cosmological and astrophysical parameters (Ω_m, σ_8, Ω_b, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}) arranged in a Sobol sequence with the same parameter ranges as in the 1P set. The variation range of Ω_b is $\Omega_b \in [0.01, 0.09]$. Unlike the TNG-SB28 set that broadly (and sparsely) samples 28 parameters in the simulation model, the main focus of the ASTRID-SBOb set is to explicitly disentangle the effect of Ω_b and Ω_m for cosmological parameter inference based on baryonic observables (e.g., Villaescusa-Navarro et al. 2022a; de Santi et al. 2023; Shao et al. 2023). The initial random seed is also unique for each simulation.

---

[13] There is one exception to this rule, when the fiducial value is zero and negative values are unphysical; hence, the four variations all use larger values than the fiducial. See details in Appendix A.1.
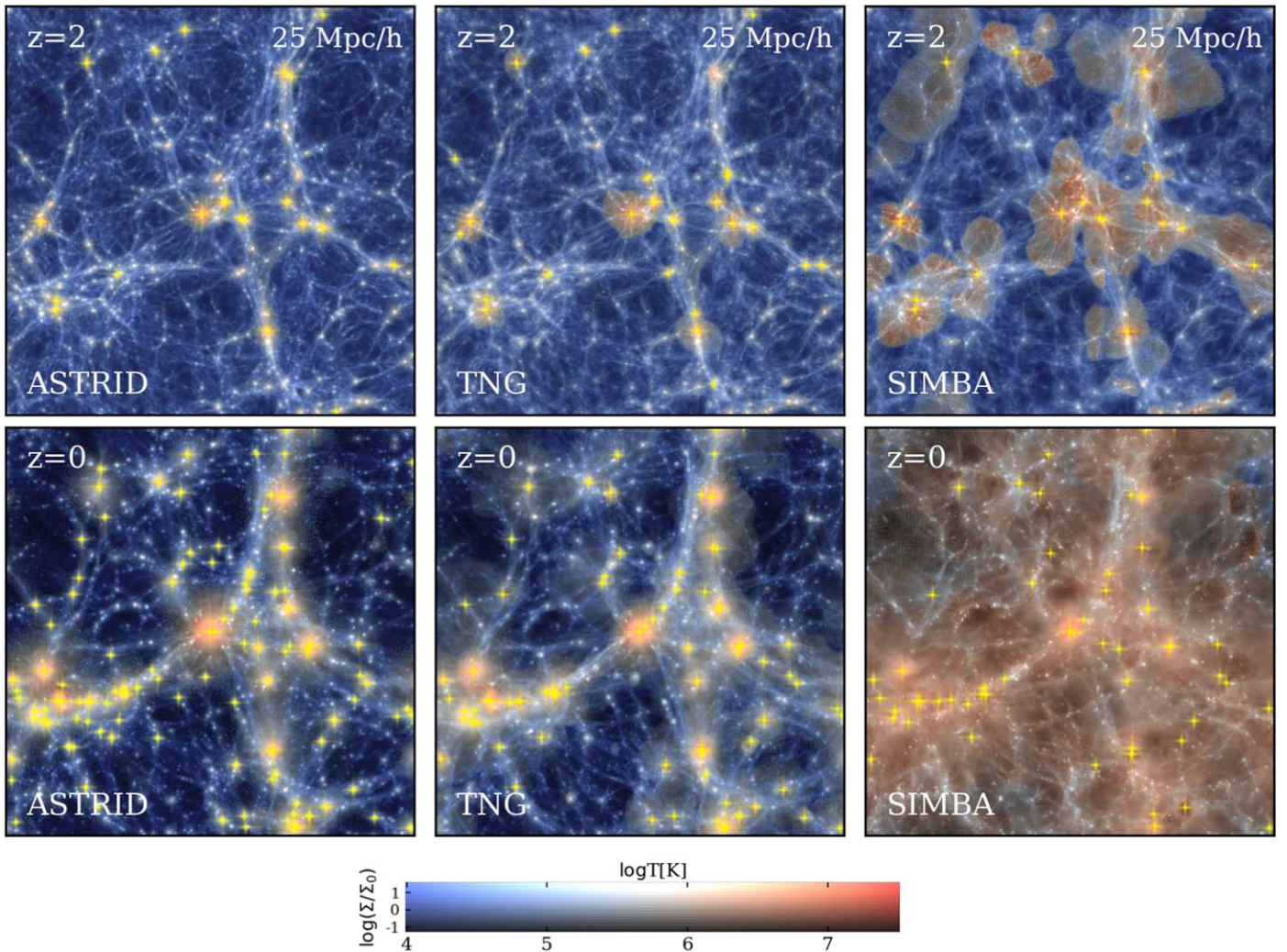
Figure 1. Illustration of the gas field of CV0 set for ASTRID (the first column), TNG (the second column), and SIMBA (the third column), centered on the most massive halo in the simulation. Each panel shows the gas density field colored by temperature (blue to red indicating cold to hot, respectively, as indicated by the 2D color map) over the full box volume of $(25\ h^{-1}\ \mathrm{Mpc})^3$ at $z = 2$ (top panels) and $z = 0$ (bottom panels). The yellow spikes mark the locations of massive BHs with $M_{\mathrm{BH}} > 10^8\ M_\odot$.

SIMBA Extension: In analogy with the TNG-1P-28 set, we introduce the SIMBA-1P-28 simulation set to investigate the impact of 22 new cosmological and astrophysical parameters beyond the six-parameter variations in the original CAMELS-SIMBA suite. The SIMBA-1P-28 set consists of 88 simulations that vary one parameter at a time, including four simulations around the fiducial model for each of the 22 new parameters (two simulations decreasing the parameter value and two simulations increasing the parameter value relative to the fiducial value). In full, the SIMBA-1P-28 suite includes variations of five cosmological parameters, three parameters controlling star formation and the ISM, eight parameters controlling galactic winds driven by stellar feedback, six parameters controlling the growth of SMBHs, and six parameters controlling AGN feedback. We describe the parameters in detail and their range of variation in Appendix A.2. We illustrate the effects of the various parameter variations on different physical quantities in Appendix C.

3. Comparison of the Three Fiducial Models

We start with a comparison between the fiducial models of the ASTRID, TNG, and SIMBA suites, to demonstrate the systematic differences between the three hydrodynamic simulations brought by their respective subgrid physical models. We highlight one of the most distinctive and interesting differences among the three simulation models: large-scale gas properties and correspondingly the influence of baryonic feedback on the matter power spectrum.

Figure 1 shows the large-scale gas density field colored by the temperature at $z = 2$ and $z = 0$ from the CV0 simulation, where the three simulation suites share the same initial conditions and adopt their corresponding fiducial astrophysical models.[14] One can see the clear differences with respect to the gas thermal properties of the intergalactic medium (IGM), where SIMBA heats up the gas surrounding large halos starting from higher redshift ($z \gtrsim 2$), producing hotter IGM gas compared to ASTRID and TNG. The striking difference in the large-scale gas field is mostly caused by the AGN feedback models, as in SIMBA AGN feedback turns on earlier and deposits kinetic and thermal energy on larger scales with

---

[14] See https://www.youtube.com/watch?v=zfxBa_Zp6WM for a movie visualization of the systematic difference between TNG, SIMBA, ASTRID, and Magneticum in the large-scale gas field.
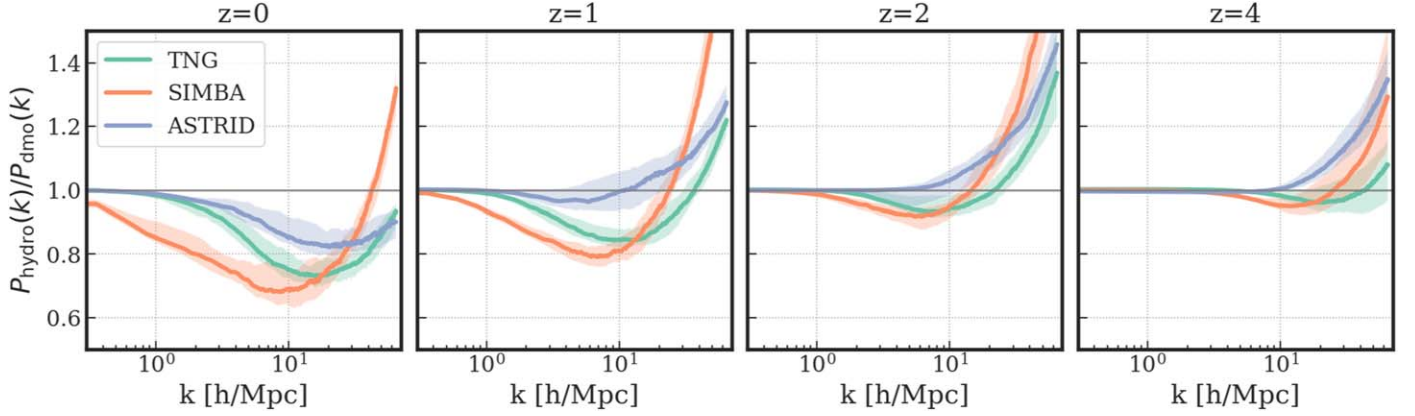
Ni et al.

Figure 2. Illustration of baryonic feedback on the total matter density field in the fiducial model of the three simulation suites. The x-axis is the k mode in comoving units, and the y-axis in each panel gives the ratio of the total matter density power spectrum in hydrodynamic simulations relative to that of the corresponding N-body simulations for the CV set, at redshifts $z = 0, 1, 2, 4$. Green, orange, and blue lines represent the results from TNG, SIMBA, and ASTRID separately. The solid line and shaded regions give the median and 10th–90th percentiles from the 27 simulations in the CV set.

fast jet outflows (see Davé et al. 2019; Borrow et al. 2020; Christiansen et al. 2020).

To illustrate where the most efficient AGN feedback takes place, Figure 1 marks the position of massive BHs with $M_{BH} > 10^8 M_{\odot}$ in yellow spikes. The bipolar lobe of hot gas in SIMBA at $z = 2$ clearly indicates the imprint of AGN jet mode feedback. On the other hand, TNG and ASTRID adopt a more localized (and also noncollimated) AGN kinetic feedback model for massive BHs at low Eddington ratio and therefore exhibit less impact on the gas field on large scales. Compared to TNG, ASTRID AGN kinetic feedback is even milder and turns on later (as it hinges on larger BH masses $M_{BH} > 5 \times 10^8 h^{-1} M_{\odot}$), and therefore it has the least impact on the large-scale gas properties among the three simulations. The gas properties have been studied previously in the work of Butler Contreras et al. (2023) for warm–hot IGM gas and Tillman et al. (2023) for the Lyα forest using the TNG and SIMBA suites, and these will be followed up with new simulation suites of ASTRID.

We compare the fiducial models of ASTRID, TNG, and SIMBA and diagnose their baryonic feedback impact across different redshifts. In Figure 2, we quantify the baryonic feedback of the fiducial models by calculating the ratio of the total matter power spectrum from the hydrodynamic simulations to that from their respective dark-matter-only simulations in the CV set, where the lines show the median of P_{hydro}/P_{dmo} for the CV set and the shaded regions give the 10th–90th percentile in each k bin.

The P_{hydro}/P_{dmo} ratio clearly shows the systematic differences in baryonic feedback between the three simulation suites with their fiducial astrophysical models. Compared to the collisionless dark-matter-only simulation, gas cooling and star formation enhance matter clustering on small scales, while the astrophysical feedback processes redistribute matter (especially the baryonic gas) and suppress clustering (as in Springel et al. 2018; van Daalen et al. 2020; Delgado et al. 2023). These two competing processes leave imprints on different scales and together shape the suppression and enhancement features in the P_{hydro}/P_{dmo} ratio as seen in Figure 2.

Among different astrophysical feedback processes, the AGN jet mode is the most important channel for efficiently redistributing the gas on large scales. Therefore, among the three simulation suites, SIMBA with its most aggressive AGN

feedback (which launches high-speed gas outflows and deposits energy at large distances that can reach hundreds of kiloparsecs away from the AGN host halos) produces the most prominent suppression of the matter power on large scales. On the other hand, ASTRID exhibits the least impact on the matter power spectrum as a consequence of having the mildest AGN feedback among the three simulation suites. The power spectrum ratio at $z = 0$ can reach $<70\%$ at $k \sim 10 h^{-1}$ Mpc in the SIMBA model while only 90% in ASTRID at the same k scale.

The baryonic impact on the matter power spectrum in the three simulation models exhibits a similar trend in time evolution. At higher redshifts ($z > 2$ in this small volume) when the majority of BHs have not grown massive enough to launch jets, baryons mostly enhance the matter power as a result of gas cooling and galaxy formation. When going to lower redshifts, the suppression of power emerges and propagates to large scales with time, as BHs become more massive and produce more powerful feedback that can affect the matter distribution on larger scales. The onset of matter power suppression is earliest in SIMBA and latest in ASTRID, mainly due to the different AGN feedback models.

4. 1P Sets: The Effects of Individual Parameters

In this section, we diagnose the effect of the six fiducial cosmological and astrophysical parameters on a selection of the key physical quantities and compare the three simulation suites. In particular, we use the 1P set of the ASTRID, TNG, and SIMBA suites to investigate how varying one parameter (with other parameters fixed) would change the matter power spectrum, the global star formation rate (SFR) history, and the galaxy and BH populations.

All the simulations in the 1P set have the same initial conditions and therefore diminish the effects of cosmic variance in the comparison. We note that the parameter study in this section focuses on the six fiducial parameters (Ω_m, σ_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}) that are varied in the LH sets of ASTRID, TNG, and SIMBA. In Appendices B and C, we show the effect of 22 additional parameters varied in the TNG and SIMBA through the extended 1P sets of TNG-1P-28 and SIMBA-1P-28.
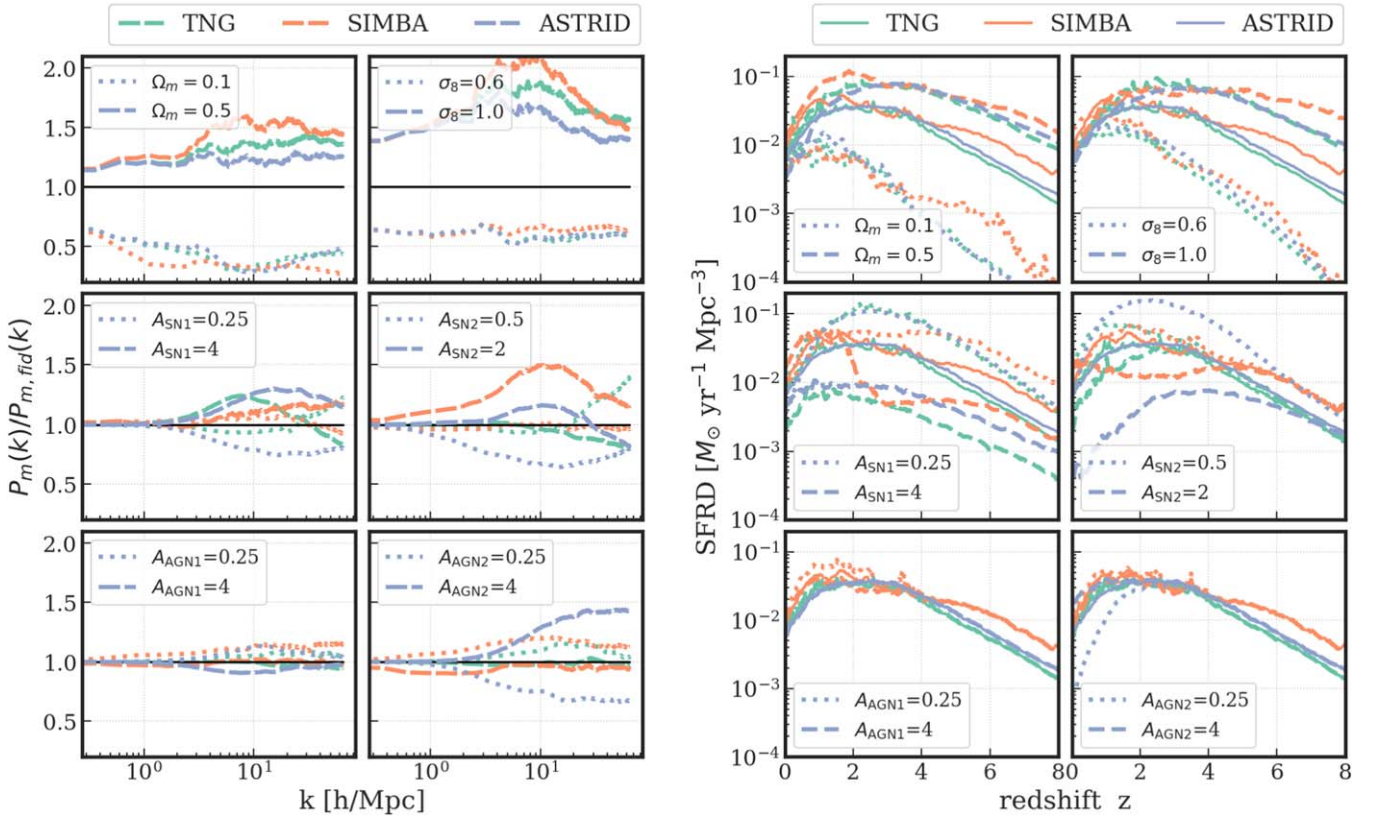
Figure 3. Left panels: ratio of the matter power spectrum of the 1P simulations (where only one parameter is varied and other parameters are fixed) to that of the fiducial model. The matter power spectra are evaluated at $z = 0$. Green, orange, and blue represent the results from TNG, SIMBA, and ASTRID, respectively. The solid and dotted lines indicate the simulations with the highest and lowest parameter value, respectively. Note that the value of A_{AGN2} is varied between 0.25 and 4.00 for ASTRID and between 0.5 and 2.0 for TNG and SIMBA. Right panels: global SFRD of the 1P simulations. Solid, dotted, and dashed lines represent the simulations with the fiducial, lowest, and highest parameter values in the variation range, respectively.

4.1. Matter Power Spectrum

The left panels of Figure 3 show the response of the matter power spectrum to the variation of the cosmological and astrophysical parameters individually. The fiducial model represents a simulation run with parameters $\{\Omega_m, \sigma_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}\} = \{0.3, 0.8, 1, 1, 1, 1\}$.

As expected, one can see that the matter power spectrum on large scales (with $k \lesssim 1\,h\,\mathrm{Mpc}^{-1}$) is mainly affected by the variation of cosmological parameters (Ω_m and σ_8), while the impact of the astrophysical parameters is limited to smaller scales of $k > 1\,h\,\mathrm{Mpc}^{-1}$ (with the exception of SIMBA, where the astrophysical parameters can mildly impact large scales owing to long-range AGN jet feedback).

We note that with the same variation in cosmological parameters the matter power spectrum responds similarly on large scales but differently on small scales between the three simulation suites. For example, with increased Ω_m or σ_8, ASTRID predicts less relative enhancement in the small-scale matter power compared to TNG and SIMBA. This is due to the different astrophysical models in each simulation suite that interact differently with the varied cosmological parameters. Notably, compared to a smaller Ω_m (or σ_8), a larger Ω_m (or σ_8) promotes the (earlier) formation of more massive halos, galaxies, and BHs, bringing nonlinear effects on the small-scale matter power into play earlier, and resulting in the different small-scale $P(k)$ variations among the three simulation models. The interaction between cosmological parameters and astrophysical processes again stresses the importance of

including different astrophysical models to improve the robustness of cosmological inference.

Variations in all four astrophysical parameters affect the amplitude and shape of the matter power spectrum in a nontrivial way. In ASTRID, A_{AGN1} exhibits the weakest influence among all four astrophysical parameters. This is mainly due to the limited volumes of the simulations, which do not form the massive objects containing the most massive BHs that can trigger AGN jet feedback. The same trend is also seen in TNG and SIMBA, as all three simulation suites adopt a M_{BH}-dependent AGN jet feedback model that biases jet feedback to more massive BHs. Given a larger volume, we might expect to see a more prominent impact of variations in AGN jet feedback on both galaxy formation and matter power spectrum.

The influence of SN and AGN feedback parameters is different between the three simulation suites, as they do not necessarily share the same physical meanings (as noted in Table 1). Even for A_{SN2}, which represents the speed of hydrodynamically decoupled galactic winds in the three simulation suites, its impact on the matter power spectrum (and on galaxy formation, as will be discussed later) can be quite different owing to other "nuisance" parameters in detailed model implementations, the interplay with other astrophysical processes, and the nonlinear evolution of galaxy formation.

It is expected that AGN feedback should make the dominant contribution to the baryonic suppression of the total matter power spectrum. Therefore, increased AGN feedback should lead to more suppression of the matter power. However, we
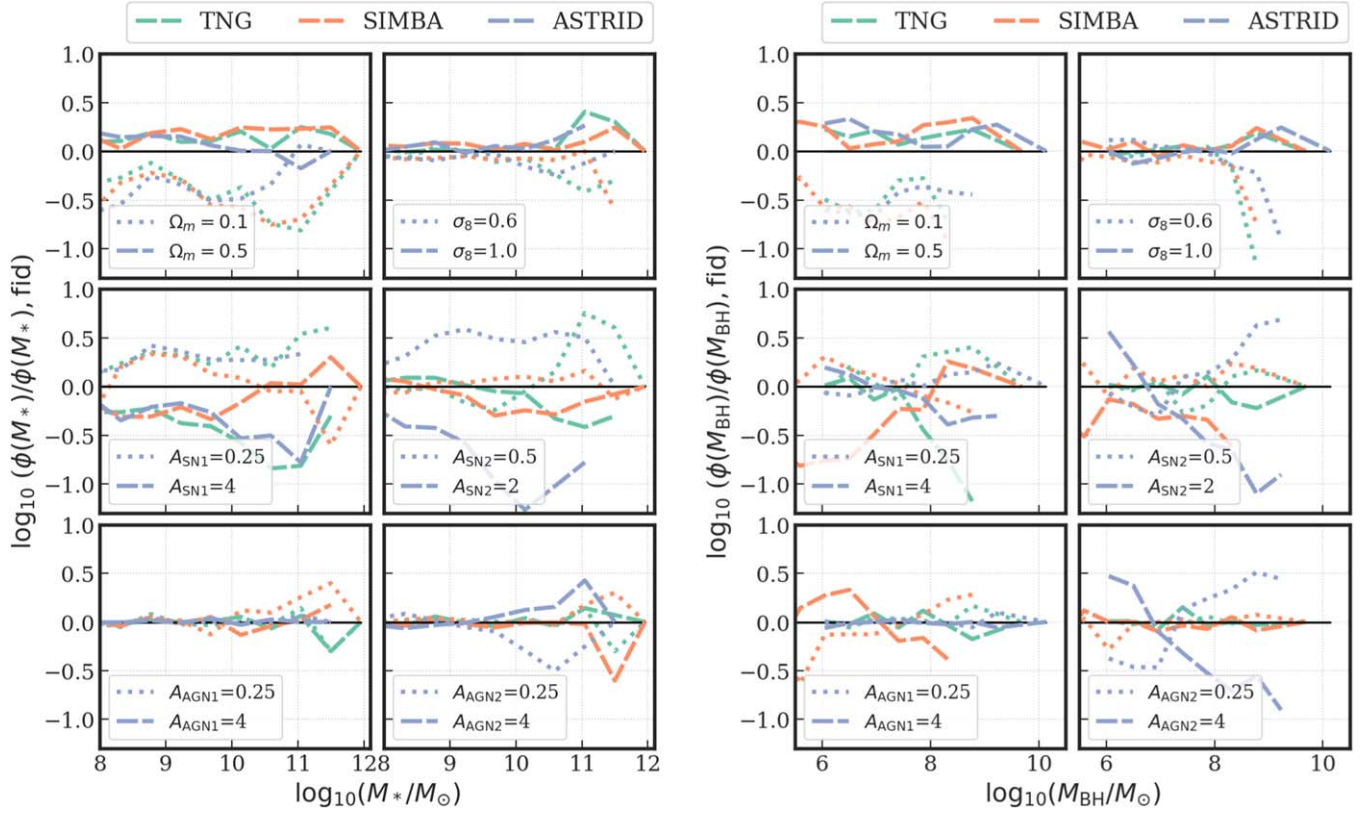
Figure 4. Impact of cosmological and astrophysical parameters on the GSMF (left panel) and BHMF (right panel). Both mass functions are evaluated at $z = 0$. The y-axis gives the ratio between the mass function of the 1P simulation (with only one parameter varied and other parameters fixed) and the mass function of the fiducial run. The convention of colors and line styles is the same as that in the left panel of Figure 3.

note that the total amount of AGN feedback cannot be simply controlled by the A_{AGN} feedback parameters, as the growth of galaxies and BHs is strongly regulated by feedback.

One example is the behavior of the A_{AGN2} parameter in the ASTRID suite. A_{AGN2} in ASTRID controls the efficiency of AGN thermal feedback, and we can see that a larger A_{AGN2} enhances (rather than suppresses) the matter power spectrum on small scales. As will be shown in Figure 4, this is because strong AGN thermal feedback aggressively suppresses the growth of the BH population and ends up curtailing the total amount of injected AGN feedback energy. In particular, by limiting the formation of the massive BHs that are able to turn on jet mode feedback, strong AGN thermal feedback prevents an efficient suppression of the matter power. Therefore, we note that varying the feedback efficiency induces nonlinear effects in the galaxy and BH populations and eventually the matter power spectrum.

4.2. Star Formation Rate Density

The right panel of Figure 3 shows the change in the SFR density (SFRD) due to individual variations of the six cosmological and astrophysical parameters. At high redshifts, $z > 2$, the SFRD is significantly affected by changes in Ω_m, σ_8, and also A_{SN1}, while at lower redshifts the SFRD begins to be noticeably impacted by the A_{SN2} parameter in all three simulation suites and by A_{AGN2} in ASTRID.

We note that the A_{SN2} parameter in the ASTRID suite drives larger variations in star formation (also at higher redshifts) compared to TNG and SIMBA. A_{SN2} modulates the speed of hydrodynamically decoupled galactic winds in all three

simulation suites. However, the detailed SN wind models and the fiducial parameters are different. Star formation in the ASTRID model turns out to be more sensitive to the SN wind speed than in TNG and SIMBA. This behavior leads to a broader variation in the galaxy population in the ASTRID-LH set compared to the LH sets of TNG and SIMBA.

The bottom left panels of Figure 3 indicate that the variation in AGN jet mode feedback (A_{AGN1} in all simulations and A_{AGN2} in TNG and SIMBA) has a limited impact on the global star formation. This result is somewhat affected by the small simulation volume, which limits the formation of the most massive systems that are subject to AGN jet feedback. The contribution of such systems to the global SFRD is, however, subdominant even without AGN feedback, except at late times, $z \lesssim 1$ (Springel & Hernquist 2003b). We can see that the SFRD in SIMBA exhibits a relatively larger response to A_{AGN1} compared to ASTRID and TNG at lower redshifts ($z < 2$). That is because the AGN jet mode in SIMBA hinges on a lower M_{BH} threshold ($M_{BH} > 10^{7.5} M_\odot$) compared to TNG and ASTRID and therefore allows an earlier onset of AGN jet feedback that can impact the SFR.

Interestingly, we find that the AGN thermal mode feedback (A_{AGN2} in the ASTRID suite) begins to impact the global star formation at low redshift ($z < 2$). The enhanced AGN thermal feedback efficiency boosts global star formation. This is partly a consequence of the nonlinear impact of AGN feedback and self-regulation of BH growth as discussed in Section 4.1, where a higher A_{AGN2} efficiency limits the growth of BHs and eventually curtails the total budget of AGN feedback energy deposited into the ISM. We note that this effect is also seen in the TNG model when varying the AGN thermal feedback efficiency in the

extended parameter study of TNG-1P-28, as shown in Appendix B (parameter 25, BlackHoleFeedbackFactor).

4.3. Galaxy and Black Hole Population at $z = 0$

We now study the impact of the variation of the six parameters on the mass functions of the galaxy and BH populations. To investigate the variation across the mass range, Figure 4 shows the ratio of the galaxy stellar mass function (GSMF) and BH mass function (BHMF) at $z = 0$ for each parameter variation compared to the fiducial model in the 1P set.

Figure 4 shows that the galaxy and BH populations respond to Ω_m in all mass intervals, while σ_8 has less impact compared to Ω_m and only affects the galaxy and BH populations at the most massive end at $z = 0$. As it sets the amplitude of the initial density fluctuations, σ_8 modulates the onset of gravitational collapse and structure formation. Due to hierarchical structure formation, in which small galaxies form earlier and merge to form more massive galaxies later, a lower σ_8 with delayed structure formation leaves an imprint on the abundance of massive galaxies, while the population of less massive galaxies has sufficient time evolution to grow down to $z = 0$. The impact of σ_8 is thus larger at higher redshifts. For example, at $z = 4$, the abundance of $M_* \sim 10^9 M_\odot$ galaxies can be 0.5 dex higher (lower) with $\sigma_8 = 1.0$ (0.6). Therefore, in general, it is more challenging to do parameter inference for σ_8 compared to Ω_m based on halo or galaxy populations at $z = 0$ in CAMELS (see, e.g., Shao et al. 2022; Villanueva-Domingo & Villaescusa-Navarro 2022; de Santi et al. 2023), due to the limited number of massive systems in the simulation volume.

The impact of the A_{SN1} parameter on the galaxy population exhibits a similar trend among the three simulation suites, with larger SN feedback efficiency leading to suppression of the galaxy abundance in all mass intervals. For the BH population, we note that a larger A_{SN1} greatly suppresses the low-mass BH population in SIMBA, as SIMBA adopts a BH seeding prescription based on the stellar mass threshold $M_* > 10^{9.5} M_\odot$ (Davé et al. 2019; Thomas et al. 2019; Habouzit et al. 2021), which is affected by SN feedback (while the TNG and ASTRID suites seed BHs based only on halo mass).

The effects of A_{SN2} in the ASTRID suite are more drastic compared to TNG and SIMBA, where a larger A_{SN2} significantly reduces the abundance of galaxies (especially at the massive end), as well as the population of massive BHs, as BHs accrete from the gas reservoir that also fuels star formation. We note that, due to the halo-based BH seeding model in CAMELS-ASTRID, the total number of BHs seeded is not affected by the astrophysical parameters. Therefore, in the cases where BHs do not grow efficiently, the BH population accumulates at the low-mass end.

The A_{AGN1} parameter, controlling the energy (in TNG and ASTRID) or the momentum flux (in SIMBA) of the AGN jet feedback, exhibits limited impact on the galaxy and BH populations, with the BH population in SIMBA showing a larger response to A_{AGN1}, as the jet mode feedback hinges on a lower M_{BH} threshold and can regulate BH growth at an earlier stage, as previously discussed in Section 4.2.

The A_{AGN2} parameter controls the BH thermal feedback efficiency in the CAMELS-ASTRID suite (as opposed to the jet speed in TNG and SIMBA). As the AGN thermal feedback is the dominant feedback channel that strongly regulates the growth of BHs before they become massive enough to turn on AGN jet feedback, A_{AGN2} in ASTRID shows a prominent impact on both the BH and galaxy populations. As discussed in earlier sections, a larger A_{AGN2} suppresses the formation of massive BHs, thereby curtailing AGN feedback on galaxy formation and producing an enhancement of the global SFR and galaxy populations in the ASTRID suite.

5. Overview of the Training Set Variation Range

In this section, we present an overview of some key cosmological and astrophysical properties for the five large hydrodynamic simulation training sets currently available in CAMELS: the three LH sets of TNG, SIMBA, and ASTRID (1000 simulations each, with six varied parameters), and the two extension sets of ASTRID-SBOb and TNG-SB28 (1024 simulations each, with additional varied parameters).

We note that many of these cosmological and astrophysical properties have been introduced in the first CAMELS presentation paper (Villaescusa-Navarro et al. 2021a) for the TNG and SIMBA-LH simulation sets. Here we revisit some of those quantities by comparing the fiducial astrophysical models of ASTRID with TNG and SIMBA and showing the range of variation of those properties covered by the five simulation sets. We present several one-dimensional statistics in Section 5.1 and various scaling relations between halos, gas, galaxies, and BHs in Section 5.2.

Various studies have been carried out to compare between CAMELS simulations and various observations, including Jo et al. (2023; the galaxy mass function and the SFR history), Moser et al. (2022; Sunyaev–Zeldovich signals from the circumgalactic medium), Parimbelli et al. (2023; emissions from warm–hot IGM), and Tillman et al. (2023; Lyα forests at low redshifts). In this section, we focus on illustrating the range of variations in different physical properties within the new training sets. We reserve detailed comparisons between the new simulation suites and observational data, as well as potential applications for model calibration, for future work.

5.1. 1D Statistics

We first look at the 1D statistics of the matter power spectrum, SFRD, and mass functions as shown in Figure 5. Note that for each of the panels in Figure 5 the solid lines give the median value in the CV sets from the fiducial astrophysical models of ASTRID, TNG, and SIMBA and the shaded regions give 10th–90th percentiles in the LH (or SB) sets to illustrate the variation range of the property considered. All quantities are measured at $z = 0$ except for the SFRD. We describe each quantity in more detail below.

Gas Power Spectrum: The first row of Figure 5 shows the power spectrum of the gas component $P_{gas}(k)$, which exhibits systematic differences between the three hydrodynamic simulation models. For the result of the CV set (the fiducial model), the clustering of gas is highest in ASTRID and lowest in SIMBA, with the difference reaching one order of magnitude at the scale of $k \sim 10\,h\,\mathrm{Mpc}^{-1}$. As already discussed in Section 3, among the three hydrodynamic simulation models, the fiducial version of ASTRID has the mildest AGN feedback and therefore leads to the lowest suppression of gas clustering.

In spite of this, the variation of the gas power in ASTRID is the largest among the simulation suites on small scales. At $k \sim 10\,h\,\mathrm{Mpc}^{-1}$, the gas power can vary across two orders of magnitudes in the ASTRID-LH set, covering the total range of SIMBA and TNG, and even TNG-SB28, which varies a
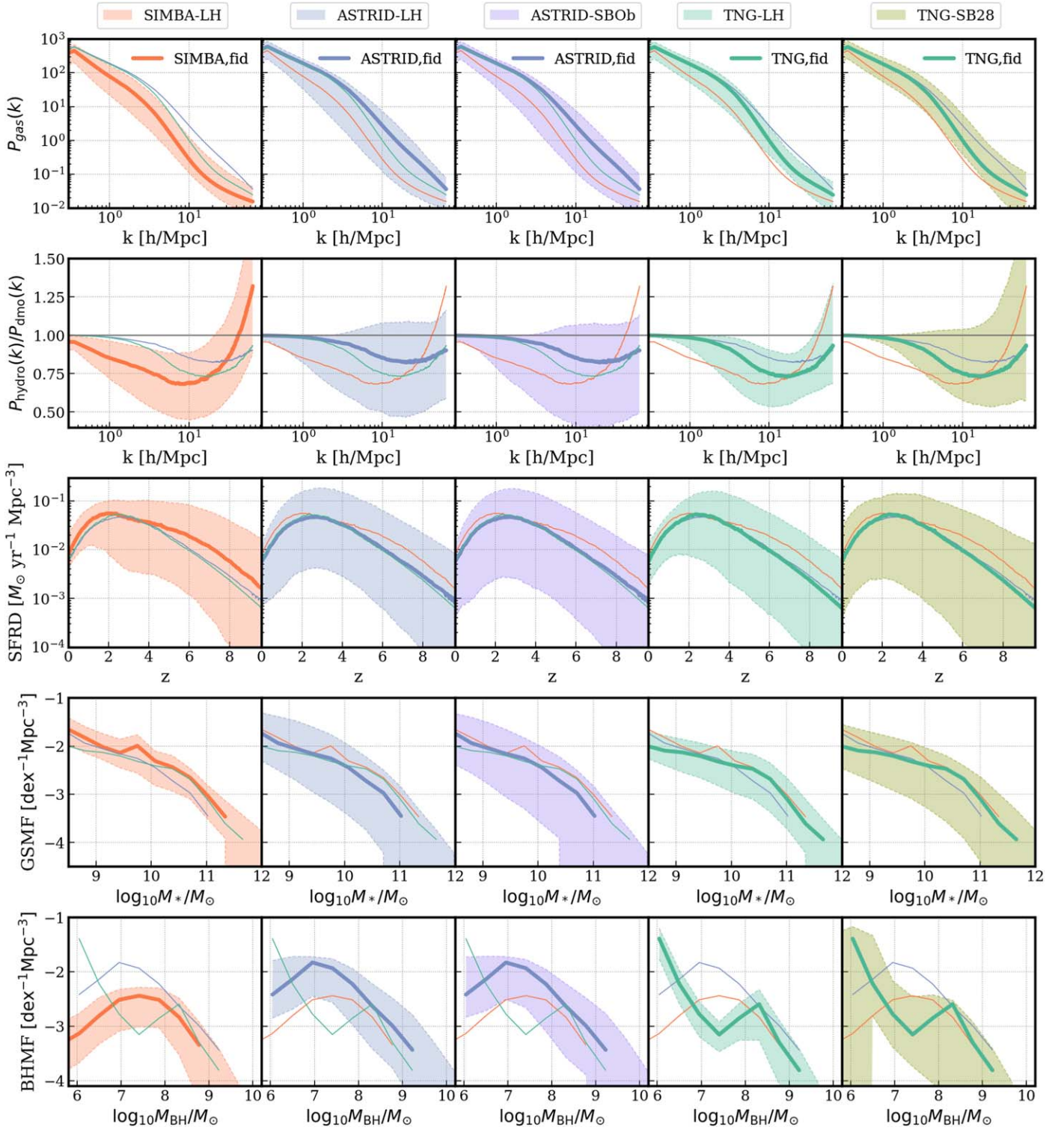
Figure 5. Illustration of the range of different cosmological and astrophysical properties in the five large simulation suites: SIMBA-LH (orange), ASTRID-LH (blue), ASTRID-SBOb (purple), TNG-LH (green), and TNG-SB28 (yellow green). From top to bottom the panels show the gas power spectrum, the ratio of the matter power spectrum in hydrodynamic simulations to that of the corresponding *N*-body simulations, the SFRD history, the stellar mass function, and the BHMF. The power spectra and mass functions are shown at $z = 0$. Note that the halo masses and stellar masses are from the SUBFIND catalogs. For each panel, the orange, green, and blue lines give the predictions from the fiducial models of SIMBA, TNG, and ASTRID, respectively, calculated by taking the median of the CV set. Each simulation suite has its corresponding fiducial model result emphasized in a thicker line. The shaded regions indicate 10th–90th percentiles from each large LH simulation set (or SB set for ASTRID and TNG extension).

significantly larger number of model parameters than ASTRID-LH. This large variation in the small-scale gas power is driven by the combination of A_{AGN2} and the two A_{SN} parameters, as can be inferred from Figure 3.

The similarity between ASTRID-LH and ASTRID-SBOb, and even between TNG-SB28 and TNG-LH, where the addition of 22 varied parameters increases the diversity of $P_{\mathrm{gas}}(k)$ responses only by a modest factor, does not demonstrate
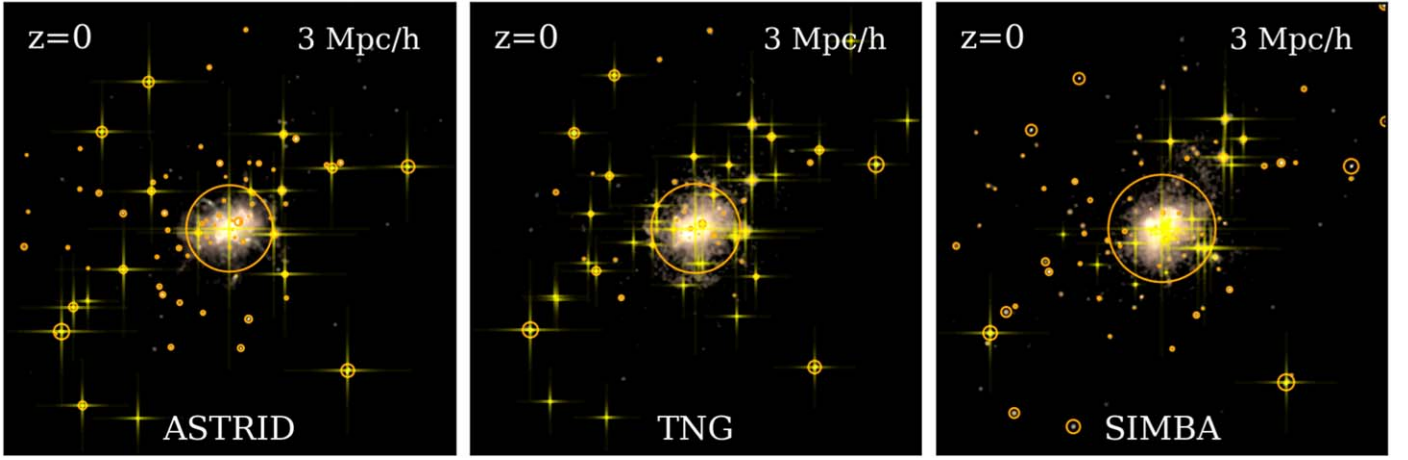
Figure 6. Illustration of galaxies and BHs in the CV0 simulation for ASTRID (first column), TNG (second column), and SIMBA (third column). Each panel shows the stellar density field over a region of $(3\,h^{-1}\,\mathrm{Mpc})^3$ centered on the most massive halo in the simulation. The orange circles mark galaxies with stellar mass $M_* > 10^8\,M_\odot$, as identified by SUBFIND. The radii of the circles correspond to the virial radii of the subhalos. The yellow spikes mark the positions of all BHs in this region, with the size scaled by M_{BH}.

that variations in Ω_b or any other of these additional parameters are necessarily subdominant for $P_{\mathrm{gas}}(k)$ with respect to the original six parameters of the LH sets. The $P_{\mathrm{gas}}(k)$ responses to the individual additional parameters are, in fact, in some cases no less significant than the original ones, as shown explicitly in Appendices B and C.

Instead, we interpret the rough similarity between the diversity of $P_{\mathrm{gas}}(k)$ results in the original and extended parameter spaces to be the result of a "regression toward the mean" effect, whereby the variations in a larger number of relevant parameters tend to roughly cancel each other out. Nevertheless, by successfully using these simulation sets in future work to learn the relations between the parameters and simulation results, such as $P_{\mathrm{gas}}(k)$, we will be able to probe the entirety of these extended parameter spaces, including regions therein that are far from the sampling points directly simulated here. We do expect this task, however, to become more challenging, as the sparsity of this sampling increases significantly owing to the "curse of dimensionality."

Matter Power Spectrum Ratio: We consider the total matter power spectra for each hydrodynamic simulation in comparison to the matter power spectrum of the corresponding dark-matter-only runs and give the value of $P_{\mathrm{hydro}}/P_{\mathrm{dmo}}$ in the second panel of Figure 5.

Compared to Figure 2, where the shaded regions give the cosmic variance driven by different realizations in the CV set, we can see that the variation of $P_{\mathrm{hydro}}/P_{\mathrm{dmo}}$ in the LH sets is significantly larger. The ASTRID-LH set can yield larger variation compared to the LH set of SIMBA and TNG, which reflects the similar trend pointed out above for $P_{\mathrm{gas}}(k)$. In some of the strong feedback scenarios, the power spectrum ratio of ASTRID can reach $\sim 50\%$ on small scales ($k \sim 10\,h\,\mathrm{Mpc}^{-1}$). On the other hand, some of the astrophysical models in the ASTRID-LH set can produce $P_{\mathrm{hydro}}/P_{\mathrm{dmo}} > 1$, which is not seen in the TNG and SIMBA suites. Simulations in those scenarios usually have a limited number of massive BHs (caused by, e.g., a large A_{AGN2} parameter), and the AGN kinetic (jet) feedback cannot efficiently counteract the small clustering from star formation. Therefore, those simulations predict enhancement rather than suppression of the total matter power.

Star Formation Rate Density: The third panel of Figure 5 shows the global SFRD for each of the simulation sets. The shaded regions indicate the 10th–90th percentiles for the given

redshift bin in the LH (or SB) sets. Comparing the three LH sets, the SFRDs in the ASTRID-LH set display larger variation at lower redshift $z < 2$ compared to the LH sets of SIMBA and TNG. We can see from the right panel of Figure 3 that the SFRD in ASTRID at low redshift is very sensitive to A_{SN2}. The same level of variation in A_{SN2} in the ASTRID suite drives a larger diversity of SFRDs compared to TNG and SIMBA. Furthermore, the SFRD at low redshifts is more sensitive to A_{AGN2} in ASTRID. Therefore, ASTRID-LH also features a larger variation in the star formation history, as well as the galaxy population at $z = 0$, compared to the other two LH sets.

With the additional parameters varied, the extension suites of ASTRID-SBOb and TNG-SB28 both show mildly larger variation in the SFRDs compared to their corresponding LH sets (more so for the larger extension represented by TNG-SB28 than that of ASTRID-SBOb, as might be expected). Our interpretation for this follows the ones we presented for $P_{\mathrm{gas}}(k)$ above.

Galaxy and Black Hole Population: The bottom two rows of Figure 5 present the GSMF and BHMF at $z = 0$. The stellar masses are summary statistics from the subhalo catalogs given by the SUBFIND algorithm, including the contribution of both host and satellite halos (galaxies). The statistics of galaxies and BHs exhibit systematic differences among the simulation suites owing to the different astrophysical models.

Figure 6 illustrates the galaxy and BH population at $z = 0$ in the three fiducial models of ASTRID, TNG, and SIMBA over a projected region of $(3\,h^{-1}\,\mathrm{Mpc})^3$ surrounding the most massive halo in the CV0 simulations. The orange circles mark the positions of galaxies with $M_* > 10^8\,M_\odot$, and the yellow spikes mark the locations of all BHs. The radii of circles and sizes of the spikes are scaled by M_* and M_{BH}, respectively.

The large-scale pattern of the galaxies is similar among the three simulations owing to the same initial conditions and therefore the resulting large-scale structure. Meanwhile, we can see differences in the population of small galaxies and BHs due to the various astrophysical models used in the three hydrodynamical simulations. Noticeably, the small satellite galaxy population is suppressed in TNG compared to ASTRID and SIMBA, as also indicated by the GSMF in Figure 5. On the other hand, small galaxies in SIMBA do not host BHs owing to its BH seeding model. Therefore, we can see that SIMBA

produces a smaller BH population and lower occupation fraction in galaxies compared to the other two simulations.

For the fiducial models, galaxy abundance in ASTRID is slightly lower than in TNG and SIMBA by ~ 0.3 dex in the stellar mass interval $M_* = 10^{10}$–$10^{11} M_\odot$. SIMBA features a spike in the galaxy mass around $M_* = 5 \times 10^9 M_\odot$. The TNG galaxy abundance is slightly lower at the low-mass end of $M_* < 10^9 M_\odot$ compared to the other two models.

The variation of the GSMF in the ASTRID-LH set is larger than the LH sets of TNG and SIMBA for the same reasons as discussed for the SFRD. At the low-mass end ($M_* < 10^{10} M_\odot$), the range of variation in ASTRID covers that of the SIMBA and TNG simulation suites, including TNG-SB28. The broad variation of the galaxy population in ASTRID-LH allows ML models trained on the ASTRID galaxy catalogs (of the LH set) to provide the best extrapolation results when applied to other simulation suites as test sets (e.g., N. Echeverri et al. 2023, in preparation).

Compared to the galaxy population, the BHMF shows much larger differences between the fiducial models of the three simulation suites, especially at the low-mass end of $M_{BH} < 10^8 M_\odot$, due to different prescriptions for BH seeding, accretion, and feedback (as well as poor observational constraints for model calibration; see also Habouzit et al. 2021). Compared to TNG and SIMBA, ASTRID has a larger BH population, especially at the low-mass end of $M_{BH} = 10^6$–$10^8 M_\odot$. The low-mass end BH population can be sensitive to the BH seed mass. We reiterate that the CAMELS-ASTRID suite does not adopt the same BH seeding prescription as the large volume ASTRID production run owing to the limited resolution of CAMELS, as discussed in Section 2.

The diversity of the BHMF as a result of parameter variations is larger in ASTRID than in the LH sets of the other suites, similarly to the case for the quantities discussed above. It is also true that the extension into Ω_b variations in ASTRID-SBOb does not noticeably increase the diversity compared to ASTRID-LH. However, in the case of the BHMF, TNG-SB28 displays a much larger diversity than TNG-LH, or even than the ASTRID sets, unlike what was seen above for other quantities. We interpret this as being a result of a high sensitivity of BH masses to a number of the new parameters introduced in TNG-SB28, which include specifically a significant number of parameters that directly control BH growth. This is as opposed to the six parameters in the original TNG-LH set, most of which made little difference for the growth of BHs.

5.2. Galaxy Catalogs and Scaling Relations

In this subsection, we inspect the halo and galaxy catalogs in the five simulation sets and compare various scaling relations of astrophysical properties.

Halo Baryon Fraction: The first row of Figure 7 shows the scaling relation between the baryon fraction $f_b = M_b / M_{halo}$ and M_{halo}, where M_{halo} is the FOF halo mass and M_b is the total baryonic mass (including the stars, gas, and BHs) in the FOF halo. We can see the systematic difference between the three simulation suites. The baryon fractions in the three simulation suites all peak at around $M_h \sim 10^{12} M_\odot$ and decrease when going to larger halo mass as a result of AGN (jet) feedback in massive halos, where ASTRID exhibits the largest baryon fraction for massive halos while SIMBA predicts the lowest. A smaller baryon fraction in halos corresponds to less clustering of gas (see, e.g., van Daalen et al. 2020; Delgado et al. 2022, 2023; Pandey et al. 2023, for more detailed studies).

Therefore, the baryon fraction in the three simulation suites exhibits a similar trend to the gas power spectrum as seen in Figure 5. Meanwhile, the ASTRID-LH set displays the largest variation in the baryon fraction within the same halo mass bin, almost covering the range of TNG and SIMBA, whether or not the LH or SB sets are considered.

Star Formation Rate versus Stellar Mass: The second row of Figure 7 shows the galaxy SFR versus the stellar mass of galaxies from the galaxy subhalo catalogs. Here SFR is the instantaneous star formation of all gas particles associated with their galaxy hosts. We only consider galaxies with nonzero SFRs to avoid the distributions being heavily affected by fully quenched galaxies. The three simulation suites exhibit a similar trend in that the SFR increases with stellar mass and flattens when going to more massive galaxies owing to quenching from feedback. As discussed in Villaescusa-Navarro et al. (2021a), in the low-M_* end, cosmic variance can make the dominant contribution to the variation in the SFR–M_* relation. Therefore, we see a similar range of the LH set variation for the three simulation suites for lower-mass galaxies at $M_* < 10^{10} M_\odot$. The variation of the SFR increases at higher galaxy masses. The SFR variation can exceed 3 orders of magnitude for $M_* > 10^{11} M_\odot$ in the three simulation suites, as the AGN feedback begins to quench the galaxies, therefore giving rise to a larger scatter in this mass regime by variations of the feedback parameters in the LH and SB sets.

Metallicity versus Stellar Mass: The third row of Figure 7 shows the relation between stellar metallicity, in units of solar metallicity ($Z_\odot = 0.012$), and stellar mass. The three simulation suites exhibit systematically different Z_*–M_* relations (or metal enrichment histories) owing to different astrophysical models. The fiducial model of TNG predicts the largest stellar metallicity in all stellar mass bins, while ASTRID gives Z_* between SIMBA and TNG. The variation in the ASTRID-LH set tends to be larger than that in the other two LH sets, especially for the massive galaxies, but TNG-SB28 shows a much higher diversity than both TNG-LH and ASTRID-LH. We believe that this is due to the particularly large effect of two of the additional parameters in TNG-SB28 on this relation, specifically the slope of the initial mass function (IMF) and the metal loading factor of the galactic winds.

Black Hole Mass versus Stellar Mass: The bottom panel of Figure 7 shows the M_{BH}–M_* relation. For each galaxy identified by SUBFIND, we consider the single most massive BH associated with that galaxy. We only include galaxies with nonzero BHs, which results in the cutoff at $M_* \lesssim 10^{9.5} M_\odot$ in SIMBA due to its M_* threshold for BH seeding.

M_{BH} in the ASTRID suite is systematically larger than that in TNG and SIMBA within a given stellar mass bin. For lower-mass galaxies $M_* < 10^{10} M_\odot$, ASTRID predicts M_{BH} about 0.5–1 dex more massive compared to TNG, as BHs in ASTRID grow more rapidly in the initial phase with boosted Bondi accretion factor $\alpha = 100$ and the possibility of super-Eddington accretion (the $\dot{M}_{BH}$ in ASTRID is capped at $2 \times L_{Edd}$). At higher M_{BH} (or M_*), BH growth starts to become self-regulated by AGN feedback. At higher masses $M_* > 10^{10} M_\odot$, ASTRID galaxies host M_{BH} about 0.4 dex more massive compared to TNG and SIMBA. We note again that the M_{BH} at $z = 0$ and the M_{BH}–M_* relation can be sensitive to the initial BH seed mass, where the CAMELS-ASTRID suite applies a different BH seeding prescription compared to the ASTRID production run because of the lower resolution of CAMELS.
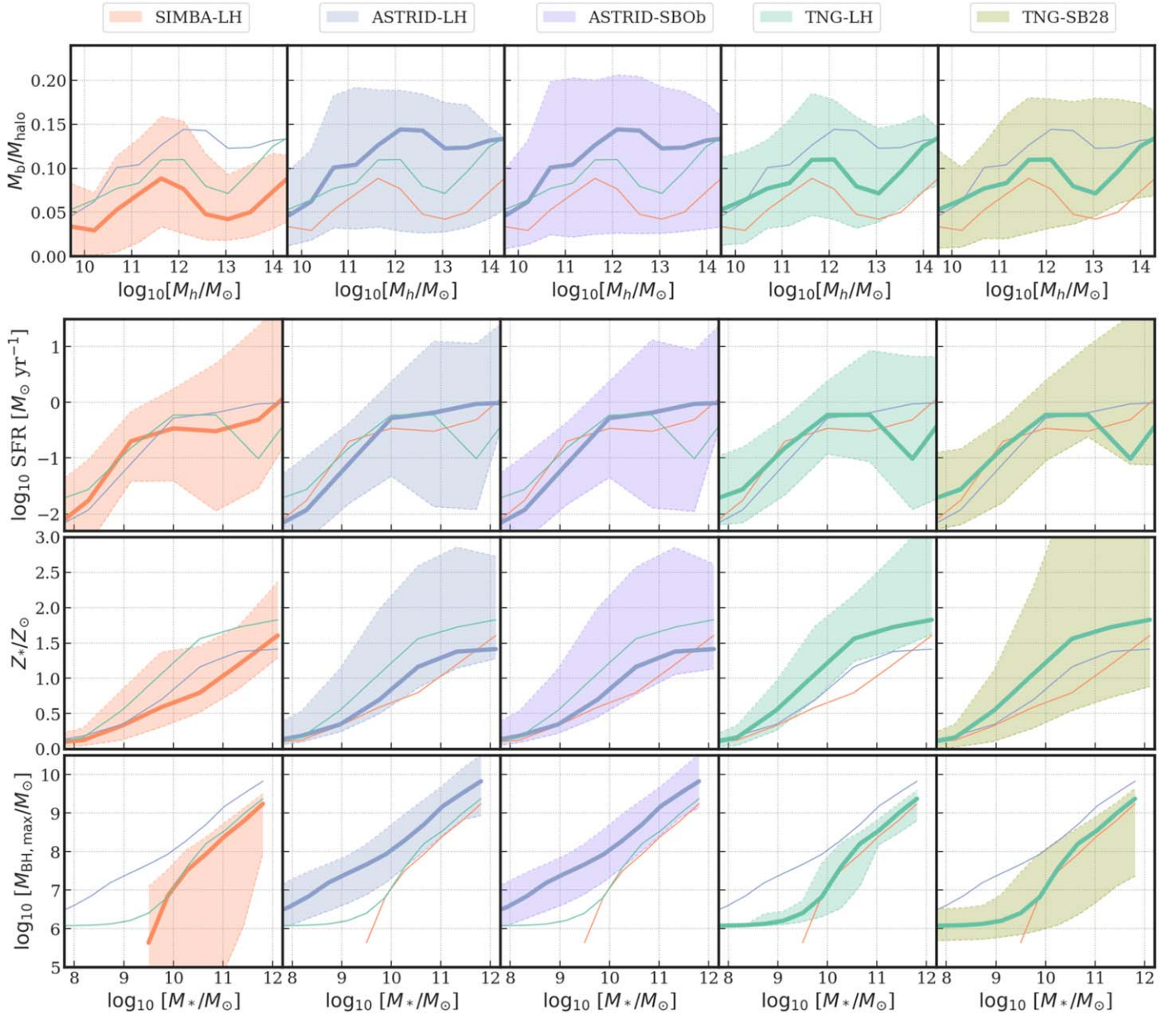
Figure 7. Various scaling relations between astrophysical properties from the (sub)halo catalogs at $z = 0$. Conventions for the colors, lines, and shaded regions are the same as in Figure 5. The top panel shows the halo baryon fraction vs. halo mass. The lower panels show the scaling relations between stellar mass and SFR (second row), mean stellar metallicity (third row), and the maximum $M_{\rm BH}$ (fourth row). For each property, we combine the subhalo catalogs from the CV set and LH set and show in solid lines the median value of y for each mass bin in x from the CV set and in the shaded area the 10th–90th percentile of the y distribution from the combined catalog from 1000 simulations of the LH set (or 1024 simulations of the SB set). See text for more details.

The range of variation in the resulting $M_{\rm BH}$–M_* relations when the simulation parameters are varied is quite distinct between the five different sets, with the exception of ASTRID-LH and ASTRID-SBOb, which are virtually the same. In particular, this diversity is a strong function of stellar mass, with different dependences on mass in the different sets. We see again, similarly to the BHMF, that the diversity in TNG-SB28 is much larger than in TNG-LH, which we associate with the addition of new parameters that directly affect BH growth and evolution.

6. Machine-learning Tests

In this section, we demonstrate several ML applications that can be performed using the new simulation suites introduced in this work. As a simple demonstration, we carry out ML tasks following our previous work on field-level likelihood-free inference of Ω_m and σ_8 using 2D maps of the total matter density field and gas temperature field. These 2D maps represent the projection of $25 \times 25 \times 5 \ (h^{-1}{\rm Mpc})^3$ slices of the simulation volume and have been created for the new CAMELS-ASTRID and TNG-SB28 simulations using the same procedure as presented in Villaescusa-Navarro et al. (2022b). These maps have been included in the CAMELS Multifield Dataset.[15]

In Section 6.1, we test previously developed ML models that are robust across the two subgrid models of TNG and SIMBA on the new ASTRID field. In Section 6.2, we demonstrate the
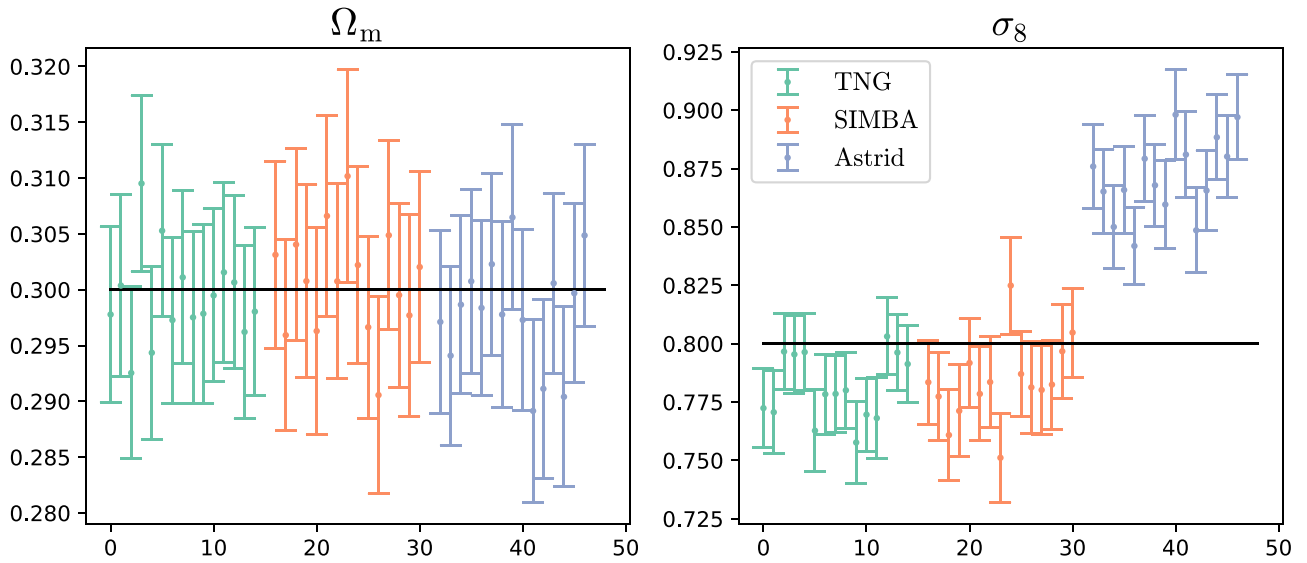
_____
[15] https://camels-multifield-dataset.readthedocs.io

Figure 8. We use the model described in Villaescusa-Navarro et al. (2021c), which was trained on total matter maps from CAMELS-TNG simulations, to conduct tests on 15 distinct maps with constant values of Ω_m and σ_8 (shown with black horizontal lines) from three different simulation suites: TNG (green), SIMBA (orange), and ASTRID (blue). As observed, the model performed effectively in inferring the value of Ω_m; however, it was not able to correctly infer σ_8 from CAMELS-ASTRID maps.

case of training ML models based on a combination of two simulation suites (LH sets of TNG + SIMBA) and testing them on the ASTRID-LH set. In Section 6.3, we present the first use of the new TNG-SB28 suite as a training set to showcase the possibility of developing ML inference models based on a simulation set embedded in a high-dimensional parameter space.

We note that in this section we use the new CAMELS-ASTRID suite to test model robustness and validate the performance of ML models trained on two simulation suites. ML models trained on the new ASTRID suite have already been presented in Shao et al. (2022), de Santi et al. (2023), and Pandey et al. (2023) and have exhibited good performance. In particular, de Santi et al. (2023) showed that, among all the available suites, ASTRID was the one with the best generalization properties when studying galaxy clustering, while N. Echeverri et al. (2023, in preparation) reached similar conclusions when studying properties of individual galaxies.

6.1. Robustness Tests

In this subsection, we investigate whether models that are robust across two simulation suites will also be accurate when tested on simulations from a third suite. For this, we take advantage of the results presented in Villaescusa-Navarro et al. (2021c), where it was shown that a convolutional neural network (CNN) model was able to perform a robust field-level likelihood-free inference on the values of Ω_m and σ_8 from 2D maps containing slices of the total matter density field.

The ML model is trained to predict the mean and standard deviation of the marginal posterior for each parameter (of Ω_m, σ_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}) from the input 2D maps. The architecture of the model consists of a set of convolutional layers followed by a fully connected layer. The training process uses the loss function modified from the moments network following Jeffrey et al. (2021). We refer to Villaescusa-Navarro et al. (2021b) for further details of the ML model.

The network was trained on maps created from the LH set of TNG simulations and tested on maps from both TNG and

SIMBA, showing that the approach was able to infer the values of Ω_m and σ_8 with comparable accuracy and precision. The same test was performed training the model on SIMBA maps yielding the same conclusion. Having a model that is robust across the TNG and SIMBA models, we investigate whether it is also robust when tested on data from ASTRID.

The result we find is that the TNG-trained model is able to infer the true value of Ω_m from the maps of all three simulations, while for σ_8 it is only able to perform such a task in the TNG and SIMBA maps, failing on the maps from ASTRID. This is demonstrated in Figure 8, which shows the results of the model trained on TNG and tested on the CV set of the three simulation suites (so the true values of Ω_m and σ_8 are 0.3 and 0.8, respectively).

To illustrate the model performance on the three simulation suites, Figure 9 gives an example of the total mass map from the TNG, SIMBA, and ASTRID CV set (with the same realization) as one of the test sets. From the visual comparison, as well as the residual plots shown in the bottom panels of Figure 9, one can see that the total matter maps of TNG and ASTRID are more similar to each other compared to the one from SIMBA. However, the network that can infer σ_8 for the TNG map (and for the SIMBA map) gives a biased result when tested on the ASTRID map. We also test the network on the CV set of the Magneticum simulation (which also contains 27 CV simulations) and find that the model also gives a biased result of σ_8 similar to that of ASTRID. We speculate that the network infers σ_8 contingent on some complex, high-order (and possibly numerical) features that happen to be shared between TNG and SIMBA while broken by ASTRID (and Magneticum) maps. We leave it to future work to interpret and understand such behavior and develop a more robust model for σ_8 inference.

We stress that the falsification of the model on the σ_8 inference task demonstrates the importance of testing ML models on a range of independent galaxy formation simulations to ensure their robustness and generalizability, as ML models that can work on two galaxy formation simulations are not guaranteed to perform well when tested on a third one.
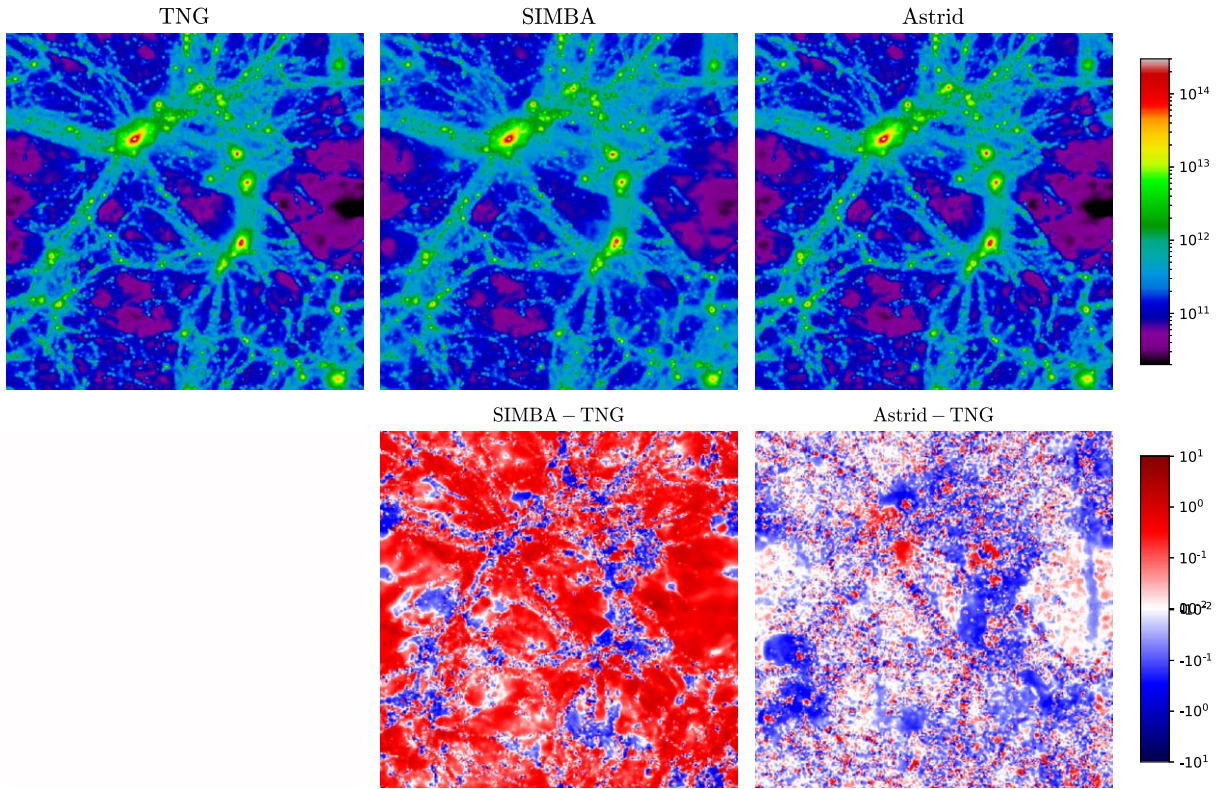
Figure 9. The top row shows three images with the total matter surface density from three different simulations run with the same initial conditions. The panels in the bottom row show differences with respect to the TNG image. From the residual plots, it can be seen that the images from the TNG and ASTRID models are more similar to the ones from SIMBA.

6.2. Training on Two Simulation Suites Combined

With a third suite of hydrodynamic simulations in CAMELS, we can train ML models on two simulation suites (i.e., based on two galaxy formation models) and test whether they can generalize to a third simulation suite that is carried out with a distinct galaxy formation model. We perform this task by training a CNN to infer Ω_m and σ_8 based on the 2D gas temperature map. By visual inspection, the gas temperature maps are one of the most divergent maps between the three CAMELS suites and therefore pose a challenging test for the robustness of ML inference models. As shown in Villaescusa-Navarro et al. (2021b), the ML model trained on TNG gas temperature maps does not generalize to SIMBA (and vice versa), due to the distinct astrophysical models utilized in each suite.

In this subsection, we train the ML model based on the combination of TNG and SIMBA maps and test that model on the ASTRID maps. We note that the ML model is only trained and tested on the 2D gas temperature maps from the LH sets of the three simulation suites, to link to our previous work. In Figure 10, the black points (with error bars) show the results of the ML model trained based on TNG+SIMBA maps and tested on TNG, SIMBA, and ASTRID individually. As a comparison, we also show the test results from the model trained only on TNG (green) and only on SIMBA (orange) that are presented in Villaescusa-Navarro et al. (2021b). To qualify the model performance, we calculate the rms error (RMSE) and reduced chi-squared (χ^2) to quantify the precision of the inference model and the accuracy of the estimated errors.

Based on Figure 10, and in particular the RMSE and χ^2 scores, it is clear that while the models trained only on one suite perform badly on the other suite (exceedingly so for σ_8), the model trained

on TNG+SIMBA performs well when tested on TNG or SIMBA. However, its performance degrades when tested on ASTRID, indicating that the model trained on two simulation suites does not generalize well when applied to a new galaxy formation model. Nevertheless, it is interesting to note that when tested on ASTRID the model trained on TNG+SIMBA (black) outperforms the models trained only on TNG (green) or SIMBA (orange). One might worry that when training a model based on two distinct suites of gas temperature maps it would categorize the input field as either TNG-like or SIMBA-like and make an inference that resembles the result from a model trained on one of the two separately. The fact that the black model outperforms both the orange and green models on ASTRID suggests that this is not the case. We also compared the inference results on individual ASTRID maps from the three ML models and found that in many cases the black model can make more accurate inferences when the other two models are biased.

It is encouraging to see that the model trained on two suites is performing better than the ones trained on a single suite, and it shows some degree of extrapolation when applied to a new suite of galaxy formation models. Our speculation is that during the training process on the two distinct simulation suites the ML model is encouraged to ignore features specific to TNG or SIMBA that make the single-suite-trained models less robust when tested on a previously unseen suite. Instead, it focuses on capturing more general features that are shared between TNG and SIMBA to make the inference. Therefore, it seems promising to improve the robustness of ML models by including more distinct galaxy formation simulations in the training set.

However, we should note that the black model's performance on ASTRID (especially on σ_8) is relatively poor
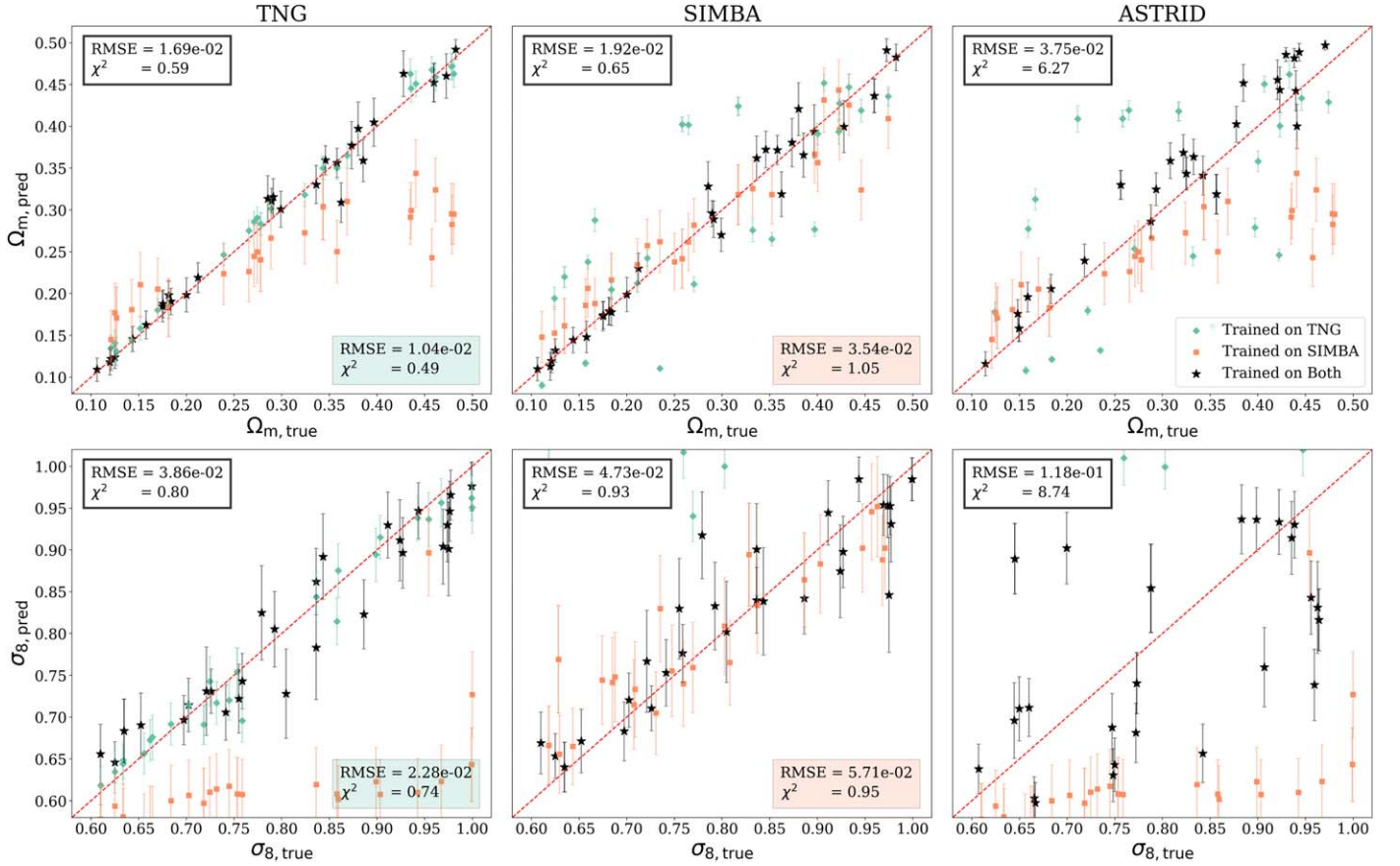
Figure 10. We train a neural network to infer Ω_m (top panels) and σ_8 (bottom panels) from gas temperature maps, tested on TNG (left column), SIMBA (middle column), and ASTRID (right column). The green, orange, and black colors represent the models trained on (the LH set of) TNG, SIMBA, and TNG+SIMBA suites, respectively. The data points and error bars give the posterior mean and variance. The scores of RMSE and χ^2 are shown in the legends.

compared to the results tested on TNG or SIMBA. Thus, we still caution that blindly training a model based on more simulation suites does not guarantee generalization to other galaxy formation models. Specially tailored machines that are trained with robustness in mind are a promising research direction, which we will explore in Y. Jo et al. (2023, in preparation).

6.3. Training on the TNG-SB28 Set

In this subsection, we present the first results of ML models trained on the new TNG-SB28 simulation set. The TNG-SB28 set sparsely samples a high-dimensional space of 28 cosmological and astrophysical parameters with 1024 simulations. It was a priori uncertain whether it is feasible to train an ML model for cosmological inference based on such a simulation set or whether many more simulations would be needed to perform such task.

In Figure 11, we show the test results for TNG-SB28, where we train a neural network to infer Ω_m and σ_8 based on the 2D total mass density field generated from TNG-SB28. The model architecture is the same as that introduced in Villaescusa-Navarro et al. (2021c). The model yields good scores for Ω_m and worse, yet still constraining, results for σ_8, despite the broad variation (and possible degeneracy) of all 28 parameters in TNG-SB28.

In our previous work (Villaescusa-Navarro et al. 2021b) training the neural network on the TNG-LH set and testing on the TNG-LH set, the mean relative error (mre) based on the 2D total

matter density maps is $\langle \delta \Omega_m / \Omega_m \rangle = 0.034$ and $\langle \delta \sigma_8 / \sigma_8 \rangle = 0.024$. Compared to the results based on the TNG-LH set, the constraining power on Ω_m does not degrade by making the parameter space larger in TNG-SB28 (with mre $\langle \delta \Omega_m / \Omega_m \rangle = 0.030$), while σ_8 is affected by it (the corresponding mre score degrades to $\langle \delta \sigma_8 / \sigma_8 \rangle = 0.060$), potentially due to the degeneracy introduced by additional parameters varied. However, there is still some constraining power on σ_8. We have also conducted a similar training and testing for the Ω_m and σ_8 inference model based on gas temperature maps, which produced slightly worse but still constraining results, with mre = 0.064 and 0.072 for Ω_m and σ_8, respectively (while the corresponding mre values based on the TNG-LH set are 0.053 and 0.037 for the gas temperature maps).

This section demonstrates the possibility of training ML models for cosmological inference based on the TNG-SB28 set. In future works, we will thoroughly investigate which categories of physical maps can infer which cosmological or astrophysical parameters in the TNG-SB28 set.

7. Summary and Conclusion

We have introduced the CAMELS-ASTRID simulation suite as the third large hydrodynamic simulation suite in the CAMELS project, carried out using the galaxy formation model of the ASTRID simulation. The core CAMELS-ASTRID suite contains the {CV, 1P, LH, EX} simulation sets with the same design as the TNG and SIMBA suites presented in Villaescusa-Navarro et al. (2021a), exploring the
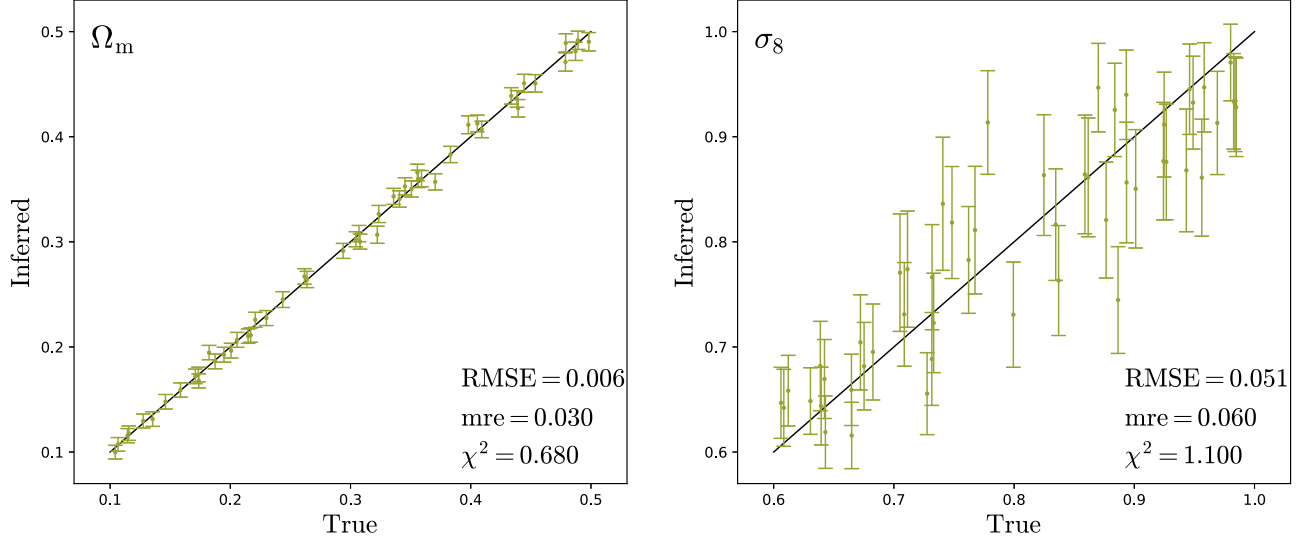
Total Matter Density

Figure 11. We show the results of a neural network trained and tested on the TNG-SB28 set to infer Ω_m and σ_8 from 2D total matter density maps. This demonstrates that it is still possible to train an ML model to infer cosmological parameters even with the simulation set of TNG-SB28 that broadly (and sparsely) samples the high-dimensional parameter space of a galaxy formation model.

cosmological and astrophysical parameter space of $\{\Omega_m, \sigma_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN2}\}$. The extension of CAMELS-ASTRID contains an additional LHOb set also varying the cosmological parameter Ω_b along with the six fiducial parameters.

We have also presented CAMELS extension simulation sets based on TNG and SIMBA that aim to explore a higher-dimensional parameter space in their respective galaxy formation models. In particular, the TNG-SB28 simulation set contains 1024 simulations that semi-uniformly sample a space of 5 cosmological parameters and 23 subgrid parameters. It is accompanied by the TNG-1P-28 set that varies one parameter at a time to study their individual impact on different astrophysical properties. The analogous set SIMBA-1P-28 is also carried out to explore the high-dimensional parameter space in SIMBA. Details of the 1P-28 sets in TNG and SIMBA are presented in the appendices.

We summarize the main features of the new simulation sets below.

1. The fiducial model of ASTRID exhibits the least impact on the large-scale gas properties and baryonic feedback on the matter power spectrum owing to a milder AGN jet mode feedback compared to TNG and SIMBA.

2. Varying feedback efficiency introduces nonlinear effects to the overall galaxy and BH populations and eventually the matter power spectrum. One example is the A_{AGN2} parameter in the ASTRID suite that controls the AGN thermal feedback efficiency. A larger A_{AGN2} suppresses the formation of massive BHs, curtails the energy budget of AGN feedback, brings less baryonic suppression on the total matter power spectrum, and enhances the global SFR and galaxy populations. Given the intricacy of the convoluted effects of feedback processes, we should view the astrophysical parameters A_{SN} and A_{AGN} as modulations of various astrophysical processes that lead to variations in different physical quantities, rather than directly infer the amount of feedback from each

astrophysical process as would be indicated by A_{SN} and A_{AGN} at face value.

3. We show that in cases where the cosmological parameters are varied identically between the ASTRID, TNG, and SIMBA suites, the responses to those parameters can show strong differences between simulation models. For example, the matter power spectrum responds differently to Ω_m and σ_8 on small scales between ASTRID, TNG, and SIMBA. In the extended 1P-28 sets, TNG and SIMBA show different responses of gas power spectrum to changes in the h and n_s cosmological parameters. This stresses the importance of including different galaxy formation models to improve and test the robustness of cosmological inference tasks.

4. The TNG-1P-28 and SIMBA-1P-28 sets allow us to develop a more complete picture of the sensitivity of the galaxy formation process to a large number of parameter variations in the TNG and SIMBA models and their effects on the cosmic matter distribution. We find that each of the 28 parameters, which are varied within reasonable ranges based on physical intuition, has a significant effect on the results of the models. Further, there is significant variability in the effect of different parameters on different quantities or "observables," where the overall impact of increasing one parameter is not necessarily linear or even monotonic. This demonstrates the enormity of the overall relevant galaxy formation model parameter space, which is sampled much more extensively here than in previous work, but it is still largely unexplored and not well understood.

5. Compared to the LH sets of TNG and SIMBA, the ASTRID-LH set drives larger variations in the gas power spectrum, the baryonic effect on the total matter power spectrum, the halo baryon fraction, the SFRD, and the galaxy population. In the ASTRID suite, the matter power spectrum is sensitive to the A_{AGN2} and A_{SN2} parameters that regulate BH growth, as well as the AGN jet mode feedback. The large variation in the cosmic SFR

and global galaxy properties is mainly driven by the sensitivity of ASTRID to the A_{SN2} parameter (which controls the SN wind speed). The broad variation of the ASTRID-LH set in some aspects promotes the robustness of ML model training. We find that ML models trained on galaxy catalogs from the ASTRID-LH set exhibit the best extrapolation performances when applied to other simulation sets.

6. The new TNG-SB28 set spans a much larger parameter space of the TNG model compared to the original TNG-LH set, and indeed it produces wider variations in the simulation results in all the individual quantities and scaling relations that we have examined, typically variations that are roughly as wide as those from the ASTRID-LH set. The degree to which this extended TNG set produces more varied results than the original TNG-LH set is, however, different between different quantities. We hypothesize that in some cases this is due to an averaging effect of many important parameters, while in some cases the new set produces significantly larger variations than the original owing to the introduction of new parameters that have a particularly significant effect on the quantity in question. The TNG-SB28 set probes a much more complete account of the flexibility in the TNG framework and can be used as a more realistic testing ground for ML models, since it does not hold fixed the values of many parameters that are fundamentally not known but are kept fixed in the original TNG-LH set.

We have presented a few ML applications allowed by the new simulation sets introduced in this work, as summarized below:

1. We evaluate the generalization performance of an ML inference model trained on 2D mass maps from the TNG suite, which generalizes to and can successfully predict Ω_m and σ_8 in the SIMBA suite. However, when tested on the ASTRID suite, the model fails to infer σ_8, despite the fact that ASTRID 2D mass maps are more similar to TNG in many statistical properties compared to SIMBA. This underscores the importance of including a diverse set of independent galaxy formation models to assess the robustness of trained ML models.

2. We develop a cosmological ML inference model by training on gas temperature maps, which have been previously shown to give rise to models that are not robust, from a combination of the TNG+SIMBA suites. While when trained on TNG+SIMBA the resulting model generalizes well between TNG and SIMBA, when we test its ability to predict Ω_m and σ_8 on the ASTRID data, our findings suggest that training a model using multiple simulation suites does not necessarily result in generalization to other galaxy formation models. Yet we observed that the two-suite trained model (TNG+SIMBA) performed better when tested on the ASTRID data than the models trained only on one suite (TNG or SIMBA separately), indicating the potential for improving the robustness of ML models by incorporating additional distinct galaxy formation models in the training sets.

3. We demonstrate the feasibility of training neural networks to infer Ω_m and σ_8 using the TNG-SB28 simulation set, which sparsely samples the high-dimensional parameter space of a galaxy formation model. Our results show that it

is possible to train accurate ML models using this set, which can successfully infer cosmological parameters.

Appendix A
Description of the 28 Varied Parameters

Here we provide brief yet explicit and complete accounts of the 28 parameters that are varied in each of the new TNG and SIMBA sets discussed in this work. In addition to a concise description of the physical meaning of each parameter, we provide references to previous work that presented the parameter in the broader context of the model, as well as the range within which we vary the parameter value in this work. We also note whether the sampling of the parameter values within said range in this work is uniform in linear or logarithmic space.

A.1. TNG 28-parameter Variations

1. Omega0 is the standard cosmological parameter Ω_m, the $z = 0$ cosmic matter density in units of the critical density. Here it is sampled linearly between 0.1 and 0.5, around the fiducial value of 0.3.

2. Sigma8 is the standard cosmological parameter σ_8, the rms of the $z = 0$ linear overdensity in spheres of radius $8\,h^{-1}$ Mpc. Here it is sampled linearly between 0.6 and 1, around the fiducial value of 0.8.

3. WindEnergyIn1e51erg (A_{SN1}) is a normalization factor for the energy in galactic winds per unit star formation and is denoted as $\bar{e}_w$ in Equation (3) in Pillepich et al. (2018). In the existing literature using CAMELS, including in this work, it is commonly referred to as A_{SN1}, albeit with a ratio of 3.6 between the two. Here it is sampled logarithmically between 0.9 and 14.4, around the fiducial value of 3.6.

4. VariableWindVelFactor (A_{SN2}) is a normalization factor for the galactic wind speed and is denoted as κ_w in Equation (1) in Pillepich et al. (2018).[16] In the existing literature using CAMELS, including in this work, it is commonly referred to as A_{SN2}, albeit with a ratio of 7.4

---

[16] Note that its description in Equation (4) in Villaescusa-Navarro et al. (2021a) is incorrect in that A_{SN2} should have been located in front of κ_w instead of the whole rhs.

between the two. Here it is sampled logarithmically between 3.7 and 14.8, around the fiducial value of 7.4.

5. RadioFeedbackFactor (A_{AGN1}) is a normalization factor for the energy in AGN feedback, per unit accretion rate, in the low-accretion state and is implemented as a pre-factor in front of the right-hand side (rhs) of Equation (8) in Weinberger et al. (2017). In the existing literature using CAMELS, including in this work, it is commonly referred to as A_{AGN1}. Here it is sampled logarithmically between 0.25 and 4.0, around the fiducial value of 1.0.

6. RadioFeedbackReorientationFactor (A_{AGN2}) is a normalization factor for the frequency of AGN feedback energy release events in the low-accretion state and is denoted as f_{re} in Equation (13) in Weinberger et al. (2017). In the existing literature using CAMELS, including in this work, it is commonly referred to as A_{AGN2}, albeit with a ratio of 20 between the two. Here it is sampled logarithmically between 10 and 40, around the fiducial value of 20.

7. OmegaBaryon is the standard cosmological parameter Ω_b, the $z = 0$ cosmic baryon density in units of the critical density. Here it is sampled linearly between 0.029 and 0.069, around the fiducial value of 0.049.

8. HubbleParam is the standard Hubble constant, in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Here it is sampled linearly between 0.4711 and 0.8711, around the fiducial value of 0.6711.

9. n_s is the standard cosmological parameter n_s, the spectral index of the initial fluctuations.[17] Here it is sampled linearly between 0.7624 and 1.1624, around the fiducial value of 0.9624.

10. MaxSfrTimescale is the timescale for star formation at the density threshold of star formation and is denoted as $t_0^\star$ in Equation (21) in Springel & Hernquist (2003a). Here it is sampled logarithmically between 1.135 and 4.54 Gyr, around the fiducial value of 2.27 Gyr.

11. FactorForSofterEQS is an interpolation factor between the effective equation of state for star-forming gas obtained from the Springel & Hernquist (2003a) subgrid model and an isothermal equation of state with 10^4 K and is denoted as q_{EOS} in Springel et al. (2005). Here it is sampled logarithmically between 0.1 and 0.9, around the fiducial value of 0.3.

12. IMFslope is the power-law index of the stellar IMF above $1\,M_\odot$ (whereas below that mass, it is kept fixed following Chabrier 2003). Here it is sampled linearly between −2.8 and −1.8, around the fiducial value of −2.3.

13. SNII_MinMass_Msun is the lower threshold for the mass of a star that produces an SN explosion and is denoted $M_{SNII,min}$ in Vogelsberger et al. (2013). It affects both the available energy for galactic wind feedback (and in that sense is degenerate with WindEnergyIn1e51erg, or A_{SN1}) and the effective (namely, stellar-population-averaged) stellar mass return and enrichment. Here it is sampled linearly between 4 and $12\,M_\odot$, around the fiducial value of $8\,M_\odot$.

14. ThermalWindFraction is the fraction of the galactic wind feedback energy that is injected thermally (the other component being kinetically) and is denoted as τ_w in

Pillepich et al. (2018). Here it is sampled logarithmically between 0.025 and 0.4, around the fiducial value of 0.1.

15. VariableWindSpecMomentum is a normalization factor for the specific momentum in galactic winds per unit star formation and is denoted as mom_w in Vogelsberger et al. (2013). Note that it affects the specific momentum of the winds only via their mass-loading factor and not via their speed. Here it is sampled linearly between 0 and $4000\,\mathrm{km\,s^{-1}}$. It is worth noting that the variations in this parameter differ from those of all the others in that the fiducial value of 0 lies at the edge of the variation range rather than in its middle, for physical reasons (negative values are unphysical).

16. WindFreeTravelDensFac sets the gas density around (collisionless) galactic wind particles at which they recouple back into the hydrodynamics. Here it is sampled logarithmically between 0.005 and 0.5, around the fiducial value of 0.05, all in units of the density threshold for star formation.

17. MinWindVel is the minimum value imposed for the galactic wind speed and is denoted as $v_{w,min}$ in Pillepich et al. (2018). Here it is sampled linearly between 150 and $550\,\mathrm{km\,s^{-1}}$, around the fiducial value of $350\,\mathrm{km\,s^{-1}}$.

18. WindEnergyReductionFactor is a normalization factor for the energy of galactic winds at high metallicity compared to low metallicity and is denoted as $f_{w,Z}$ in Pillepich et al. (2018). Here it is sampled logarithmically between 0.0625 and 1.0, around the fiducial value of 0.25.

19. WindEnergyReductionMetallicity sets the metallicity at which the transition from high- to low-energy galactic winds occurs and is denoted as $Z_{w,ref}$ in Pillepich et al. (2018). Here it is sampled logarithmically between 0.0005 and 0.008, around the fiducial value of 0.002, in terms of fractional metal mass.

20. WindEnergyReductionExponent controls the abruptness in metallicity of the transition between high- and low-energy galactic winds and is denoted as $\gamma_{w,Z}$ in Pillepich et al. (2018). Here it is sampled linearly between 1 and 3, around the fiducial value of 2.

21. WindDumpFactor is the fraction of the metals in a star-forming cell getting ejected into a galactic wind that get deposited in neighboring star-forming cells prior to the ejection and is denoted as $1 - \gamma_w$ in Pillepich et al. (2018). Here it is sampled linearly between 0.2 and 1.0, around the fiducial value of 0.6.

22. SeedBlackHoleMass is the mass of seed SMBHs, as described in Vogelsberger et al. (2013). Here it is sampled logarithmically between $2.5 \times 10^5\,M_\odot$ and $2.5 \times 10^6\,M_\odot$, around the fiducial value of $8 \times 10^5\,M_\odot$.

23. BlackHoleAccretionFactor is a normalization factor for the Bondi rate for the accretion onto SMBHs and is implemented as a pre-factor in front of the rhs of Equation (2) in Weinberger et al. (2017). Here it is sampled logarithmically between 0.25 and 4, around the fiducial value of 1.

24. BlackHoleEddingtonFactor is a normalization factor for the limiting Eddington rate for the accretion onto SMBHs and is implemented as a pre-factor in front of the rhs of Equation (3) in Weinberger et al. (2017). Here it is sampled logarithmically between 0.1 and 10, around the fiducial value of 1.

---

[17] Note that since the power of the fluctuations at a scale of $8\,h^{-1}$ Mpc serves as a pivot point that is set by Sigma8, most scales available in our $25\,h^{-1}$ Mpc boxes are enhanced when n_s is increased, with smaller scales enhanced more than larger ones.

25. `BlackHoleFeedbackFactor` is a normalization factor for the energy in AGN feedback, per unit accretion rate, in the high-accretion state and is denoted as $\epsilon_{f,\mathrm{high}}$ in Equation (7) in Weinberger et al. (2017). Here it is sampled logarithmically between 0.025 and 0.4, around the fiducial value of 0.1. We note that this parameter is similar to the A_{AGN2} parameter in ASTRID suite.

26. `BlackHoleRadiativeEfficiency` is the radiative efficiency of AGN feedback, namely the fraction of the accretion rest mass that is released in the accretion process. In the high-accretion state, it is denoted as ϵ_r in Equation (7) in Weinberger et al. (2017). In the low-accretion state, it plays a dual role in setting $\epsilon_{f,\mathrm{kin}}$ in Equation (9) in Weinberger et al. (2017) by (i) serving as an upper limit on its value (replacing the constant 0.2 that appears in that equation) and (ii) modulating the density at which that upper limit is reached, by scaling the density in the numerator of the first term in that equation, where a factor (`BlackHoleRadiativeEfficiency`/0.2) is missing for the general case of `BlackHoleRadiativeEfficiency` $\neq 0.2$. Here it is sampled logarithmically between 0.05 and 0.8, around the fiducial value of 0.2.

27. `QuasarThreshold` is the Eddington ratio (at the "pivot mass" of $10^8 M_\odot$) that serves as the threshold between the low-accretion and high-accretion states of AGN feedback and is denoted as χ_0 in Equation (5) in Weinberger et al. (2017). Here it is sampled logarithmically between .000063 and .063, around the fiducial value of 0.002.

28. `QuasarThresholdPower` is the power-law index of the scaling of the low- to high-accretion state threshold with BH mass and is denoted as β in Equation (5) in Weinberger et al. (2017). Here it is sampled linearly between 0 and 4, around the fiducial value of 2.

A.2. SIMBA 28-parameter Variations

1. `Omega0` is the $z = 0$ cosmic matter density in units of the critical density. Here it is sampled linearly between 0.1 and 0.5 around the fiducial value $\Omega_m = 0.3$.

2. `Sigma8` is the rms of the $z = 0$ linear overdensity in spheres of radius $8\,h^{-1}$ Mpc. Here it is sampled linearly between 0.6 and 1 around the fiducial value of $\sigma_8 = 0.8$.

3. A_{SN1} is a normalization factor for the mass loading of galactic winds relative to the scaling from FIRE simulations (Anglés-Alcázar et al. 2017b) implemented in SIMBA. This is one of the feedback parameters originally varied in CAMELS (see Equation (7) in Villaescusa-Navarro et al. 2021a), and it is sampled logarithmically between 0.25 and 4.0 around the fiducial value $A_{\mathrm{SN1}} = 1.0$.

4. A_{SN2} is a normalization factor for the speed of galactic winds relative to the scaling from FIRE simulations (Muratov et al. 2015) implemented in SIMBA. This is one of the feedback parameters originally varied in CAMELS (see Equation (8) in Villaescusa-Navarro et al. 2021a), and it is sampled logarithmically between 0.5 and 2.0 around the fiducial value $A_{\mathrm{SN2}} = 1.0$.

5. A_{AGN1} is a normalization factor for the momentum flux of kinetic AGN-driven outflows (Anglés-Alcázar et al. 2017a) in the radiative and jet modes. This is one of the feedback parameters originally varied in CAMELS

(see Equation (9) in Villaescusa-Navarro et al. 2021a), and it is sampled logarithmically between 0.25 and 4.0 around the fiducial value $A_{\mathrm{AGN1}} = 1.0$.

6. A_{AGN2} is a normalization factor for the speed of AGN-driven outflows in jet mode. This is one of the feedback parameters originally varied in CAMELS (see Equation (10) in Villaescusa-Navarro et al. 2021a), and it is sampled logarithmically between 0.5 and 2.0 around the fiducial value $A_{\mathrm{AGN2}} = 1.0$.

7. `OmegaBaryon` is the $z = 0$ cosmic baryon density in units of the critical density. Here it is sampled linearly between 0.029 and 0.069 around the fiducial value of $\Omega_b = 0.049$.

8. `HubbleParam` is the standard Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Here it is sampled linearly between 0.4711 and 0.8711 around the fiducial value $h = 0.6711$.

9. `n_s` is the spectral index of the initial fluctuations. Here it is sampled linearly between 0.7624 and 1.1624 around the fiducial value of $n_s = 0.9624$.

10. `SfrCritDens` is the gas number density threshold for star formation. Here it is sampled logarithmically between 0.02 and $2\,\mathrm{cm^{-3}}$ around the fiducial value of $0.2\,\mathrm{cm^{-3}}$.

11. `SfrEfficiency` is the star formation efficiency per freefall time, denoted as $\epsilon_\star$ in Davé et al. (2019). Here it is sampled logarithmically between 0.01 and 0.04 around the fiducial value $\epsilon_\star = 0.02$.

12. `ISMJeansFac` determines the level of artificial pressurization in the ISM such that the Jeans mass is resolved with a number of gas resolution elements given by `ISMJeansFac` $\times N_{\mathrm{ngb}}$, where $N_{\mathrm{ngb}} = 64$ is the number of neighbors in the smoothing kernel (Davé et al. 2016). Here `ISMJeansFac` is sampled logarithmically between 0.25 and 4.0 around the fiducial value of 1.0.

13. `WindTravTime` is the maximum amount of time that galactic winds can propagate before recoupling hydrodynamically to the surrounding gas, in units of the Hubble time at launch (Davé et al. 2016). Here it is sampled logarithmically between 0.2 and 0.002 around the fiducial value of 0.02.

14. `WindDensFac` defines a gas number density threshold below which decoupled galactic winds are forced to recouple hydrodynamically to the surrounding gas (Davé et al. 2016). Here it is sampled logarithmically between 0.1 and $0.001\,\mathrm{cm^{-3}}$ around the fiducial value of $0.01\,\mathrm{cm^{-3}}$.

15. `WindColdTemp` defines the temperature assumed for the cold phase of galactic winds (Davé et al. 2016). Here it is sampled logarithmically between 100 and 10^4 K around the fiducial value of 1000 K.

16. `WindHotFrac` is the fraction of wind particles ejected in the hot phase, where the wind temperature is determined by the fraction of the SN energy not used for kinetic ejection (Davé et al. 2016). Here `WindHotFrac` is sampled linearly between 0.0 and 0.6 around the fiducial value of 0.3.

17. `WindVelSlope` sets the power-law dependence of the galactic wind velocity on the circular velocity of the galaxy such that $\log_{10}(v_{\mathrm{wind}}) \propto (1+\texttt{WindVelSlope}) \times \log_{10}(v_{\mathrm{circ}})$, as defined in Equation (3) of Davé et al. (2019). Here `WindVelSlope` is sampled

linearly between -0.38 and 0.62 around the fiducial value of 0.12.

18. `AGBWindHeatVel` is the velocity of winds from AGB stars that are assumed to thermalize with the ambient ISM following the model of Conroy et al. (2015). Here `AGBWindHeatVel` is sampled logarithmically between 25 and $400\,\mathrm{km\,s^{-1}}$ around the fiducial value of $100\,\mathrm{km\,s^{-1}}$.

19. `BHSeedMass` is the mass of the SMBH seeds, which is varied logarithmically between 10^3 and $10^5\ h^{-1}M_\odot$ around the fiducial value of $10^4\ h^{-1}M_\odot$ (Davé et al. 2019).

20. `BHSeedRatio` defines the minimum galaxy stellar mass required for BH seeding as $\mathrm{BHSeedRatio} \times M_{\mathrm{seed}}$ (Anglés-Alcázar et al. 2017a). Here `BHSeedRatio` is sampled logarithmically between 3×10^4 and 3×10^6 around the fiducial value of 3×10^5.

21. `BHAccrFac` is a normalization factor for the BH growth rate applied to both accretion of hot gas following Bondi (1952) and accretion of cold gas driven by gravitational torques (Hopkins & Quataert 2011; Anglés-Alcázar et al. 2017a). Here `BHAccrFac` is sampled logarithmically between 0.25 and 4.0 around the fiducial value of 1.0.

22. `BHAccrMaxR` is the maximum size of the BH kernel used when searching for neighboring gas elements. Here it is sampled logarithmically between 2 and $8\ h^{-1}$ kpc around the fiducial value of $4\ h^{-1}$ kpc.

23. `BHAccrTempThr` is the temperature threshold above which gas accretes onto the BH at the Bondi rate and below which gas accretion proceeds at the gravitational torque accretion rate (Davé et al. 2019). Here it is sampled logarithmically between 10^4 and 10^6 K around the fiducial value of 10^5 K.

24. `BHEddingtonFac` is a normalization factor for the limiting Eddington rate applied to the accretion of cold gas onto BHs through gravitational torques (hot-mode accretion is always limited to the Eddington rate). Here `BHEddingtonFac` is sampled logarithmically between 0.75 and 12.0 around the fiducial value of 3.0.

25. `BHNgbFac` sets the desired effective number of gas resolution elements within the BH kernel as $\mathrm{BHNgbFac} \times N_{\mathrm{ngb}}$, where $N_{\mathrm{ngb}} = 64$ is the number of neighbors in the gas smoothing kernel. Here it is sampled linearly from 2 to 6 around the fiducial value of 4.

26. `BHRadiativeEff` is the radiative efficiency of the BH accretion disk, denoted as η in Davé et al. (2019), which determines the fraction of the rest-mass energy of the accreting material lost to radiation, the amount of energy injected as X-ray feedback, and the momentum flux of outflows in radiative and jet modes (Davé et al. 2019). Here `BHRadiativeEff` is sampled logarithmically from 0.025 to 0.4 around the fiducial value of 0.1.

27. `BHJetTvirVel` sets the outflow velocity above which gas ejected in jet mode is heated to the virial temperature of the host halo (Davé et al. 2019). Here `BHJetTvirVel` is sampled logarithmically from 500 to $8000\,\mathrm{km\,s^{-1}}$ around the fiducial value of $2000\,\mathrm{km\,s^{-1}}$.

28. `BHJetMassThr` is the BH mass threshold above which BHs are allowed to transition into the jet feedback mode. It is sampled logarithmically from $4.5 \times 10^6\,M_\odot$ to $4.5 \times 10^8\,M_\odot$ around the fiducial value of $4.5 \times 10^7\,M_\odot$ (Davé et al. 2019).

Appendix B
TNG Extended 1P Set

Here we present a few examples for how individual model parameters affect the results of simulations in the TNG framework, by way of showing three distinct quantities based on the TNG-1P-28 set—the cosmic SFRD, the M_{BH}–M_* relation, and the gas power spectrum $P_{\mathrm{gas}}(k)$ normalized by that of the fiducial model. Figures 12, 13, and 14 each show one of these quantities, respectively, for 57 simulations from the TNG-1P-28 set, which are the two extremes for each of the 28 parameters (red for the highest and blue for the lowest value), as well as the fiducial model (black, repeating in all panels). The panels are organized such that the first six correspond to the original six parameters from the TNG-LH set, followed by the three new cosmological parameters, followed by the star formation and then the galactic wind parameters, and finally in the bottom row the BH growth and AGN feedback parameters.

This presentation of the dependencies of these three quantities on the model parameters in Figures 12, 13, and 14 serves to support the interpretation of the diversity of results in the TNG-SB28 set that is discussed in Section 5, to contribute to the justification for the choice of parameter variation ranges, and to explore the degree to which the various parameters are degenerate with one another in terms of their effects on the simulation results, as discussed below.

Figure 12, for the cosmic SFRD versus redshift, demonstrates that some of the original six parameters, namely Ω_m, σ_8, and A_{SN1}, are some of the most influential ones. However, two of the new cosmological parameters, h and n_s, have effects of similar overall scale. It is notable, though, that the response to each of these five most influential parameters has a unique shape, such that they are not degenerate with one another, even for just this single quantity. A few additional galactic wind parameters also have fairly significant effects; it is worth noting that one of those, `WindFreeTravelDensFac`, can be considered rather "numerical" in nature, as it controls the ambient density at which collisionless "wind particles" recouple into the hydrodynamics. This highlights a situation where even "nuisance" parameters without a very clear analog in physical reality, which are often "buried" deep inside the model, can be consequential for modern galaxy formation simulations. Finally, we note that BH parameters have a mild effect on this quantity, as expected, given that they tend to affect the most massive galaxies but that star formation is dominated by low- and intermediate-mass galaxies.

Figure 13, for the $z = 0$ stellar-to-halo mass ratio, shows a much broader set of parameters that matter than is the case for the SFRD in Figure 12. Essentially all the parameters except the original AGN parameters A_{AGN1} and A_{AGN2} make a visible effect in the figure. On the other hand, of the galactic wind parameters, the original ones, A_{SN1} and A_{SN2}, are the most significant, joined by `WindFreeTravelDensFac`. These types of parameters, as well as the cosmological ones, tend to affect a wide range of halo masses, while the BH and AGN parameters meaningfully affect only halos more massive than $\sim 10^{12}\,M_\odot$, as expected. Newly varied star formation parameters, the star formation timescale and the IMF slope, also have a significant impact. It is interesting to explore the degeneracy, and breaking thereof, across the two quantities shown in Figures 12 and 13. For example, the effects of n_s and `MaxSfrTimescale` (second and third panels in the second row) appear to be almost degenerate for the M_*–M_h relation,
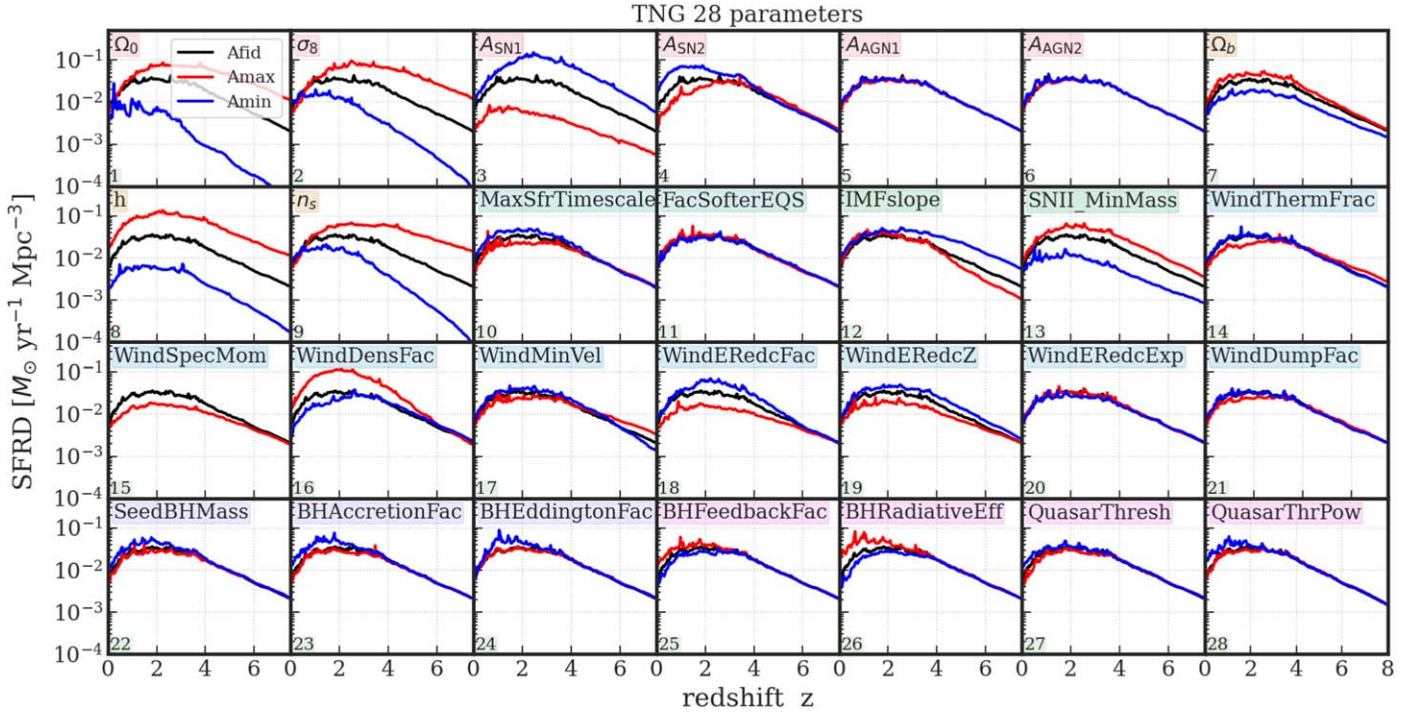
Figure 12. Global SFRD of the TNG extended 1P set. The 28 panels reproduce the fiducial model as the black line. The red and blue lines in each panel represent the simulations run with the highest and lowest parameter value (except for the `WindSpecMom` parameter in TNG; see text for more details).
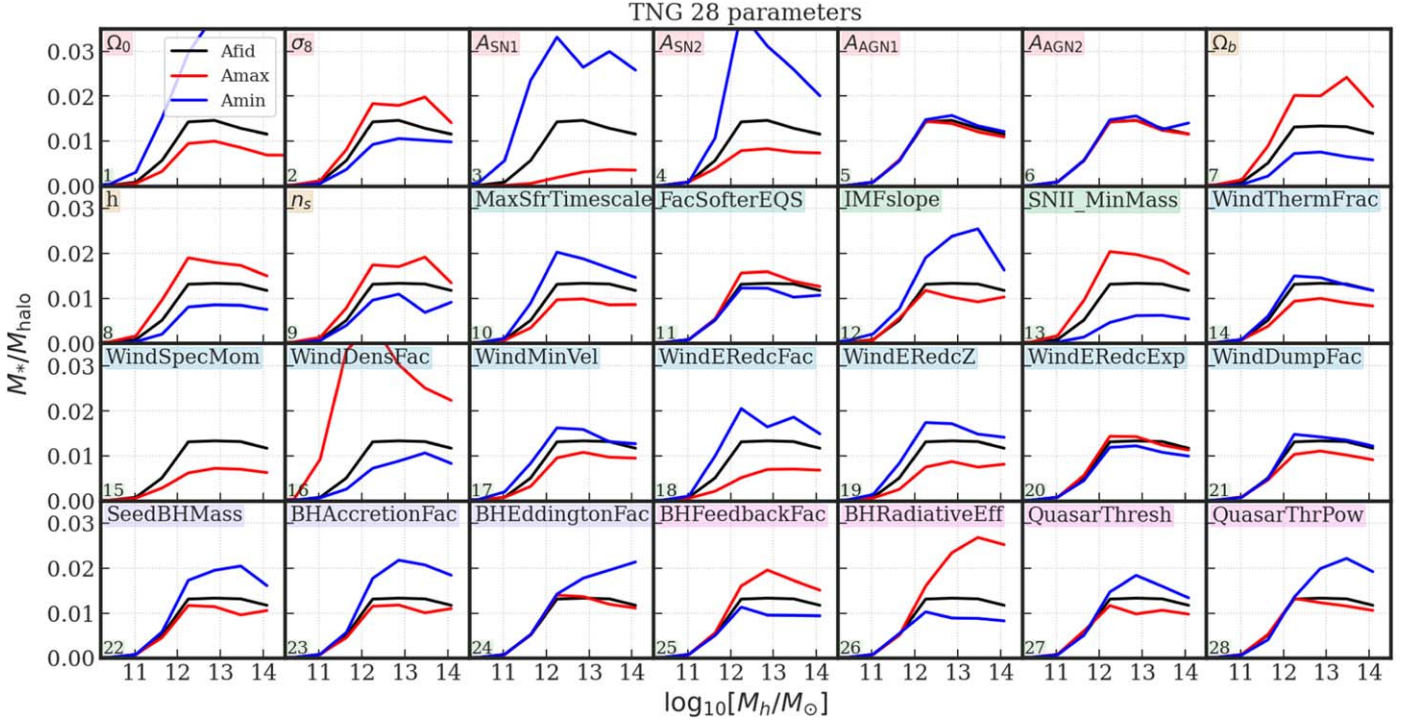

Figure 13. Same as Figure 12, but this plot gives the $z = 0$ stellar mass fraction of the TNG extended 1P set. $M_{\rm halo}$ here is the FOF halo mass, and M_* is the stellar mass within the FOF halo. The y-axis gives the averaged $(M_*/M_{\rm halo})$ for each halo mass bin for each simulation.

but they have radically different consequences for the SFRDs. Conversely, $A_{\rm SN2}$ and `BlackHoleRadiativeEffi-ciency`, which have a similar effect on the SFRDs, have rather distinct impacts on the M_*–M_h relation.

Finally, Figure 14, for the gas power spectrum $P_{\rm gas}(k)$ (relative to the fiducial model), demonstrates its own distinctive trends. For example, the newly added cosmological parameters

h and n_s that prove so significant for the SFRDs in Figure 12 make barely a dent on $P_{\rm gas}(k)$. In contrast, `QuasarThre-sholdPower`, which does not stand out as the most important AGN parameter in Figures 12 and 13, has a very strong effect on $P_{\rm gas}(k)$, in particular when its value is small, likely by moving many massive BHs from the kinetic to the thermal feedback mode. Overall, $P_{\rm gas}(k)$ shows a large variation in
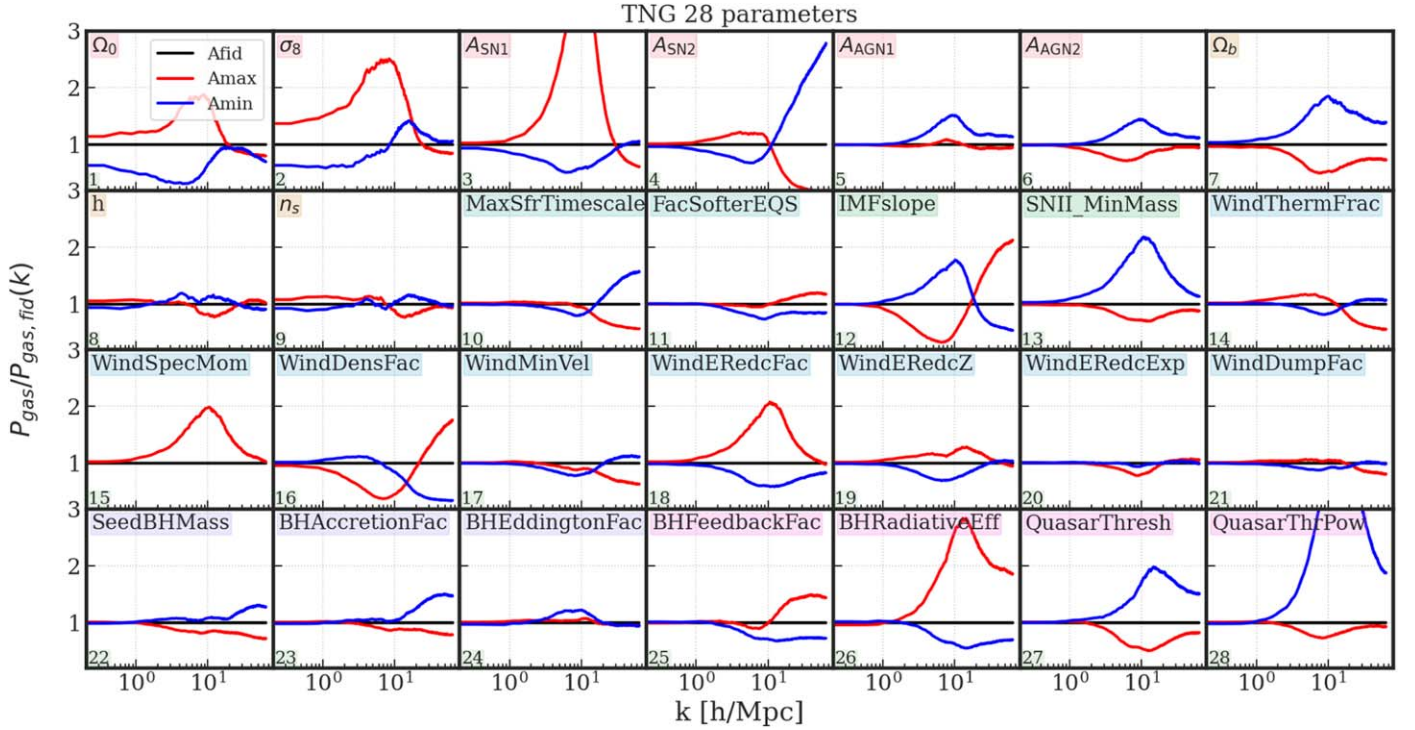
Figure 14. Same as Figure 12, but this plot gives the ratio between the gas power spectrum of the TNG extended 1P simulations and that of the fiducial model.

terms of the strength of its dependence on different parameters, while both the important ones and the nonimportant ones sample all parameter types: cosmological, star formation and ISM, galactic winds, and BH parameters.

Appendix C
SIMBA Extended 1P Set

As in Appendix B for TNG, here we illustrate how individual parameters in the SIMBA model affect global properties of galaxies and the large-scale distribution of baryons in simulations. Figure 15 shows the evolution of the cosmic SFRD, Figure 16 shows the stellar mass–halo mass relation at $z = 0$, and Figure 17 shows the gas power spectrum $P_{gas}(k)$ at $z = 0$ normalized by that of the fiducial model. Each figure consists of 28 panels corresponding to all parameter variations in the SIMBA-1P-28 set, where the fiducial model is indicated by the black line and the red and blue lines correspond to the results for the highest and lowest parameter values, respectively. The first six panels correspond to the original SIMBA parameter variations in CAMELS (Villaescusa-Navarro et al. 2021a), including Ω_m, σ_8, and four parameters controlling the efficiency of kinetic outflows driven by stellar feedback (A_{SN1} and A_{SN2}) and AGN feedback (A_{AGN1} and A_{AGN2}). Then, we show variations of three additional cosmological parameters (Ω_b, h, and n_s), followed by astrophysical parameters controlling star formation, galactic winds, BH accretion, and AGN feedback.

Figure 15 shows that the cosmic SFRD increases systematically with higher Ω_b, h, and n_s, as seen for TNG in Figure 12. Interestingly, the SFRD increases not only with the star formation efficiency, as expected, but also with a higher threshold density for star formation, suggesting that stellar feedback is less efficient at regulating star formation when the star-forming gas reaches higher densities. Besides the normalization of the mass loading (A_{SN1}) and velocity (A_{SN2}) of

galactic winds, the next most important wind parameter appears to be the power-law dependence of wind speed with the circular velocity of the galaxy `WindVelSlope`. Numerical parameters related to the hydrodynamic treatment of winds also have a noticeable effect, with longer hydrodynamic decoupling times decreasing the SFRD and higher recoupling densities having the opposite effect. Interestingly, increasing the BH seed mass at fixed BH-to-host mass ratio has similarly strong effects on the SFRD to increasing the BH-to-host mass ratio by the same factor at fixed BH seed mass. This suggests that the seed mass itself is not important, as expected given the weak dependence of gravitational torque accretion on BH mass (Anglés-Alcázar et al. 2013, 2017a), but the stellar mass at which galaxies are seeded is crucial, with significantly higher SFRD when BH seeding occurs only in higher-mass galaxies. Relatedly, increasing the BH accretion normalization systematically decreases the SFRD. Other important AGN feedback parameters besides the momentum flux (A_{AGN1}) and the jet speed (A_{AGN2}) include the BH mass threshold for jet feedback, with higher SFRD when only the most massive BHs are allowed to transition to the jet mode.

Figure 16 shows that most parameter variations have a noticeable effect in the stellar mass–halo mass relation at $z = 0$, as seen for TNG. Increasing Ω_m at fixed Ω_b decreases the stellar-to-halo mass ratio, as expected given the decline in the cosmic baryon fraction (Delgado et al. 2023), while increasing Ω_b at fixed Ω_m has the opposite effect. Increasing the momentum flux of AGN-driven outflows (A_{AGN1}) results in a reduction of M_*/M_{halo} across the halo mass range, while increasing the jet speed (A_{AGN2}) only affects the higher-mass galaxies (Davé et al. 2019). Interestingly, increasing the speed of galactic winds (A_{SN2}) drives a clear decrease in M_*/M_{halo} at all masses, while increasing the mass loading of winds (A_{SN1}) increases the stellar-to-halo mass ratio in massive galaxies. This counterintuitive effect could result from an increase in the
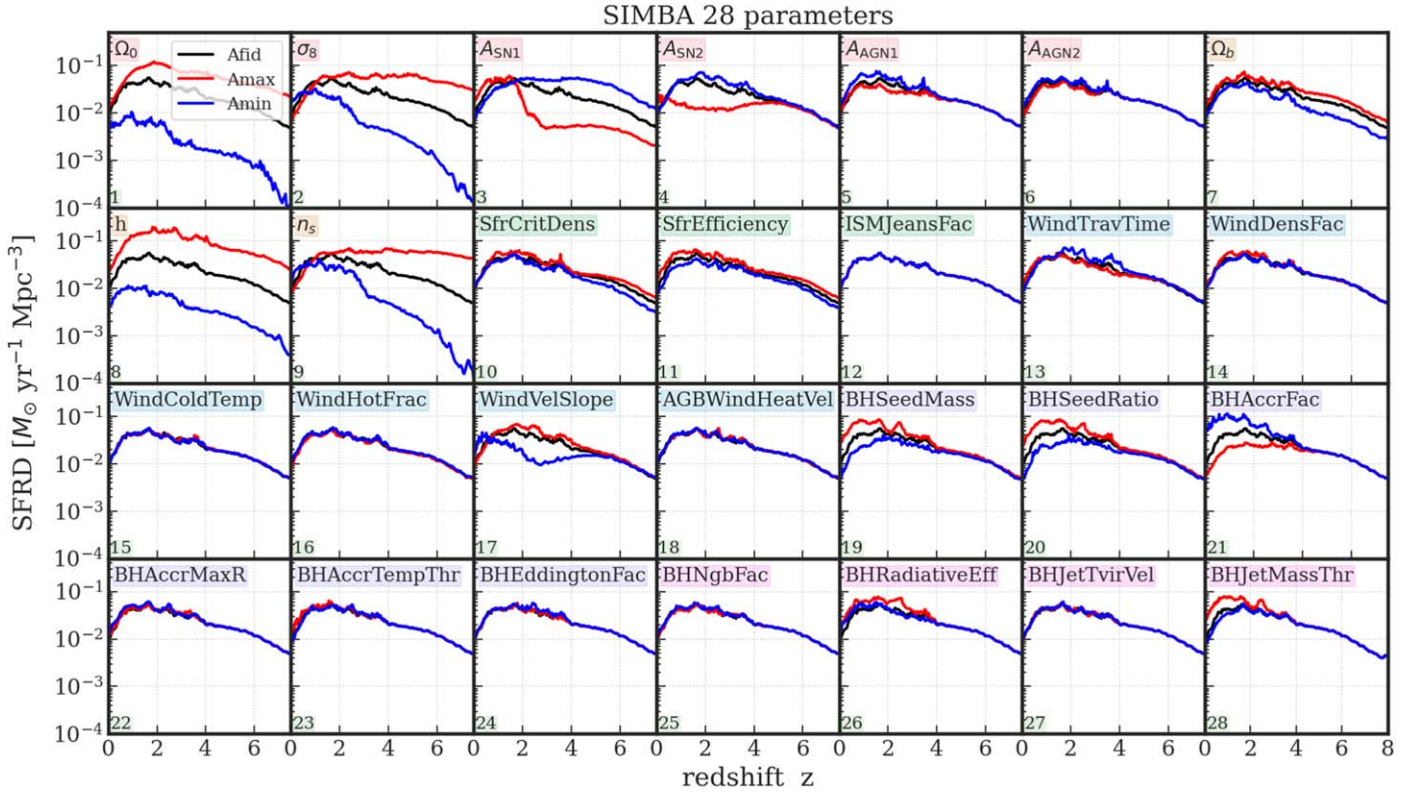
Figure 15. Global SFRD of the SIMBA extended 1P set. The 28 panels share the same black line that corresponds to the result from the fiducial model of SIMBA.
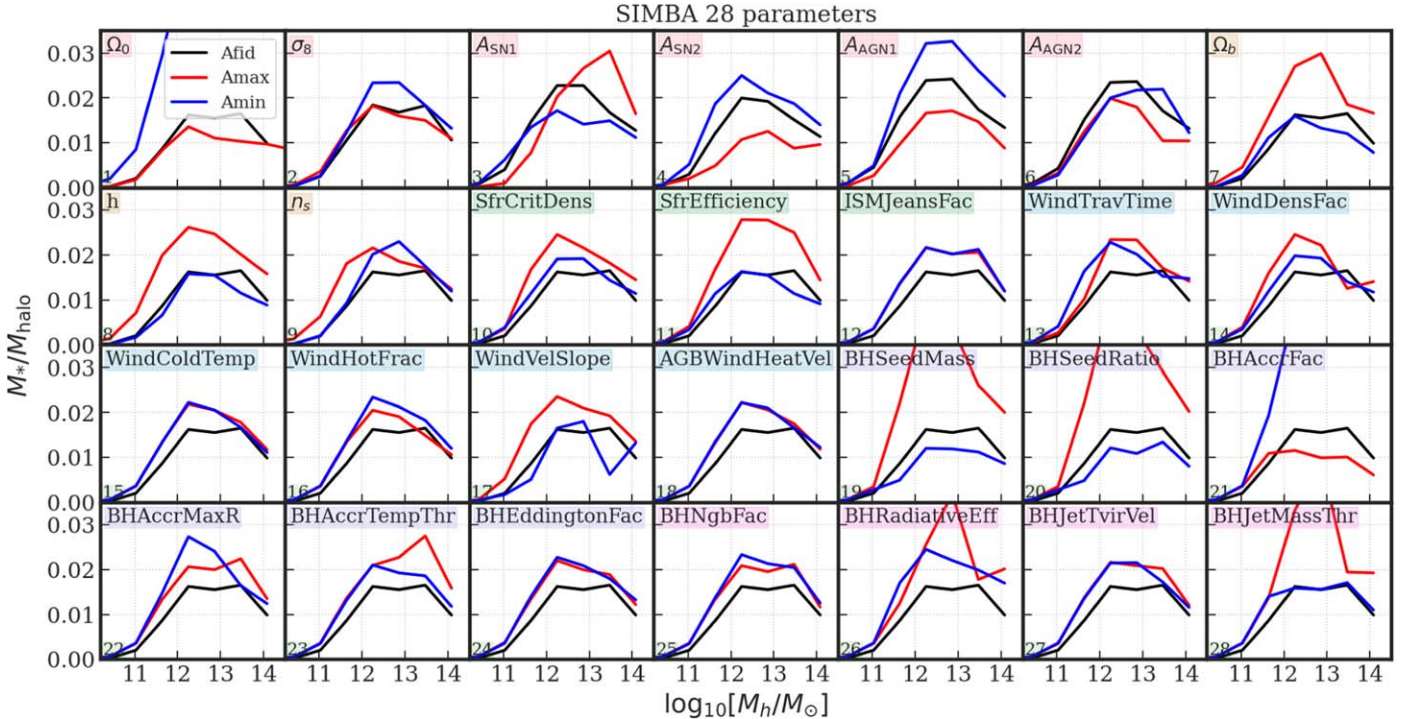

Figure 16. Same as Figure 15, but this plot gives the $z = 0$ stellar mass fraction of the SIMBA extended 1P set. M_{halo} here is the FOF halo mass, and M_* is the stellar mass within the FOF halo. The y-axis gives the averaged (M_*/M_{halo}) for each halo mass bin for each simulation.

intergalactic transfer of gas from lower-mass galaxies, providing an additional source of gas to higher-mass galaxies (Anglés-Alcázar et al. 2017b), or could reflect the complex nonlinear interplay between stellar and AGN feedback as noted in previous works (Booth & Schaye 2013; van Daalen et al.

2020; Nicola et al. 2022; Delgado et al. 2023; M. Gebhardt et al., in preparation). As expected from Figure 15, increasing the galaxy stellar mass above which BHs are seeded (i.e., higher BHSeedMass or BHSeedRatio) results in significantly higher M_*/M_{halo} across the halo mass range, while
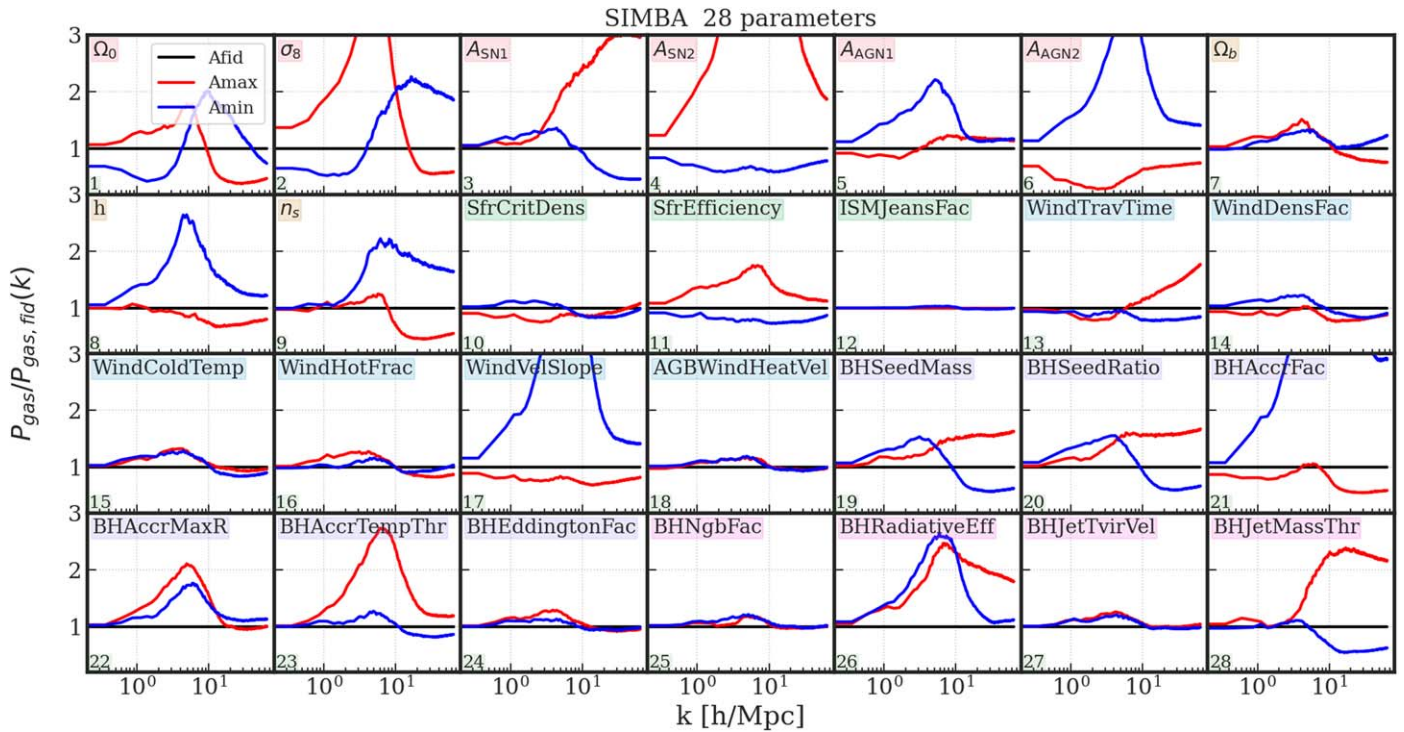
Figure 17. Same as Figure 15, but this plot gives the ratio between the gas power spectrum of the SIMBA extended 1P simulations and that of the fiducial model.

increasing the BH accretion normalization decreases M_*/M_{halo}. Another expected trend is the strong increase in stellar mass at fixed halo mass when only the most massive BHs are allowed to switch into the jet feedback mode (Davé et al. 2019). Nonetheless, several parameter variations indicate complex nonlinear behavior, with nonmonotonic trends when systematically increasing parameter values.

Figure 17 illustrates the very complex response of the gas power spectrum in SIMBA to changes in model parameters, where the relative impact of either increasing or decreasing parameter values often depends on and can even revert trend with scale. In addition to the original six parameters (1–6) varied in CAMELS, other key parameters affecting the clustering of gas include WindVelSlope, BHSeedRatio, BHAccFac, and BHJetMassThr, which also have important effects on the cosmic SFRD and the M_*–M_{halo} relation. Interestingly, increasing the threshold temperature for accretion of "hot" gas via Bondi accretion (instead of "cold" gas via gravitational torques) had only a minor impact on the SFRD but increased the M_*/M_{halo} ratio and therefore decreased the efficiency of AGN feedback in the most massive halos, resulting in enhanced clustering of gas on scales $k \lesssim 1\,h\,\mathrm{Mpc}^{-1}$ relative to the fiducial model. It is also interesting to note that the variations of h and n_s in SIMBA have a significant impact on the gas power spectrum, while for TNG their effect on $P_{gas}(k)$ was negligible, despite h and n_s having similar large effects on the cosmic SFRD for both models. In SIMBA, the impact on $P_{gas}(k)$ is primarily driven by AGN jet feedback, which comes from the most massive BHs, whose growth and abundance are very sensitive to cosmology. The fact that different simulation models predict different responses of the gas (and hence the total matter) power spectrum to the same variation of the cosmological parameters stresses the importance of incorporating diverse galaxy formation models to improve and test the robustness of cosmological inference.

ORCID iDs

Yueying Ni https://orcid.org/0000-0001-7899-7195
Shy Genel https://orcid.org/0000-0002-3185-1540
Daniel Anglés-Alcázar https://orcid.org/0000-0001-5769-4945
Francisco Villaescusa-Navarro https://orcid.org/0000-0002-4816-0455
Yongseok Jo https://orcid.org/0000-0002-4728-6881
Simeon Bird https://orcid.org/0000-0001-5803-5490
Tiziana Di Matteo https://orcid.org/0000-0002-6462-5734
Rupert Croft https://orcid.org/0000-0003-0697-2583
Nianyi Chen https://orcid.org/0000-0001-6627-2533
Natalí S. M. de Santi https://orcid.org/0000-0002-4728-6881
Matthew Gebhardt https://orcid.org/0009-0003-0953-8931
Helen Shao https://orcid.org/0000-0002-0152-6747
Shivam Pandey https://orcid.org/0000-0001-5780-637X
Lars Hernquist https://orcid.org/0000-0001-6950-1629
Romeel Dave https://orcid.org/0000-0003-2842-9434

References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, PhRvD, 98, 043526
Anglés-Alcázar, D., Davé, R., Faucher-Giguère, C.-A., Özel, F., & Hopkins, P. F. 2017a, MNRAS, 464, 2840
Anglés-Alcázar, D., Faucher-Giguère, C.-A., Kereš, D., et al. 2017b, MNRAS, 470, 4698
Anglés-Alcázar, D., Özel, F., & Davé, R. 2013, ApJ, 770, 5
Banerjee, A., & Abel, T. 2021, MNRAS, 500, 5479
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013a, ApJ, 762, 109
Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013b, ApJ, 763, 18
Bird, S., Ni, Y., Di Matteo, T., et al. 2022, MNRAS, 512, 3703
Bondi, H. 1952, MNRAS, 112, 195
Booth, C. M., & Schaye, J. 2013, NatSR, 3, 1738
Borrow, J., Anglés-Alcázar, D., & Davé, R. 2020, MNRAS, 491, 6102
Butler Contreras, A., Lau, E. T., Oppenheimer, B. D., et al. 2023, MNRAS, 519, 2251

Chabrier, G. 2003, PASP, 115, 763
Christiansen, J. F., Davé, R., Sorini, D., & Anglés-Alcázar, D. 2020, MNRAS, 499, 2617
Conroy, C., van Dokkum, P. G., & Kravtsov, A. 2015, ApJ, 803, 77
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265
de Santi, N. S. M., Shao, H., Villaescusa-Navarro, F., et al. 2023, ApJ, 952, 69
Delgado, A. M., Angles-Alcazar, D., Thiele, L., et al. 2023, MNRAS, 526, 5306
Delgado, A. M., Wadekar, D., Hadzhiyska, B., et al. 2022, MNRAS, 515, 2733
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Natur, 433, 604
Faucher-Giguère, C.-A. 2020, MNRAS, 493, 1614
Feng, Y., Di-Matteo, T., Croft, R. A., et al. 2016, MNRAS, 455, 2778
Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, MNRAS, 503, 1940
Hahn, C., Villaescusa-Navarro, F., Castorina, E., & Scoccimarro, R. 2020, JCAP, 2020, 040
Hopkins, P. F. 2015, MNRAS, 450, 53
Hopkins, P. F., & Quataert, E. 2011, MNRAS, 415, 1027
Jeffrey, N., Alsing, J., & Lanusse, F. 2021, MNRAS, 501, 954
Jo, Y., Genel, S., Wandelt, B., et al. 2023, ApJ, 944, 67
Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Khandai, N., Di Matteo, T., Croft, R., et al. 2015, MNRAS, 450, 1349
Knollmann, S. R., & Knebe, A. 2009, ApJS, 182, 608
Krumholz, M. R., & Gnedin, N. Y. 2011, ApJ, 729, 36
Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, MNRAS, 432, 1544
Moser, E., Battaglia, N., Nagai, D., et al. 2022, ApJ, 933, 133
Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al. 2015, MNRAS, 454, 2691
Ni, Y., Di Matteo, T., Bird, S., et al. 2022, MNRAS, 513, 670
Nicola, A., Villaescusa-Navarro, F., Spergel, D. N., et al. 2022, JCAP, 2022, 046
Pandey, S., Lehman, K., Baxter, E. J., et al. 2023, MNRAS, 525, 1779
Parimbelli, G., Branchini, E., Viel, M., Villaescusa-Navarro, F., & ZuHone, J. 2023, MNRAS, 523, 2263
Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077
Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013, MNRAS, 430, 2427
Seljak, U., Aslanyan, G., Feng, Y., & Modi, C. 2017, JCAP, 2017, 009
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337

Shao, H., de Santi, N. S. M., Villaescusa-Navarro, F., et al. 2023, ApJ, 956, 149
Shao, H., Villaescusa-Navarro, F., Villanueva-Domingo, P., et al. 2022, arXiv:2209.06843
Springel, V. 2005, MNRAS, 364, 1105
Springel, V. 2010, MNRAS, 401, 791
Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776
Springel, V., & Hernquist, L. 2003a, MNRAS, 339, 289
Springel, V., & Hernquist, L. 2003b, MNRAS, 339, 312
Springel, V., Pakmor, R., Pillepich, A., et al. 2018, MNRAS, 475, 676
Springel, V., Pakmor, R., Zier, O., & Reinecke, M. 2021, MNRAS, 506, 2871
Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
Srisawat, C., Knebe, A., Pearce, F. R., et al. 2013, MNRAS, 436, 150
Thomas, N., Davé, R., Anglés-Alcázar, D., & Jarvis, M. 2019, MNRAS, 487, 5764
Tillman, M. T., Burkhart, B., Tonnesen, S., et al. 2023, ApJL, 945, L17
Vallés-Pérez, D., Planelles, S., & Quilis, V. 2022, A&A, 664, A42
van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2020, MNRAS, 491, 2424
Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021a, ApJ, 915, 71
Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., et al. 2021b, arXiv:2109.09747
Villaescusa-Navarro, F., Ding, J., Genel, S., et al. 2022a, ApJ, 929, 132
Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., et al. 2021c, arXiv:2109.10360
Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., et al. 2022b, ApJS, 259, 61
Villaescusa-Navarro, F., Wandelt, B. D., Anglés-Alcázar, D., et al. 2020, arXiv:2011.05992
Villanueva-Domingo, P., & Villaescusa-Navarro, F. 2022, ApJ, 937, 115
Villanueva-Domingo, P., Villaescusa-Navarro, F., Anglés-Alcázar, D., et al. 2022, ApJ, 935, 30
Villanueva-Domingo, P., Villaescusa-Navarro, F., Genel, S., et al. 2021, arXiv:2111.14874
Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, MNRAS, 436, 3031
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Wadekar, D., Thiele, L., Hill, J. C., et al. 2023, MNRAS, 522, 2628
Weinberger, R., Springel, V., Hernquist, L., et al. 2017, MNRAS, 465, 3291