이학박사 학위논문

Synergetic bridge of simulations, observations, and theory using machine learning

우주론적 모의 실험, 관측, 이론간의 상승 작용을 위한 기계 학습 활용

2022년 8월

서울대학교 대학원

물리천문학부

조 용 석

Synergetic bridge of simulations, observations, and theory using machine learning

우주론적 모의 실험, 관측, 이론간의 상승 작용을 위한 기계 학습 활용

지도교수 김 지 훈

이 논문을 이학박사 학위논문으로 제출함

2022년 7월

서울대학교 대학원
물리천문학부
조 용 석

조용석의 이학박사 학위논문을 인준함

2022년 7월

위 원 장	최 선 호	(인)
부위원장	김 지 훈	(인)
위 원	양 운 기	(인)
위 원	황 호 성	(인)
위 원	홍 성 욱	(인)

Abstract

Synergetic bridge of simulations, observations, and theory using machine learning

Yongseok Jo
Department of Physics and Astronomy
Graduate School
Seoul National University

Over the decades, the remarkable progress of cosmological simulations and observations has greatly extended our understanding of a wide variety of astrophysical and cosmological phenomena, such as the formation and evolution of large-scale structure, and cosmological simulations have become indispensable for studying astrophysics. However, simulations and observations have not been reconciled due to several factors, such as various uncertainties in both observation and simulation, limitations of physical models for both observation and simulation, computational cost of simulations, and the absence of delicate and sophisticated comparisons. Meanwhile, the emerging power of machine learning has shown the full potential to solve most of the problems above. Harnessing the power of machine learning, we aim to mainly address two issues: (1) we establish a model that can assist cosmological hydrodynamic simulation in a computationally-highly efficient way, and (2) we also build a pipeline that calibrates cosmological simulations against observations.

In the first part, we present a pipeline to estimate baryonic properties of a galaxy inside a dark matter (DM) halo in DM-only simulations using a machine trained on high-resolution hydrodynamic simulations. An extremely randomized tree (ERT) algorithm is used together with multiple novel improvements we introduce here such as a refined error function in machine training and two-stage learning. Aided by these improvements, our model demonstrates a significantly increased accuracy in predicting baryonic properties. By applying our machine to the DM-only simulation of a large volume, we then validate the pipeline that rapidly generates a galaxy catalogue from a DM halo catalogue. Our machine may become a promising method to transplant the baryon physics of galaxy-scale hydrodynamic calculations onto a larger-volume DM-only run. We discuss the benefits that machine-based approaches like this entail, as well as suggestions to raise the scientific potential of such approaches.

In the second part, employing simulation-based inference (SBI), also known as likelihood-free inference, we calibrate the parameters of cosmological simulations against observations, which has previously been unfeasible due to the high computational cost of these simulations. Using the cosmic star formation rate density (SFRD) and, separately, stellar mass functions (SMFs) at different redshifts, we perform SBI on select cosmological and astrophysical parameters ($\Omega_{\rm m}$, σ_8, stellar wind feedback and kinetic black hole feedback) and obtain full 6-dimensional posterior distributions. We find that there exist degeneracies between the parameters inferred from the SFRD, which is confirmed with new full cosmological simulations. We also find that the SMFs can break the degeneracy in the SFRD, which indicates that the SMFs provide stronger constraints for the parameters. Further, we find that the parameter set inferred from an observationally-inferred SFRD reproduces the target observed SFRD very well, whereas in the case of the SMFs, the inferred and observed SMFs show significant discrepan-

cies, pointing to limitations of the galaxy formation modelling framework.

Keywords : Cosmological simulation, Simulation-based inference, Galaxy formation and evolution, Large-scale structure formation
Student Number : 2014-22379

List of Contents

Chapter 1

Introduction

Since long before the development of notion of astrophysics and cosmology, the sky pitch-dark night has been adorned with countless magically-shining stars, and the mystery of somewhere at the edge of the sky has stimulated people's imagination. Questions about the extraterrestrial world beyond the Earth on which we are living have been constantly attracting people's attention amidst controversy such as geocentrism and heliocentrism, and long-standing questions about the universe are finally being bearing fruit in the 21st century.

Many observations and theories have made it possible to predict the origin and age of the universe, and provided clues to the formation of the universe. For instance, the discovery of the cosmic wave background radiation plays a major role in defining the *present* in which we live, drifting in the infinite space-time of the expanding universe, and makes it possible to study the density fluctuations of the early universe, which has become the basis for the growth of the universe after inflation. Furthermore, the development of observational instruments has led to the discovery of literally astronomical numbers of galaxies and galaxy clusters, which has made it possible to help understand the characteristic large-scale structure of the universe.

By combining observations and various theories, numerous theoretical models for formation and evolution of the universe that do not significantly violate observations and principal theories have been proposed. Among them, the most familiar ΛCMD cosmology is currently the most reliable theoretical model (i.e., most compatible with observations). In addition, since the 2011 Nobel Prize, the expanding universe has become a known fact, but the expansion rate, H_0, together with the cosmological pa-

rameters such as $\Omega_{\rm m}$ and σ_8, is still not well-constrained, and the precise estimation of the cosmological parameters remains as one of the biggest challenges in cosmology and astrophysics.

Cosmological simulation as virtual observatory

Behind this success story, there was an indispensable assistant, along with observation and theory, which is numerical simulation—hereafter simply simulation. Numerical simulation complements one of the most difficult aspects of cosmology and astrophysics study: irreproducibility of the universes in a laboratory. Simulation (or mock experiment), which has been developed since the mid-1990s, starting with E. Holmberg's laboratory experiment in 1941 [67], has grown into an indispensable existence in astrophysical research through the late 19's to the present.

In the early days of simulation, the number of particles, volume, and dimension that can be simulated were quite limited due to the computational limits. For example, in the mid-1900s, two-dimensional simulations only with hundreds of particles were mainly performed. However, with the rapid development of computers in the late 1900s, which was in line with Moore's law, simulations on a cosmological scale became doable. At that time, so-called cosmological N-body simulations that solve gravitational interactions were used to study the evolution of dark matter[1] from the beginning of the universe to the present. These cosmological simulations have played a significant role in studying the large-scale structure of the universe or the structure of individual halos.

Even so, up to that point, studies using N-body simulations that solve only gravitational interactions had been limited so far because the physics of baryons subject to hydrodynamics (or fluid dynamics) could not be ac-

[1] Dark matter is a hypothetical form of matter thought to account for approximately 85% of the matter in the universe (except dark energy). Dark matter is called "dark" because it does not appear to interact with the electromagnetic field, but only with gravitational field, and is, therefore, difficult to detect.

counted for. The development of computers enables combining N-body simulation with hydrodynamics, and it is possible to reproduce the observable properties of baryons in the universe. The advent of cosmological hydrodynamic simulations gave another impetus to numerical study in astrophysics, and has produced large-size, high-resolution simulated universes that can mimic the observed universe. Researchers have compared and analyzed the simulated universes with actual observations as if it were another virtual universe, which helps us improve our understanding of the universe.

However, cosmological hydrodynamic simulations also have two major drawbacks: (1) expensive computational cost and (2) ambiguity in subgrid models. Numerical hydrodynamics, unlike gravity, requires solving nonlinear physics that requires numerous computations, which limits the size and resolution of the simulation. In addition, due to the resolution limits, hydrodynamics simulations employ physical approximations—so-called subgrid model—to realize astrophysical phenomena, such as star formation and explosion, radiative cooling, black hole accretion, and AGN feedback. These subgrid models inevitably entail free parameters subject to calibration and lead to uncertainty of cosmological hydrodynamic simulations.

The high computational cost of hydrodynamic simulations leads to several limits. First, the size of the simulations is limited. For instance, cosmological N-body simulations are performed on a Gpc scale, while cosmological hydrodynamics simulation are performed on a Mpc scale. Note that the size of Milky Way Galaxy is roughly several hundreds kpc, while the Hubble horizon of the universe is ~ 4 Gpc. Secondly, the resolution of the simulations is limited. The simulation resolution is directly connected to accuracy of simulations. Thus, there are many attempts to reduce or overcome the computational cost of the hydrodynamics. In particular, in the era of machine learning, many studies have been conducted to overcome these computational limitations using machine learning.

Machine learning: computationally-efficient, versatile function

Since the notion of machine learning is first conceived in 1943 by logician Walter Pitts and neuroscientist Warren McCulloch in terms of mapping the decision-making process in human cognition and neural networks, machine learning along with its remarkable growth has established itself as a rapidly emerging and revolutionary keyword in almost all fields. The key idea of machine learning is to establish mapping from input to output directly from training data using parametrized functions such as decision trees and neural networks. Conceptually, it is not much different from function fitting out of data points. However, the application of various loss functions and optimization methods, such as a gradient descent, makes it able to handle high-dimension and highly-nonlinear problems. In addition, having many hyper-parameters—i.e., deeper neural networks—leads to expand the function space that the machines can have and increase the possibility that the optimal function that the problem can ideally have is inside the function space of the machines. In addition, computational efficiency is one of the major merits of machine learning. Once the machine is trained, the computation time of the machine is generally less than a second. Thus, there have been mounting studies that aim at building a machine that can estimate output of complex and computationally-heavy models by training it on the input-output pairs of the models.

In this dissertation, we summarize the physics of cosmological simulations in Chapter 2 and mathematics of machine learning in Chapter 3. We exploit machine learning to address the problem of computational cost of cosmological simulations: In Chapter 4, we transplant physics of small-size hydrodynamic simulations to large-size N-body simulations that lack of hydrodynamics. In Chapter 5, we calibrate simulation parameters against observations using simulation-based inference—one of the machine learning methods.

Chapter 2

Numerical framework for evolution of universe

In this chapter, we provide the preliminaries, including the implementations of the cosmological hydrodynamic simulations, for the main body of this dissertation.

2.1 Evolution of the universe

Initial conditions for cosmological simulations

In the early universe at $z \sim 1100$, recombination occurred—photons decouple from baryons—leading to small-amplitude *linear Gaussian* density fluctuations in baryons, photons, and dark matter based on the inflation model [61]. The *Gaussian* density fluctuations, $\delta(\mathbf{x}, t)$, can fully be described by its matter power spectrum $P(k,t) \equiv \int \mathrm{d}^3 r\, e^{i\mathbf{k}\cdot\mathbf{r}}\, \xi(r,t)$ where $\xi(r,t)$ is the two point correlation function, $\langle \delta(\mathbf{x},t)\delta(\mathbf{y},t)\rangle$, with $r = |\mathbf{x} - \mathbf{y}|$.

Initial conditions for cosmological simulations consist of the cosmological model and its perturbations. The cosmological model is generally taken to be a spatially flat or open Robertson-Walker spacetime with specified composition of dark matter, baryons, a possible cosmological constant. Once cosmological model is specified, the standard approach for generating initial conditions for simulations is to displace particles from a Cartesian lattice using the Zel'dovich approximation [202] as follows:

$$\mathbf{r} = \mathbf{q} + D(t)\boldsymbol{\psi}(\mathbf{q}),\ \mathbf{v} = a\frac{\mathrm{d}D}{\mathrm{d}t}\boldsymbol{\psi} \tag{2.1}$$

where $\mathbf{q}$ and $D(t)$ are the unperturbed lattice position and is linear growth

factor, respectively. $\boldsymbol{\psi}$ is the irrotational displacement field computed by the linear continuity equation

$$\nabla \cdot \boldsymbol{\psi} = -\frac{\delta}{D(t)}. \tag{2.2}$$

Accelerating expansion of the universe

Cosmological simulations generally evolve the initial density fluctuations of dark matters and baryons by solving the (Newtonian) gravity and (magneto-)hydrodynamics on a uniformly expanding Friedmann-Lemaître-Robertson-Walker spacetime. With the help of Enstein's general relativity, the evolution of the scale factor a that explains the expansion of the universe is given by the Friedmann equation as

$$\frac{H^2}{H_0^2} = \Omega_{\mathrm{r}} a^{-4} + \Omega_{\mathrm{m}} a^{-3} + \Omega_{\mathrm{k}} a^{-2} + \Omega_{\Lambda}. \tag{2.3}$$

Here, $H = \dot{a}/a$ the fractional expansion rate, and H_0 the Hubble constant today. Once the cosmological parameters are specified, dark matter and baryon, together with the scale factor a, can be evolved forward in time using Poisson's equation for the Newtonian gravity

$$\nabla^2 \Phi = -4\pi G a \left[\rho(\boldsymbol{x}, t) - \bar{\rho}(t)\right], \tag{2.4}$$

where $\bar{\rho}(t)$ is the mean density of the universe at time t.

2.2 Magneto-hydrodynamics

In contrast to dark matter that is subject to collisionless Boltzmann equation, physics of baryon—i.e., dynamics of gas—is highly non-linear and requires collisional Boltzmann equation (hearafter hydrodynamics). In addition to baryon, it is known that the magnetic field is ubiquitous in the universe, which can be approximately evaluated by magneto-hydrodynamics

(MHD). In the expansion of the universe, the equations of the ideal MHD can be written as [128]:

$$\frac{\partial \rho}{\partial t} + \frac{1}{a}\nabla \cdot (\rho \mathbf{v}) = 0, \tag{2.5}$$

$$\frac{\partial \rho(a\mathbf{v})}{\partial t} = \nabla \cdot \left(\rho \mathbf{v}\mathbf{v}^{\mathrm{T}} + \mathbf{I}p_{\mathrm{tot}} - \frac{\mathbf{B}\mathbf{B}^{\mathrm{T}}}{a} \right) = -\frac{\rho}{a}\nabla\Phi, \tag{2.6}$$

$$\frac{\partial (a^2 E)}{\partial t} + a\nabla \cdot \left[\mathbf{v}(E + p_{\mathrm{tot}}) - \frac{1}{a}\mathbf{B}(\mathbf{v}\cdot\mathbf{B}) \right] = \frac{\dot{a}}{2}\mathbf{B}^2 - \rho(\mathbf{v}\cdot\nabla\Phi) + a^2(\mathcal{H} - \Lambda), \tag{2.7}$$

$$\frac{\partial \mathbf{B}}{\partial t} + \frac{1}{a}\nabla \cdot (\mathbf{B}\mathbf{v}^{\mathrm{T}} - \mathbf{v}\mathbf{B}^{\mathrm{T}}) = 0, \tag{2.8}$$

where $p_{\mathrm{tot}} = p_{\mathrm{gas}} + B^2/2$ is the total gas pressure, and $E = \rho u_{\mathrm{th}} + \rho v^2/2 + B^2/2$ is the total energy per unit volume, with u_{th} denoting the thermal energy per unit mass. ρ, v and B represent the local gas density, velocity and magnetic field strength, respectively. Note that the ideal MHD requires several assumptions, such as quasi-neutrality of plasma and a low-frequency, long-wavelength approximation.

2.3 Physical models

Although principal equations—such as the Friedmann equation, Newtonian gravity, and hydrodynamics—are implemented and solved in cosmological simulations, the limited resolution of simulations constrains resolving small-scale physics. Thus, modern cosmological hydrodynamic simulations typically include a sub-resolution model for radiative cooling, formation, the explosion of stars, and formation and evolution of massive black holes. Among them, we introduce implementations of stellar and black hole physics in Sections 2.3.1 and 2.3.2.

2.3.1 Stellar physics

Stellar formation and evolution

The cosmological simulations that we use throughout this dissertation model the star-forming dense inter-stellar gas (ISM) using an effective equation of state, where stars form stochastically above a gas density of $\rho_{\rm sfr}$ with a star formation time scale of $t_{\rm sfr}$. the ISM is believed to be governed by small-scale effects, such as turbulence and thermal instability, leading to self-regulated equilibrium state of the ISM. Here, the adopted parameters are $\rho_{\rm sfr} = 0.13\,{\rm cm}^{-3}$ and $t_{\rm sfr} = 2.2{\rm Gyr}$. The formed stars represent a population of stars, not a single star, with an initial mass function given by Chabrier [25]. The stars are allowed to evolve and age, with a return of mass to the surrounding ISM as stellar winds. The detailed implementations of the evolution of stars can be found in Vogelsberger et al. [187], Wiersma et al. [201].

Stellar winds

At the end of the stellar evolution, stars return its mass as a stellar wind. The star formation-driven galactic wind is isotropically injected in a kinetic form [136, for details]. The winds are characterized by a mass loading factor η_w which is defined by

$$\eta_w \equiv \frac{\dot{M}_w}{\dot{M}_{\rm SFR}} = \frac{2}{v_w^2} e_w (1 - \tau_w), \tag{2.9}$$

where $\dot{M}_w$ and $\dot{M}_{\rm SFR}$ are the rate of gas mass to be converted into wind particles and the instantaneous, local star formation rate, respectively. With a fixed thermal energy fraction τ_w, the mass loading factor is determined by the total energy injection rate per unit star-formation e_w and the wind speed

v_w that involve $A_{\rm SN1}$ and $A_{\rm SN2}$, respectively as follows:

$$e_w = A_{\rm SN1} \times \bar{e_w} \left[f_{w,Z} + \frac{1 - f_{w,Z}}{1 + (Z/Z_{w,ref})^{\gamma_w,Z}} \right] \times N_{\rm SNII} E_{\rm SNII,51} 10^{51} {\rm erg M}_\odot^{-1} \tag{2.10}$$

and

$$v_w = A_{\rm SN2} \times \max \left[\kappa_w \sigma_{\rm DM} \left(\frac{H_0}{H(z)} \right)^{1/3}, v_{w,\rm min} \right], \tag{2.11}$$

where details on the parameters $\bar{e_w}$, $f_{w,Z}$, $Z_{w,ref}$, Z, $\gamma_{w,Z}$, $N_{\rm SNII}$, $E_{\rm SNII,51}$, κ_w, σ_8, and $v_{w,\rm min}$ can be found in Table 1 of Pillepich et al. 2018 [138]. Here, $A_{\rm SN1}$ and $A_{\rm SN2}$ are set to 1 in the chapter 4 but subject to change in the work of chapter 5.

2.3.2 Black hole physics

Various lines of observations have indicated the presence of black holes and strong feedback of massive black hole at the center of the galaxy, known as active galactic nuclei (AGN). However, the formation and evolution of black hole still remain as open question. Thus, many cosmological simulation typically implant a black hole seed at the center of a halo when the halo attains a certain mass threshold. In this dissertation, a super-massive black hole (SMBH) particle with mass $M_{\rm seed} = 8 \times 10^5 h^{-1} {\rm M}_\odot$ is seeded on-the-fly at the center of any halo with mass $M_{\rm halo} > 5 \times 10^{10} h^{-1} {\rm M}_\odot$ that does not yet contain a SMBH. To prevent the SMBH particles from artificially wandering around the galaxy, the SMBH particles are kept close to the potential minimum of their host dark matter haloes using an ad hoc prescription. Also, their velocities are set to the mean mass-weighted velocity of the region around the host haloes.

Black hole accretion

Once implanted, the SMBHs grow by accreting surrounding gas ac-

cording to Bondi-Hoyle-Lyttleton accretion [15, 16, 68]:

$$\dot{M}_{\rm Bondi} = \frac{4\pi G^2 M_{\rm BH}^2 \rho}{c_{\rm s}^3} \tag{2.12}$$

where $M_{\rm BH}$, ρ, and $c_{\rm s}$ are the mass of the SMBHs, gas density in the vicinity of the SMBHs, and sound speed of the gas, respectively. The accretion of SMBH is thought to form a gaseous disk around the SMBH, which emits radiation. The pressure of radiation suppresses gas accretion by the Thomson scattering, and when the accretion rate reaches a certain threshold, called Eddington limit where the pressure of radiation is equivalent to the gravitational pressure of the SMBH, the accretion of the SMBH halts:

$$\dot{M}_{\rm Edd} = \frac{4\pi G M_{\rm BH} m_{\rm p} c}{\epsilon_{\rm r} \sigma_{\rm T}} \tag{2.13}$$

where m_p, c, $\sigma_{\rm T}$, and $\epsilon_{\rm r}$ are the mass of a proton, the speed of light, and the cross section of Thomson scattering, respectively. Note that the power of radiation of accretion disk can be written as $\dot{E}_{\rm r} = \epsilon_{\rm r} \dot{M}_{BH} c^2$ where $\epsilon_{\rm r}$ is radiation efficiency. Therefore, the accretion rate of the SMBH is given by $\dot{M}_{\rm BH} = \min(\dot{M}_{\rm Bondi},\, \dot{M}_{\rm Edd})$.

Black hole kinetic feedback

The states of SMBHs are distinguished into high accretion (a classical thin disc) and low accretion (hot accretion flow) based on a threshold of

$$\chi = \min\left[0.002 \left(\frac{M_{\rm BH}}{10^8 {\rm M}_\odot}\right)^8, 0.1\right] \tag{2.14}$$

in units of Eddington accretion limit. According to the state of the accretion, the feedback mode is determined. For the high-accretion state, the feedback energy is injected as pure thermal energy into the vicinity of the SMBH (thermal mode). For the low-accretion state, feedback energy is released

kinetically in a random direction (kinetic mode) as

$$\dot{E}_{\rm low} = A_{\rm AGN1}\epsilon_{\rm f,kin}\dot{M}_{\rm BH}c^2, \tag{2.15}$$

where

$$\epsilon_{\rm f,kin} = \min\left[\frac{\rho}{0.05\rho_{\rm SFthresh}}, 0.2\right]. \tag{2.16}$$

Here, ρ and $\rho_{\rm SFthresh}$ are the gas density around the SMBH and the density threshold for star-formation. The injection of the kinetic feedback occurs every time the accumulated energy has reached the energy threshold since the last feedback. The energy threshold for the kinetic feedback is parametrized as

$$E_{\rm inj,min} = A_{\rm AGN2} \times f_{\rm re}\frac{1}{2}\sigma^2_{\rm DM}m_{\rm enc}, \tag{2.17}$$

where $\sigma^2_{\rm DM}$ is the one-dimensional dark matter velocity dispersion around the SMBH, $m_{\rm enc}$ is the enclosed gas mass within the feedback sphere, and $f_{\rm re}$ is a free parameter which is set to 20 for the fiducial TNG model. $A_{\rm AGN2}$ controls the frequency and speed of the SMBH feedback. $A_{\rm AGN1}$ and $A_{\rm AGN2}$ are fixed as 1 in Chapter 4 and subject to change in Chapter 5. The details of prescription for the SMBH physics are described in [194].

Chapter 3

Machine learning framework as bridge

In this chapter, we describe the machine learning frameworks that we employ in the main body of this dissertation.

3.1 Extremely randomized tree

Extremely randomized tree (ERT) is a commonly-used machine learning that combines the output of multiple decision trees to reach a single result. ERT is a branch of random forest (RF) algorithms which itself is a type of ensemble learning. We introduce the regressor's basic concept and inner workings here to later explain the improvements we made in the machine. At the heart of an ERT lies a "decision tree" that is constructed top-down from a root node. The tree partitions the data into subsets which contain instances of similar values; a (leaf) node generally has more than one attribute. A "forest" refers to an ensemble of decision trees — i.e., a collection of trees makes a forest. Compared to a plain RF, ERT's additional randomization step arises as the tree nodes are split (i.e., the points of split are randomly chosen), which makes an ERT perform mostly faster than a plain RF.

Here, we briefly summarize terminology used in this section: the attribute denotes a particular input variable. The candidate attributes stand for all input variables available. Output variables denotes the target variable for training. N refers to the size of the training sample.

Since the ERT is made of decision trees, it would be useful to describe the decision tree. Decision trees is a flowchart-like tree structure that helps decision-making process. Every node (or leaf) of decision trees asks "yes" or "no" questions—e.g., the attribute A is greater than B?—and splits the

Algorithm 1: Building Extra Tree Ensemble

Input: Training set $\mathcal{D}$
Initialize: Tree ensemble $\mathcal{T} = \{\}$
for $i = 1$ *to* N **do**
  Build a tree $T \cup \{t_i = \textbf{\textit{Build\_Tree}}(\mathcal{D})\}$;
end
Output: Tree ensemble $\mathcal{T}$

Build_Tree$(\mathcal{D})$
Input: Training set $\mathcal{D}$
if $\mathcal{D} < n_{\min} \| $*all candidate attributes are constant in* $\mathcal{D} \| $*the output variable is constant in* $\mathcal{D}$ **then**
  Output: D
else
  Select randomly K attributes, $\{a_1, \ldots, a_K\}$, without replacement, among all (non constant in $\mathcal{D}$ candidate attributes;
  Generate K splits, $\{s_1, \ldots, s_K\}$, where $s_i = \textbf{\textit{Pick\_Split}}(\mathcal{D}, a_i)$, $\forall i = 1, \ldots, K$;
  Select a split s_* s.t. $\mathcal{L}(s_*, \mathcal{D}) = \max_i \mathcal{L}(s_i, \mathcal{D})$;
  Split $\mathcal{D}$ into subsets $\mathcal{D}_l$ and $\mathcal{D}_r$ according to the test s_*;
  Build $t_l = \textbf{\textit{Build\_Tree}}(\mathcal{D}_l)$ and $t_r = \textbf{\textit{Build\_Tree}}(\mathcal{D}_r)$ from these subsets;
  Create a node with the split s_*, attach t_l and t_r as left and right subtrees of this node and return the resulting tree t.
end

Pick_Split$(\mathcal{D}, a)$
Input: Training set $\mathcal{D}$, attribute a
Find $a_{\max} = \max(a)$ and $a_{\min} = \min(a)$ $a \in \mathcal{D}$;
Draw a cut-point a_c uniformly in $[a_{\min}, a_{\max}]$;
Output: the split$[a < a_c]$

data accordingly into two subsets (left node and right node). Therefore, decision trees seek to find the best split to the data in a way that it can reach the best answer in the end. To best split the nodes, the ERT can be trained through the Classification and Regression Tree (CART) algorithm. Also, loss functions, such as Gini impurity, information gain, or mean square er-

ror (MSE), can be used to estimate the quality of the split. In this section, we adopt the loss function $\mathcal{L}$—called relative variance reduction—for two subsets of training data ($\mathcal{D}_r$ and $\mathcal{D}_l$) and split s defined as

$$\mathcal{L} = \frac{\mathrm{Var}_{\mathcal{D}}(y) - \frac{|\mathcal{D}_l|}{|\mathcal{D}|}\mathrm{Var}_{\mathcal{D}_l}(y) - \frac{|\mathcal{D}_r|}{|\mathcal{D}|}\mathrm{Var}_{\mathcal{D}_r}(y)}{\mathrm{Var}_{\mathcal{D}}(y)} \tag{3.1}$$

where $\mathrm{Var}_{\mathcal{D}}(y) \equiv \mathbb{E}_{\mathcal{D}}[(y-\bar{y})^2]$ is the variance of the output y in the sample $\mathcal{D}$ where $\bar{y}$ is the mean of y in sample $\mathcal{D}$. The "decision tree" is constructed to minimize $\mathcal{L}$ for training set $\mathcal{D}$. The minimization of the loss function $\mathcal{L}$ goes to the reduction of relative variance (e.g., $|\mathcal{D}_l|\mathrm{Var}_{\mathcal{D}_\updownarrow r}(y)/|\mathcal{D}|\mathrm{Var}_{\mathcal{D}}(y)$). If the reduction succeeds, we can split the data $\mathcal{D}$ into two subsets $\mathcal{D}_r$ and $\mathcal{D}_l$ that have similar y (or similar kinds) , respectively, in their own sets. This resembles the classification of y into two classes. Detailed algorithm of building trees can be found in Algorithm 1.

3.2 Simulation-based inference

Simulation-based inference (SBI) aims at retrieving the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$ using forward models that estimates observable $\boldsymbol{x}$ taking parameters $\boldsymbol{\theta}$ given observations $\boldsymbol{x}_0$ [29, 177]. In contrast to the conventional Bayesian inference that adopts an explicit likelihood, SBI employs a neural network that outputs a probability distribution—called a neural density estimator (NDE). The NDE takes observations $\boldsymbol{x}_0$ and parameters $\boldsymbol{\theta}$ as input and estimates a conditional probability distribution $\tilde{p}_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\boldsymbol{x}_0)$ where $\boldsymbol{\phi}$ represents hyper-parameters of neural network.

One of the important features is that SBI aims at training the NDE with the least number of evaluations of forward models by narrowing down the parameter space of interest. To this end, we must choose a proposal distribution $q(\boldsymbol{\theta})$ from which we draw the sets of parameters for the new simulations in the next iteration. However, the choice of the optimal proposal generally

remains as an open question.

3.2.1 Neural Density Estimator: Learning Density Ratio

The neural density estimator (NDE) estimates the posterior density $p(\boldsymbol{\theta}|\mathbf{x})$ by computing the density ratio $r(\boldsymbol{\theta}, \mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})}{p(\mathbf{x})} = \frac{p(\boldsymbol{\theta}|\mathbf{x})}{p(\boldsymbol{\theta})}$, where a simulator defines a valid probability density function $p(\mathbf{x}|\boldsymbol{\theta})$ over observations $\mathbf{x}$ [64, 113, 175].

Consider a binary random variable $Y : \Omega \to \mathbb{R}$ where $\Omega \in \{y = 0, y = 1\}$ and each outcome is equally likely a priori (i.e., $p(y = 0) = p(y = 1)$). Then,

$$\begin{aligned} p(y = 1|\boldsymbol{x}, \boldsymbol{\theta}) &= \frac{p(\boldsymbol{x}, \boldsymbol{\theta}|y = 1)}{p(\boldsymbol{x}, \boldsymbol{\theta}|y = 1) + p(\boldsymbol{x}, \boldsymbol{\theta}|y = 0)} \\ &= \frac{r(\mathbf{x}, \boldsymbol{\theta})}{1 + r(\mathbf{x}, \boldsymbol{\theta})} \end{aligned} \tag{3.2}$$

where the density ratio $r(\boldsymbol{x}, \boldsymbol{\theta})$ is defined by $r(\boldsymbol{x}, \boldsymbol{\theta}) \equiv \frac{p(\boldsymbol{x}, \boldsymbol{\theta}|y=1)}{p(\boldsymbol{x}, \boldsymbol{\theta}|y=0)}$ (refer to the derivation [1]). That is, the binary classifier that computes $p(y = 1|\boldsymbol{x}, \boldsymbol{\theta})$ or $p(y = 0|\boldsymbol{x}, \boldsymbol{\theta})$ can be exploited to compute the ratio density $r(\mathbf{x}, \boldsymbol{\theta})$.

In case of the conditional probability density, $y = 1$ class represents $(\boldsymbol{x}, \boldsymbol{\theta}) \sim p(\boldsymbol{x}, \boldsymbol{\theta})$ that $\boldsymbol{\theta}$ is drawn from the given prior $p(\boldsymbol{\theta})$ and $\boldsymbol{x}$ is obtained subsequently by the simulations with the sampled parameters. In the meantime, $y = 0$ class is pairs $(\boldsymbol{x}, \boldsymbol{\theta}) \sim p(\boldsymbol{x})p(\boldsymbol{\theta})$ with parameters and observations sampled independently. By training the neural classifier $\tilde{p}_\phi(\boldsymbol{x}, \boldsymbol{\theta})$ that takes $(\mathbf{x}, \boldsymbol{\theta})$ as input and outputs the probability density $p(y = 1|\boldsymbol{x}, \boldsymbol{\theta})$ or $p(y = 0|\boldsymbol{x}, \boldsymbol{\theta})$ where ϕ stands for hyper-parameters (e.g., weights and

[1] By definition of conditional probability, $p(y = 1|\boldsymbol{x}) = p(\boldsymbol{x}|y = 1)p(y = 1)/p(\boldsymbol{x})$. Using the law of total probability, we can rewrite $p(y = 1|\boldsymbol{x}) = p(\boldsymbol{x}|y = 1)p(y = 1)/(p(\boldsymbol{x}|y = 0)p(y = 0) + p(\boldsymbol{x}|y = 1)p(y = 1))$. Since each class is equally likely a priori, $p(y = 1|\boldsymbol{x}) = p(\boldsymbol{x}|y = 1)/(p(\boldsymbol{x}|y = 0) + p(\boldsymbol{x}|y = 1))$.

Algorithm 2: Sequential Ratio Estimation

Input: Simulator $\tilde{p}(\boldsymbol{x}|\boldsymbol{\theta})$, prior $p(\boldsymbol{\theta})$, target observation $\boldsymbol{x}_0$, neural density estimator $\tilde{p}_\phi(\boldsymbol{x}, \boldsymbol{\theta})$, iterations N, simulations per iteration M

Initialize: Proposal $q^{(1)}(\boldsymbol{\theta}|\boldsymbol{x}_0) = p(\boldsymbol{\theta})$, training set $\mathcal{T} = \{\}$

for $n = 1$ *to* N **do**

- Draw $\boldsymbol{\theta}_m \sim q^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0)$, $m = 1, \ldots, M$;
- Simulate $\boldsymbol{x}_m \sim \tilde{p}(\boldsymbol{x}|\boldsymbol{\theta}_m)$, $m = 1, \ldots, M$;
- Construct $\mathcal{T} = \mathcal{T} \cup \{(\boldsymbol{x}_m, \theta_m)|m = 1, \ldots, M\}$;
- **while** $\tilde{p}_\phi$ *not converged* **do**
  - Sample mini-batch $\{(\boldsymbol{x}_b, \boldsymbol{\theta}_b)\} \sim \mathcal{T}$;
  - Optimize a neural density estimator $\tilde{p}_\phi(\boldsymbol{x}, \boldsymbol{\theta})$ using stochastic gradient descent;
- **end**
- Save the neural density estimator $\tilde{p}_\phi(\boldsymbol{x}, \boldsymbol{\theta})$
- Update posterior $p^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0) \propto \tilde{p}_\phi(\boldsymbol{x}_0|\boldsymbol{\theta})p(\boldsymbol{\theta})$
- Update proposal $q^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0) \propto p^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0)p(\boldsymbol{\theta})$

end

Output: Posterior $p^{(N)}(\boldsymbol{\theta}|\boldsymbol{x}_0) \propto \tilde{p}_\phi(\boldsymbol{x}_0|\boldsymbol{\theta})p(\boldsymbol{\theta})$

biases) of the neural network, we can compute the ratio density

$$r(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})p(\boldsymbol{\theta})} = \frac{p(\mathbf{x}|\boldsymbol{\theta})}{p(\mathbf{x})} = \frac{p(\boldsymbol{\theta}|\mathbf{x})}{p(\boldsymbol{\theta})}. \tag{3.3}$$

Therefore, if the prior $p(\boldsymbol{\theta})$ is known, the posterior density $p(\boldsymbol{\theta}|\mathbf{x})$ can be obtained as $p(\boldsymbol{\theta}|\mathbf{x}) = r(\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})$.

3.2.2 Sequential Ratio Estimation

In this section, we introduce the sequential ratio likelihood estimation (SRE) [47], which is one of the simulation-based inference (SBI) methods. The SRE retrieves a posterior density using the neural density estimator described in Sec. 3.2.1. The workflow of the SRE is as follows: (1) The SRE samples M parameters $\boldsymbol{\theta}_m$ from the prior $p(\boldsymbol{\theta})$ (or proposal density) and simulates M observables $\boldsymbol{x}_m$ with the parameters $\boldsymbol{\theta}_m$ where $m = 1, \ldots, M$;

(2) The SRE constructs a training set $\mathcal{T}$ that consists of two classes of pairs: $(\boldsymbol{x}_m, \boldsymbol{\theta}_m)$ for $y = 1$ and $(\boldsymbol{x}_k, \boldsymbol{\theta}_m)$ for $y = 0$ by permutations where $k \neq m$. (3) The NDE $\tilde{p}_\phi(\boldsymbol{x}, \boldsymbol{\theta})$ is trained on $\mathcal{T}$ until it converges; (4) The structure of trained NDE is saved as an external file for future use. (5) The posterior and the proposal densities are updated as $p^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0) \propto \tilde{p}_\phi(\boldsymbol{x}_0|\boldsymbol{\theta})p(\boldsymbol{\theta})$ and $q^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0) \propto p^{(n)}(\boldsymbol{\theta}|\boldsymbol{x}_0)p(\boldsymbol{\theta})$, respectively. (6) The SRE repeats from (1) with the newly updated proposal. Please refer to Algorithm 2 for details.

The SRE has two hyper-parameters subject to optimization: the number of simulations per iteration M and the number of iterations N. Since M determines the amount of information that the training set can carry, it can vastly affect the accuracy of the posterior density. For instance, if M is over-biased, the resultant posterior density can also be biased. M should be set to a sufficiently large value that the sampled training set can carry enough information about the trained posterior.

Chapter 4

Part I: Painting cosmological simulations with baryons

The results of this chapter are published in Monthly Notices of the Royal Astronomical Society, 2019, Volume 489, Issue 3.

4.1 Introduction

In the context of the ΛCMD cosmology, the gravitational N-body simulations that contains only dark matter (DM) to describe the evolution of large scale structures (LSS) are successfully performed [7, 18, 62, 71, 82, 146, 160, 193]. DM-only simulations has also provided considerable insights into halo properties, such as the spatial and velocity correlations [72, 198, 199], density profiles [23, 81, 120, 142, 143], angular momentum profiles and shapes [14, 22, 27, 97], and halo substructures [79, 105, 114, 171].

However, gravity alone is not sufficient to extend our understanding of the universe beyond the DM properties which is generally not observable. Baryon physics must be taken into account via one of the two popular methods: (magneto-)hydrodynamic simulations, or semi-analytic models (SAMs). On the one hand, with the advent of high-performance computing units with a large amount of memories, fully hydrodynamics, high-resolution cosmological simulations have become one of the major tools in studying baryonic contributions in the Universe's evolution. Hydrodynamic simulations that treat baryon physics such as individual galaxy formation from $\sim$Mpc scales down to $\lesssim$100 pc scales have emerged in recent years despite the expensive computational costs. Prominent examples includes ILLUSTRIS [56, 189, 191], ILLUSTRISTNG [123, 137, 172];, HORIZON-

AGN [45], EAGLE [156], ROMULUS [180], MUFASA [36] and SIMBA [37]. On the other hand, in SAMs and empirical models, halos from DM-only simulations are "colored" with baryons based on relatively simple physical recipes [9, 10, 12, 28, 31, 115, 150]. While SAMs inevitably require a set of tunable parameters, the computational cost of typical SAMs is much less than that of high-resolution hydrodynamic simulations. In addition, SAMs make it easy to test and appreciate the importance of physical interactions and parameters in play [159].

Even with the cutting-edge computing technologies that have allowed us to simulate individual galaxies with high fidelity, the contemporary computation power is insufficient to describe a larger volume of the Universe (i.e., $\sim$Gpc scale) with detailed baryon physics resolved at $\lesssim$100 pc resolution. To obtain "observable" baryonic signatures populating such a large volume, combining DM-only simulations with a SAM has traditionally been the only strategy that is computationally feasible. But, now with the arrival of machine learning technology, preliminary studies have been carried out to combine DM-only simulations with machine learning algorithms such as random forest (RF) to produce galaxy catalogues [1, 73, 74].

Here, in what we call a machine-assisted semi-simulation model (MSSM), a machine — suitable for big data regression — is trained to first establish correlations between DM and baryonic properties in fully hydrodynamic simulations (e.g., DM mass and stellar mass in a halo). The machine is then tested and used to estimate various baryonic properties of a DM halo (either in hydrodynamic simulations or in DM-only simulations) based purely on its DM content. A well-constructed machine can generate an extensive galaxy catalogue out of a DM-only simulation of a large volume, within a fraction of time needed for a high-resolution hydrodynamic simulation. Furthermore, this method can be one of the most promising ways to accurately transplant the baryon physics of galaxy-scale hydrodynamic calculations (e.g., ILLUSTRISTNG in a $(75\ h^{-1}\mathrm{Mpc})^3$ volume) onto a larger-volume

DM-only simulation (e.g., MULTIDARK-PLANCK in a $(1\ h^{-1}\mathrm{Gpc})^3$ volume [81]). Training the machine with a RF-type algorithm, we could also grasp the degree of contribution or "feature importance" by each of the input features (e.g., halo mass vs. halo angular momentum) in estimating a particular property (e.g., stellar mass). From the intuition gained by feature importances and by comparing the resulting catalogues with SAMs', we will be able to provide insights to improve the SAMs as well.

In this chapter, we first describe how we establish the MSSM in Section 4.2.1. The cosmological simulations that we adopt are explained in Section 4.2.2, and its pre-processing for training and application is described in Section 4.2.3. We then focus on improving the machine training for MSSM, and compare our machine's accuracy with a simpler baseline model's (Section 4.3.1). Lastly in Section 4.3.2, our machine learning and application pipeline, MSSM, is shown to be largely compatible with popular SAMs when generating a galaxy catalogue using the DM-only simulation database MULTIDARK-PLANCK.

4.2 Methodology

In this section, we describe the pipeline of our model and how we build and train our machine. In particular, we focus on the structure of our pipeline, and how we pre-process the input dataset to improve the machine's accuracy.

Overview

In our so-called MSSM, we exploit the results of full hydrodynamic, high-resolution simulations to establish correlations or mappings — not analytic prescriptions — between DM and baryonic properties. Machine learning means training a machine for a task that typically deals with a large amount of data. If we assign two sets of data as "input" and "output", the machine by itself searches for a model and model parameters to take in

the input and produce the output. In general, the more amount of data one gives, the more accurate the model becomes. The large datasets from modern cosmological simulations are thus ideal to exploit the novelty of machine learning.

In the *supervised* learning phase of our work,[1] we first divide the halo-galaxy catalogue from a large hydrodynamic simulation into a "training set" and a "test set" (see Section 4.2.2). The machine learns a structure that maps an input to an output based on example input–output pairs, i.e., the training set (e.g., DM mass and stellar mass). The machine looks for an optimized mapping by constantly evaluating the current mapping with an "error function" (or "cost function"; e.g., a widely used metric in public packages is mean square error or MSE). Based on this evaluation, the machine returns positive or negative feedback to itself. When the training is completed, one can "score" how well the machine can match the *actual* features in the simulation using the test set. Based on this score, one may choose to update the learning algorithm or replace it with a different method.

4.2.1 Machine-assisted semi-simulation model

The flowchart of our MSSM, the machine learning and application pipeline, is illustrated in Figure 4.1. Our goal is to construct a machine to produce a galaxy catalogue by combining a DM-only N-body simulation and a machine learning technique, that is on a par with or better than catalogues made with popular SAMs. Our pipeline is divided into two main parts — (1) the learning phase: train a machine to estimate baryonic data out of DM data using a fully hydrodynamic simulation, and (2) the application phase: apply the trained machine to a DM-only simulation to produce catalogues of galactic baryonic properties.

[1] Machine learning algorithms are divided into several categories based on the amount and type of supervision in training: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

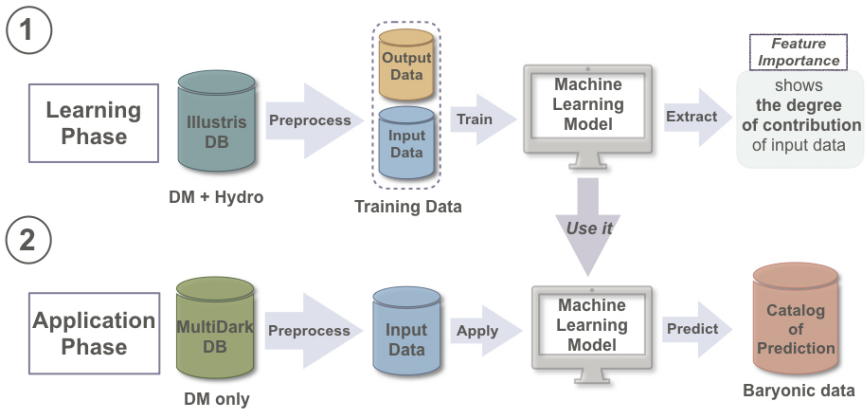

Figure 4.1: Flowchart of our machine-assisted semi-simulation model (MSSM). In the learning phase (*top panel*), we train our machine with a fully hydrodynamic simulation database that contains both dark matter (DM) and baryon data to predict the baryonic properties ("output") based on the DM properties ("input"). In the application phase (*bottom panel*), by feeding a DM-only N-body simulation to the trained machine, we produce a catalogue of baryonic predictions.

Learning phase

In the learning phase, we use only the DM-related features extracted from the IllustrisTNG hydrodynamic simulation of a $(75\ h^{-1}\mathrm{Mpc})^3$ volume ("TNG100" [124]; see Section 4.2.2 for more information) as input data. We take these DM features such as DM halo mass and halo velocity dispersion as inputs, and the baryonic features such as stellar mass and gas mass of the halo as desired outputs. These input–output pairs is used to train the machine via `ExtraTreeRegressor` (Section 3.1). Note that several historical and environmental characteristics of each halo not included in the native catalogue are computed in the pre-processing step (see Section 4.2.3 and Table 4.1 for more information). During the training process, 20% of the IllustrisTNG data is spared for a test — a "test set" — to score the accuracy of the machine afterwards. Fed with the test set, the resulting ma-

chine makes a set of predicted output data (e.g., stellar masses predicted from DM masses); and, by comparing it with the *actual* values in the simulation (e.g., the *actual* stellar masses in ILLUSTRISTNG) we "score" the machine. Common metrics for scoring the linear regression are MSE and Pearson correlation coefficient (PCC); but, in the reported study different measures are also used to evaluate the machine accuracy.

It is also worth to mention that ERT in our MSSM not only builds a map connecting inputs and outputs, but also provides the "feature importance" that shows which input feature contributes how much to predict a particular output (e.g., which input feature is more important to predict stellar mass, halo mass or halo angular momentum?). From the feature importance we may update the set of input parameters to increase the machine's accuracy, understand the underlying physics, and potentially provide insights to improve SAMs (see also Section 4.4).

Application phase

In the application phase, the machine from the learning phase is fed with a DM-only simulation data. Here, the MULTIDARK-PLANCK DM-only simulation of a large $(1\ h^{-1}\mathrm{Gpc})^3$ volume is used as an input ("MDPL2" [83]; see Section 4.2.2 for more information). Needless to say, this input data needs to be pre-processed so that it is exactly in the same format and structure as the input used in the learning phase (Section 4.2.3). A well-optimized machine can swiftly generate a galaxy catalogue once the DM-only simulation dataset is pre-processed. In our study, the machine is able to "paint" baryonic features on $\sim 10^6$ halos in a large cosmological volume in just a few tens of minutes. This is a miniscule amount of time when contrasted with what is typically needed for a high-resolution hydrodynamic simulation that resolves each galaxy-size halo with proper baryon physics. In Section 4.4, we will discuss more on how to utilize MSSM for science.

Table 4.1: All DM-related input parameters utilized to predict baryonic properties of a halo in our machine.

Input Parameter	Definition
DM mass of a halo	Total mass of all DM particles bound to a halo
Velocity dispersion of a halo	Dispersion of all member particles' velocities
Maximum velocity of a halo	Maximum of spherically-averaged circular velocity
Angular momentum of a halo	Halo spin parameter
Number of all mergers	Number of all mergers throughout the halo's entire history
Number of all major mergers	Number of all mergers in which the mass ratios of the participating halos are less than 3:1
Last major merger mass ratio	The mass ratio of the most recent major merger along the merger tree
Local density	The local density, $(\sum M_i)/V_{\rm box}$, estimated for all local halos within a $(2\,{\rm Mpc})^3$ volume
Number of local halos	Number of all local halos whose mass is larger than 80% of the target halo's mass
Sum of mass over distance	Sum of mass over distance, $\sum M_i/R_i$, of all local halos within a $(2\,{\rm Mpc})^3$ volume
Maximum mass over distance	Mass over distance, $M_{\rm max}/R_{\rm max}$, for the most massive halo in the local volume

4.2.2 Cosmological simulations

As noted in Section 4.2.1 and Figure 4.1, two types of simulations are considered in our MSSM pipeline — (1) in the learning phase: a fully hydrodynamic simulation is used to train a machine, and (2) in the application phase: the trained machine is applied to a DM-only simulation to produce galaxy catalogues.

Hydrodynamic simulation for learning phase

ILLUSTRISTNG [123, 137] are gravito-hydrodynamic simulations performed with a moving-mesh code AREPO [164] (see Chapter 2 for the detailed implementation). Both simulations include all relevant galaxy-scale physics to follow the evolution of dark matter, cosmic gas, stars and super massive black holes (SMBHs) from $z = 127$ to 0, such as radiative gas cooling [75, 155, 166, 200], stellar evolution and chemical enrichment based on stellar synthesis models [201], stellar feedback [165] and SMBH and Active Galactic Nuclei (AGN) feedback [168, 169]. The more recent ILLUTRISTNG (The Next Generation) updates ILLUSTRIS by including magneto-hydrodynamics [119, 129, 130, 195, 196], various computational improvements [137], as well as updated cosmology consistent with Planck Collaboration [141]: $\Omega_{\mathrm{m},0} = 0.3089, \Omega_{\Lambda,0} = 0.6911, \Omega_{\mathrm{b},0} = 0.0486, \sigma_8 = 0.8159, n_{\mathrm{s}} = 0.9667$, and h =0.6774.

ILLUSTRISTNG is one of the most successful hydrodynamic calculations to date resolving individual galaxies with sophisticated baryon physics in a large enough volume. For this reason, we employ ILLUTRISTNG in the learning phase of our MSSM pipeline (Section 4.2.1). In particular, among the three different box sizes the ILLUTRISTNG database offers, the "TNG100" simulation of a $(75\ h^{-1}\mathrm{Mpc})^3$ volume is adopted ("TNG100" dataset as designated in Nelson et al. [124]), where 100 denotes the simulation's approximate box size in Mpc). The TNG100 run was performed at three different resolutions: TNG100-1, -2 and -3 with TNG100-1 being the highest resolution run. At $z = 127$, the TNG100-1 data consists of 1820^3 DM particles with $m_{\mathrm{DM}} = 7.5 \times 10^6\,\mathrm{M}_\odot$, and 1820^3 hydrodynamic cells with $m_{\mathrm{gas}} = 1.4 \times 10^6\,\mathrm{M}_\odot$. At $z = 0$ the simulation box holds 4371211 (sub)halos identified with the friends-of-friends halo finder (FoF) [39] and the SUBFIND subhalo finder [167]. The publicly available halo catalogue also includes the merger trees built with the SUBLINK code [148].[2]

[2]The ILLUSTRISTNG data is available at http://www.tng-project.org/.

DM-only simulation for application phase

MULTIDARK-PLANCK [81, 147, 149] is a DM-only gravitational dynamics simulation using L-GADGET-2, a version of GADGET-2 optimized for a run with large number of particles [163]. The cosmological model adopted is consistent with Planck Collaboration [140]: $\Omega_{\mathrm{m},0} = 0.3071$, $\Omega_{\Lambda,0} = 0.6929$, $\Omega_{\mathrm{b},0} = 0.0482$, $\sigma_8 = 0.8228$, $n_{\mathrm{s}} = 0.96$, and h =0.6777.

In the application phase of our MSSM, the later version of MULTIDARK-PLANCK is used as an input ("MDPL2" [83]). Run on a volume of $(1\ h^{-1}\mathrm{Gpc})^3$ that is large enough to match observational surveys, MDPL2 depicts the large-scale evolution of a significant chunk of the Universe from $z = 65$ to 0 using 3840^3 DM particles with $m_{\mathrm{DM}} = 1.5 \times 10^9 h^{-1}\mathrm{M}_\odot$ each. The MDPL2 database publicly provides a halo catalogue for each redshift snapshot identified with the ROCKSTAR code, along with the merger trees built with the CONSISTENT TREES code [11].[3]

4.2.3 Pre-processing simulation dataset

Data pre-processing is a pivotal step in machine learning. As noted in Figure 4.1, we transform the raw database — the ILLUSTRISTNG data for the learning phase, and the MULTIDARK-PLANCK data for the application phase — into a desired input format for the machine.

Pruning the input datasets

Due to resolution discrepancy between MULTIDARK-PLANCK and ILLUSTRISTNG, we need to trim input datasets accordingly to reconcile it. MDPL2 resolves dark matter halos down to $\sim 2.23 \times 10^9\,\mathrm{M}_\odot$. TNG100-1 resolves dark matter halos down to $7.5 \times 10^6\,\mathrm{M}_\odot$ while resolving baryon down to $1.4 \times 10^6\,\mathrm{M}_\odot$. Therefore, we exclude the halos of masses below $10^9\,\mathrm{M}_\odot$ in TNG100-1 to be compatible with MDPL2. In addition, since ha-

[3]The MULTIDARK-PLANCK data can be found in the COSMOSIM database at http://www.cosmosim.org/.

los which do not contain star or gas are not our targets of interest, we have excluded halos whose stellar or gas mass is zero. With these cuts, the actual training set for the learning phase is reduced to $\sim 3\%$ of the original TNG100-1 halo catalogue. In Section 4.2.1, we demonstrate that this training set is still sufficiently large for our learning process.

Extracting historical and environmental factors

The "baseline" input features to predict baryonic properties include: DM mass, velocity dispersion, and maximum circular velocity of a halo (see Table 4.1). This set of parameters — straight from public halo catalogues — is largely what prior attempts have used [74]. In addition to the baseline parameters, in the present study we aim to capture what we refer to as "historical" and "environmental" factors, and add them to the input dataset. The new features for each halo are extracted (1) from the halo's merger history, and (2) from the halo's local volume.

First, from the halo's merger tree, the following three features are obtained (Table 4.1): the number of all mergers, the number of all major mergers, and the mass ratio of the last major merger. These characteristics are chosen to *explicitly* quantify the evolution history of a halo imprinted in the merger tree (unlike Agarwal's work [1] where the merger-related parameters are implicit). Here, the mass ratio of participating halos must be less than 3:1 to be considered as major merger. Analogous to Rodriguez-Gomez's work [148], the mass ratio is calculated when the secondary progenitor reaches its maximum halo mass, $t_{\rm max}$, before the two halos merge into one in the tree. We take this point $t_{\rm max}$ as the moment of merger.

Second, from the target halo's local volume of $(2\,{\rm Mpc})^3$, the following four features are extracted (Table 4.1): the local density, the number of local halos whose masses are greater than 80% of the target halo's mass, the sum of mass over distance ("semi-potential") of all local halos $\Phi_{\rm s} = \sum M_i/R_i$, and the mass over distance for the most massive local halo. These parame-

ters aim to characterize the target halo's local environment which has likely affected how the halo has evolved. Extracting these features from the raw dataset leads to the nearest neighbor search and range search problem. It requires us to construct a *k*-d tree that partitions the space into tree structure so that neighboring halos are efficiently located.

Indeed, the *value-added* input datasets containing the additional input features improve the MSSM's accuracy for several feature predictions. This will be discussed in detail in Section 4.4.

4.3 Results

4.3.1 Improving a machine that predicts baryonic properties

In Sections 4.3.1 and 4.3.2, we present the results of our study focusing on the learning phase and the application phase of the MSSM pipeline (Figure 4.1), respectively. For the rest of the paper, unless the redshift of the data is specified, we only discuss the $z = 0$ result. We also note that we will focus on the halos of DM masses in the range of approximately $[10^{10}, 10^{13.5}]\,\mathrm{M}_{\odot}$ when presenting our results (but *not* necessarily when training the machine; see Section 4.2.3). It is because (1) for halos of DM masses below $10^{10}\,\mathrm{M}_{\odot}$, the resolutions of ILLUSTRISTNG (Section 4.3.1) and MULTIDARK-PLANCK simulations (Section 4.3.2) are too coarse for the machine to extract reliable mappings between DM and baryonic features, and (2) ILLUSTRISTNG contains insufficient number of halos of DM masses above $10^{13.5}\,\mathrm{M}_{\odot}$ due to a small simulation box size. It should be noted that the limitation here is not about our model but about the available simulations; $[10^{10}, 10^{13.5}]\ \mathrm{M}_{\odot}$ is indeed also the range for which the SAMs are best optimized.

How accurate is the machine's prediction?

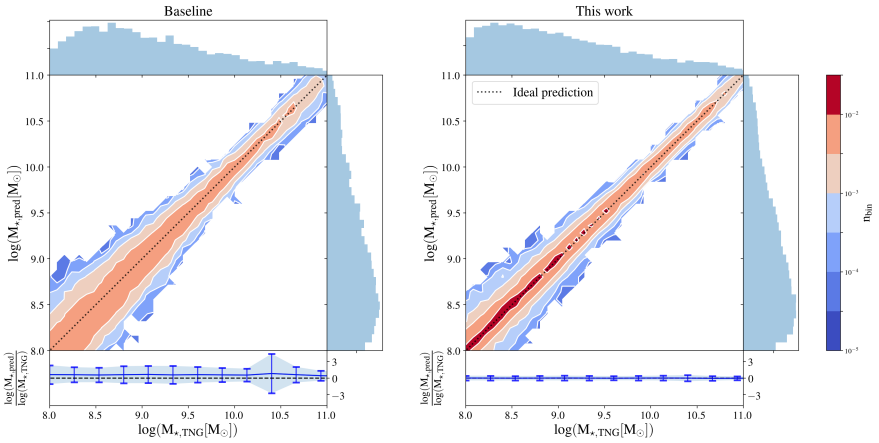

Figure 4.2: Normalized two-dimensional contour plot comparing the *actual* stellar masses of halos in the IllustrisTNG test set, $M_{\star,\mathrm{TNG}}$, and the stellar masses predicted from input DM features of the test set, $M_{\star,\mathrm{pred}}$. Colors indicate the normalized frequency, $n_{\mathrm{bin}} = N_{\mathrm{bin}}/N_{\mathrm{tot}}$, where N_{tot} is the total number of halos and N_{bin} is the number of halos in each two-dimensional bin. Results from two machine learning models are shown: the "baseline" model similar to previous studies (*left panel*) and our model improved for its performance (*right panel*). The black dotted line indicates an ideal prediction, $M_{\star,\mathrm{pred}} = M_{\star,\mathrm{TNG}}$.

We first discuss how well our machine from the learning phase can predict halos' baryonic properties based purely on their DM features. Shown in Figure 4.2 are normalized two-dimensional histograms comparing the predicted stellar masses ("predicted output") and the *actual* stellar masses in a simulation ("desired output" or "answer"), when a test set from the IllustrisTNG run is used. First, shown on left is the "baseline" model that considers only mass, velocity dispersion, and maximum circular velocity of a DM halo as inputs (similar to previous studies such as Kamdar's work [74]; see Section 4.2.3). Shown on right is our model that improves the baseline one in various ways to be discussed in the following sections. We test both models to predict the following baryonic properties: gas mass, stellar mass, central black hole mass, star formation rate (SFR), metallicity, and

	Gas mass	Stellar mass	BH mass	SFR	Metallicity	Stellar mag. g
Baseline	0.0015	0.0018	0.0047	1.71	0.022	0.0012
	(0.023)	(0.017)	(0.020)	(36.10)	(0.099)	(0.0121)
Logarithmic scaling	0.0010	0.0045	0.0126	1.70	0.010	–[5]
	(0.021)	(0.017)	(0.025)	(30.42)	(0.076)	(–)
history and environment	0.0014	0.0014	0.0042	1.5	0.018	0.0010
	(0.023)	(0.016)	(0.018)	(28.27)	(0.093)	(0.0100)
Two-stage learning	0.0014	0.0016	0.0036	1.11	0.013	0.0005[11]
	(0.021)	(0.011)	(0.017)	(20.15)	(0.078)	(0.0064)
Best combination	0.0010	0.0013	0.0034	1.00	0.010	0.0005
	(0.020)	(0.010)	(0.016)	(20.23)	(0.070)	(0.0053)

Table 4.2: The mean binned error (MBE), Equation 4.4, quantifying how well the machine predicts each of the six baryonic properties — gas mass, stellar mass, central black hole mass, SFR, metallicity, and stellar magnitude (g band). Each row indicates the MBE score within the respective x-range in Figure 4.3 when the machine is improved by a single improvement—except the "Best combination" row[9]. Numbers in the parentheses are mean binned standard deviation (MBSD), Equation 4.5.

stellar magnitudes.[4]

Both histograms in Figure 4.2 are around the ideal prediction line (*black dotted*), but in the bottom panels, the distribution is markedly tighter in our improved model resulting in the emergence of more concentrated region (*red region*) around the ideal prediction line. To quantify the machine's accuracy we first score each model with two common measures — (1) mean square error (MSE),

$$\text{MSE} = \frac{1}{N_{\text{tot}}} \sum_{i}^{N_{\text{tot}}} \left(y^{i}_{\text{pred}} - y^{i}_{\text{TNG}} \right) , \tag{4.1}$$

[4]Stellar magnitudes are the luminosities of all star particles in eight photometric bands — *U, B, V, K, g, r, i, z* [122].

and (2) Pearson correlation coefficient (PCC),

$$\mathrm{PCC} = \frac{\mathrm{cov}\left(y_{\mathrm{pred}}\,,\, y_{\mathrm{TNG}}\right)}{\sigma_{y_{\mathrm{pred}}}\sigma_{y_{\mathrm{TNG}}}}\,, \tag{4.2}$$

where $\mathrm{cov}(A, B)$ is the covariance of two variables A, B and σ is the standard deviation. In both equations, y^i_{pred} is the predicted *logged* output, and y^i_{TNG} is the desired *logged* output in the simulation. Note that we take the logarithm of the output data—except for stellar magnitudes where y^i_{pred} and y^i_{TNG} are simply the raw data (i.e., not logged).[5]

We find that both measures are significantly improved in our model: MSE decreased from 2.0×10^{-2} to 1.9×10^{-4}, and PCC increased from 0.971 to 0.987.

We have also tried — and eventually adopted — another metric to measure the machine accuracy.[6] To compute what we call the "mean binned error" (MBE), first, the predicted and desired output pairs, $\left(y^i_{\mathrm{pred}}, y^i_{\mathrm{TNG}}\right)$, are binned into $\mathcal{N}_{\mathrm{bins}}$ bins according to the y^i_{TNG} values. Then, in each bin, the *normalized* mean error is

$$\Gamma_j = \frac{1}{N_j}\sum_i^{N_j} \frac{\left|y^i_{\mathrm{pred}} - y^i_{\mathrm{TNG}}\right|}{y^i_{\mathrm{TNG}}}\,, \tag{4.3}$$

where N_j is the number of data in the *j*-th bin. Finally, by averaging Γ_j's across all bins we obtain the MBE as

$$\mathrm{MBE} = \frac{1}{\mathcal{N}_{\mathrm{bins}}}\sum_j^{\mathcal{N}_{\mathrm{bins}}} \Gamma_j\,. \tag{4.4}$$

If we replace the mean error in each bin, Γ_j, with the standard deviation

[5] Unlike other baryonic properties we consider, the stellar magnitudes are already logged and lie in the range of [-25, -13].

[6] This is inspired by the case in which MSE or PCC does not aptly represent the entire $y^i_{\mathrm{pred}} - y^i_{\mathrm{TNG}}$ distribution — i.e., PCC can be high even when the datapoints are widely spread out around the $y^i_{\mathrm{pred}} = y^i_{\mathrm{TNG}}$ line in Figure 4.2.

in each bin, s_j, then we acquire another accuracy measure "mean binned standard deviation" (MBSD),

$$\mathrm{MBSD} = \frac{1}{\mathcal{N}_{\mathrm{bins}}} \sum_{j}^{\mathcal{N}_{\mathrm{bins}}} s_j \,. \tag{4.5}$$

We find that, in general, MBE captures the accuracy of a trained machine better than other metrics do. When predicting stellar masses, our model improves the MBE score from the baseline model's 0.0018 to 0.0013, and MBSD from 0.017 to 0.010. We will extensively use MBE and MBSD in Table 4.3.1.

In addition to reducing the machine accuracy down to a numeric score, we also inspect the machine's performance across the output's entire value range. In Figure 4.3, for six baryonic properties we predict, we compare the probability distribution functions (PDFs) of the two machine learning models, and the *actual* data in the simulation.[7] Again, both the baseline (*blue dot dashed lines*) and our model (*red solid lines*) predict the baryonic properties well, but in general our improved model's PDFs better match the *actual* PDFs in ILLUSTRISTNG — as demonstrated by the residual plots.

Factors that improved our model

Having overviewed our machine's overall accuracy by comparing it with the *actual* data and with the baseline model, we now focus on each of the factors that improved our model. In what follows, we explain each of the three major improvements we made to our MSSM pipeline, followed by how we identify the best combination of these improvements that exhibits the highest accuracy.

[7] To make the PDF in Figure 4.3, we sum up the test results of 5 ($= 1/0.2$) trials of machine learning and testing, where 0.2 is the fractional size of the ILLUSTRISTNG test set. Then, the fractional halo numbers in each bin match the number density in the real Universe. For this reason, the black dashed line in Figure 4.3 is slightly different from that of Figure 4.4, the actual halo number density in the ILLUSTRISTNG volume (TNG100-1).

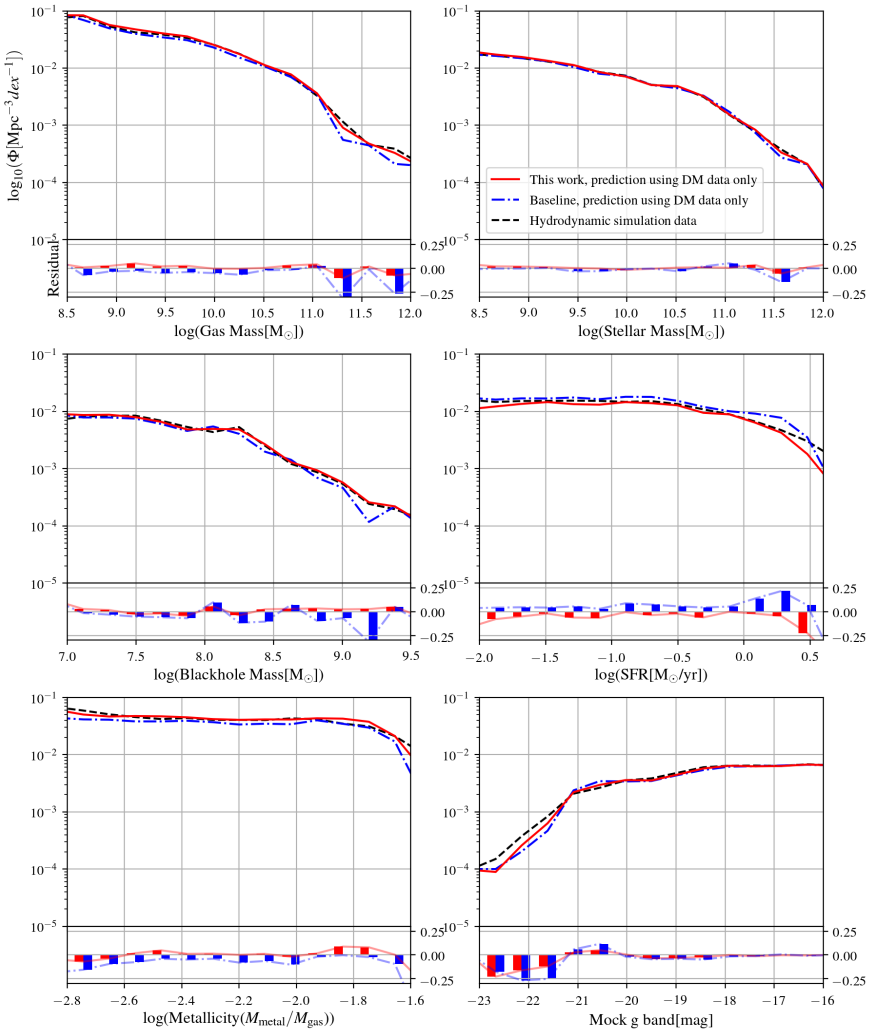

Figure 4.3: Probability distribution functions Φ (PDFs) of six baryonic properties — gas mass, stellar mass, central black hole mass, star formation rate (SFR), metallicity, and stellar magnitude (g band) — predicted from input DM features in the ILLUSTRISTNG test set. The baseline model and our improved model are shown in *blue dot dashed lines* and *red solid lines;*. The residuals between the predicted and the *actual* PDF, $\log \Phi_{\rm pred} - \log \Phi_{\rm TNG}$, are displayed in the bottom chart of each panel.

Using A Refined Error Function with Logarithmic Scaling

One of the most common choices for an error function in the machine learning algorithm — including our choice, ERT — is the MSE (see Section 3.1),

$$\mathrm{MSE}_{\mathrm{node}} = \frac{1}{N_{\mathrm{node}}} \sum_{i \in \mathrm{node}} \left(y^i - y_{\mathrm{node}}\right)^2. \tag{4.6}$$

However, a severe problem may arise when our prediction target property has a large dynamic range (e.g., halo gas masses ranging from $10^8\,\mathrm{M}_\odot$ to $10^{12}\,\mathrm{M}_\odot$). A simple mathematical argument tells that when naively used with raw y values, MSE could be disproportionately more sensitive to larger y values. For example, a small fractional error in the $10^{12}\,\mathrm{M}_\odot$ range may completely dominate over even a very large fractional error in the $10^8\,\mathrm{M}_\odot$ range. This has caused the naive baseline model to perform poorly in the lower value range (see e.g., the left panel of Figure 4.2).

To amend the problem, in the learning phase, we apply *logarithmic scaling* to desired outputs of the training set (i.e., actual baryonic properties in ILLUSTRISTNG — except stellar magnitudes).[5] Or equivalently, the y variables in the MSE error function, Equation 4.6, now mean *logged* outputs, brining y values to the range of $O(1)$. As a result, the equation is no longer heavily biased towards larger y values. Hence, our fix alleviates the inaccuracy in the lower end of the predicted outputs (see e.g., the right panel of Figure 4.2).[8] As seen in the 2nd row of Table 4.3.1 where we assemble the scores by each of the improvements, predictions such as gas mass, SFR, and metallicity benefit from the refined error function (e.g., MBE for gas mass prediction decreased from 0.0015 to 0.0010).[9] On the other hand, pre-

[8] An alternative to the logarithmic scaling could be to normalize the raw y values. However, the normalized variables lose their physical meanings, so the physically meaningful quantities must be carefully recovered afterwards. In contrast, logarithmic scaling does not lead to the loss of physical meaning.

[9] Each of the MBE/MBSD scores in the table is an average over 200 trials. A machine built in each trial is different due to the randomness in building an ERT, and in choosing a training set (80% of the ILLUSTRISTNG data).

dictions for stellar and central black hole masses do not benefit as much from the refined error function alone.

Using historical and environmental factors

As discussed in Section 4.2.3, we extract and add "historical" and "environmental" factors to the input features when we pre-process the data for our MSSM pipeline. The newly added features are extracted (1) from the halo's merger history, and (2) from the halo's local volume, aimed at directly and indirectly capturing the halo's evolution history. The resulting *value-added* dataset includes seven additional input features such as: number of all mergers, number of all major mergers, mass ratio of the last major merger, local density, number of local halos whose masses are greater than 80% of the target halo's mass, etc. (see Section 4.2.3 for details). It improves our model's MBE and MBSD scores when predicting features like stellar mass, central black hole mass, and SFR (see the 3rd row of Table 4.3.1). For other features, including these extra factors is not as effective by itself.

Two-stage learning with stellar magnitudes as an intermediary

Broadly speaking, the accuracy of the ERT machine learning algorithm improves as the number of decision trees or the "size" of each tree increases[10] (Section 3.1).Since the increased tree size requires exponentially more computing resources, we often need to limit the "depth" of a tree, and/or prune the nodes that are not functional. In practice, however, it is difficult to grow a large tree and prune them into an efficient shape.

Here we introduce a scheme that "links" two machines, by using a predicted output from one machine as an input to the next. The "two-stage learning" scheme works as follows. Imagine building a machine to predict

[10]ERT includes various hyper-parameters to optimize: e.g., maximum depth of a tree, minimum samples split, maximum number of nodes, etc. The "depth" of a decision tree refers to the distance from a root node to a farthest leaf node. The "size" of a tree is the number of all nodes.

SFR based only on DM features (e.g., DM mass or velocity dispersion). To increase the machine accuracy the tree must be both deep and large, requiring copious computing resources. A training set may not be informative enough for a machine to establish a meaningful direct mapping between the DM properties and SFR within a practical time limit. Instead, here we first build a machine estimating stellar magnitudes based on DM properties, then use the predicted stellar magnitudes as part of inputs to another machine estimating SFR.[4,11] By supervising what to estimate first (stellar magnitudes) in order to predict the eventual output (SFR), we effectively "guide" the machine to build one combined, large — yet efficient — ERT. Readers should note that we select stellar magnitudes as an "intermediary" because (1) the stellar magnitudes are relatively accurately predicted only from DM features and (2) the stellar magnitudes and SFR are highly correlated in the simulation data.[12] Thus, we argue that in the two-stage machine training, new astrophysical information is provided to the machine by a human supervisor that the stellar magnitudes are a good intermediary between DM properties and SFR. For more discussion on how stellar magnitudes and the two-stage learning can improve the performance of MSSM.

We find that the two-stage learning technique described here is one of the best strategies to construct a large and efficient ERT, and is also arguably the most effective way to improve the MSSM's accuracy. As an example, for the SFR prediction, the two-stage learning scheme improves both MBE and MBSD scores the most when compared with any other improvements (e.g., MBE for SFR prediction decreased from 1.71 to 1.11, and MBSD from 36.10 to 20.15; see the *4th row* of Table 4.3.1).

Combining improvements to construct the best model

[11] To predict the g band, the other seven bands are used to link the machines.

[12] For example, SFR is more strongly correlated with the stellar magnitudes (e.g., g band) than with any other DM features like DM mass. In other words, when predicting SFR, stellar magnitudes' "feature importances" dominate ($> 50\%$; see Figure 4.7) over other DM features'.

Finally, we combine all three improvements discussed above. Rather than using all the improvements at once, we have carefully tested various combinations of improvements per each of baryonic properties. This is because, when combined, one improvement may hurt the other and lead to an unexpected decrease in machine accuracy. The MBE scores for the identified best combinations are shown in the last row of Table 4.3.1. The best combinations identified here have been referred as our "improved model", and are used to produce the figures presented throughout this chapter.

In Table 4.3.1, readers may notice that the score of a best combination is sometimes the same as that of a single improvement. For example, the MBE for a stellar magnitude prediction is 0.0005 for the best combination, but also for the two-stage learning alone. This means that the two-stage learning technique is the most important and dominant factor in improving the accuracy of stellar magnitude prediction.

4.3.2 Predicting baryonic properties in dark matter-only simulations

We now turn to the application phase of our MSSM pipeline (Figure 4.1), and use the machine to estimate baryonic properties for halos in a DM-only N-body simulation data. The improved machine from Section 4.3.1 trained with the ILLUSTRISTNG data in the learning phase, is fed with the MULTIDARK-PLANCK DM-only simulation (MDPL2; see Section 4.2.2).[13] The machine is asked to generate a galaxy catalogue with multiple baryonic properties — gas mass, stellar mass, central black hole mass, SFR, metallicity, and stellar magnitudes — filling the entire MULTIDARK-PLANCK volume of $(1\ h^{-1}\mathrm{Gpc})^3$.

[13]Note that the DM halos in DM-only simulations and hydrodynamic simulations have experienced different physical processes so are inevitably different but the baryonic back-reaction effect is relatively small.

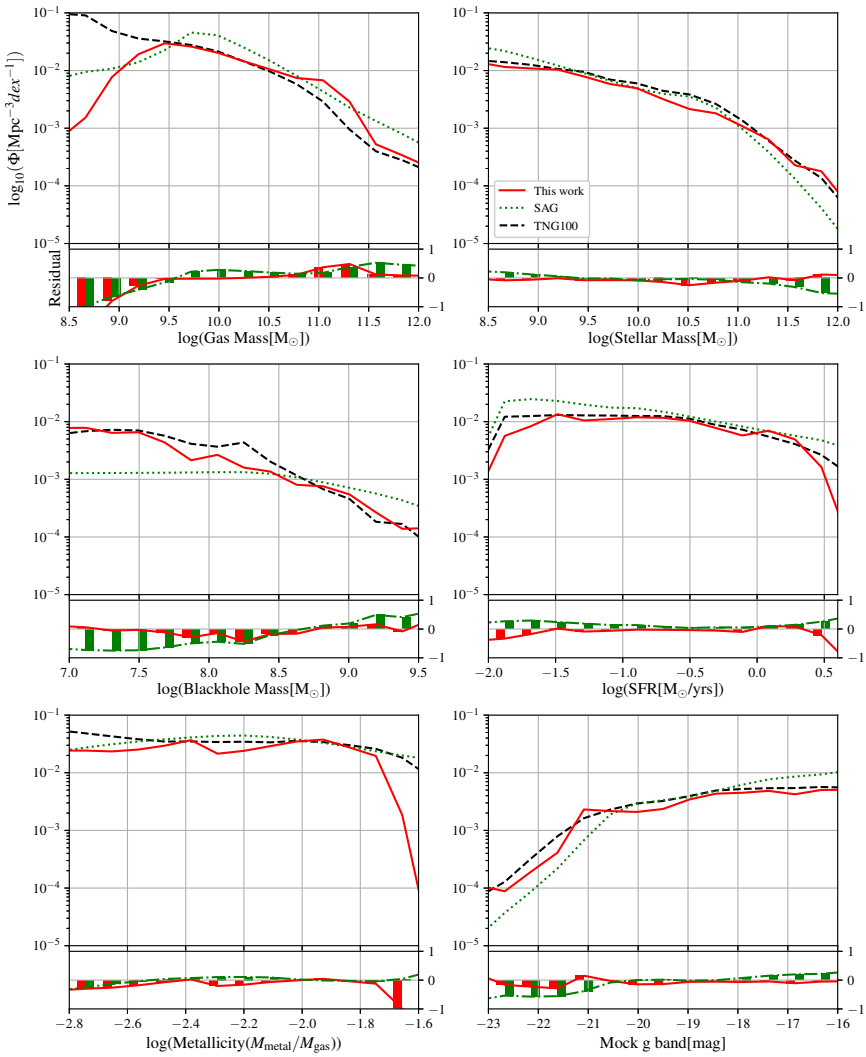

Figure 4.4: Probability distribution functions Φ (PDFs) of six baryonic properties predicted using a DM halo catalogue from the MULTIDARK-PLANCK database. Our improved machine trained with ILLUSTRISTNG is applied to a MULTIDARK-PLANCK dataset to make predictions (*red solid lines*). The residuals between the predicted PDF and the simulation's PDF (ILLUSTRISTNG's), $\log \Phi_{\mathrm{pred}} - \log \Phi_{\mathrm{TNG}}$, are displayed in the bottom chart of each panel.

Is the Machine-assisted Semi-Simulation Model (MSSM) compatible with Semi-Analytic Models (SAMs)?

In Figure 4.4, for six baryonic properties we estimate, we compare the PDFs of our machine learning model (*red solid lines*), and of a SAM (*green dotted lines*). For a representative SAM, we utilize the MDPL2-SAG catalogue (Cora et al. [28]), one of the three SAM-generated galaxy catalogues in the MULTIDARK-GALAXIES database (Knebe et al. [83]).[14] We also add the *actual* baryon data in the ILLUSTRISTNG for comparison (TNG100-1; *black dashed lines*). Overall, we find that our MSSM and the SAM (SAG) exhibit largely compatible distribution functions. For certain properties like black hole masses, star formation rate, and stellar magnitudes, there is a sign that the MSSM mimics the distribution of ILLUSTRISTNG more closely — which is what MSSM is specifically designed to do. Yet, there are some clear mismatches due in large part to the small number statistics. For example, in the gas mass distribution, at $M_{\rm gas} \lesssim 10^{9.5}\,{\rm M}_\odot$, the MSSM's prediction deviates from ILLUSTRISTNG leading to a sizable gap at the lowest mass end (*1st row, left panel*). The MSSM's prediction for metallicity drops drastically at $\log({\rm Metallicity}) \gtrsim -1.8$, too (*3rd row, left panel*).

We then consider the relation between the predicted stellar mass and the halo mass, $M_\star - M_{\rm halo}$, in Figure 4.5. This plot shows how the two halo properties are correlated on a two-dimensional plane (two-dimensional PDF). Since stellar mass is one of the properties the machine can estimate well, our MSSM prediction (*red-blue contours* in the *upper right panel*) replicates the actual $M_\star - M_{\rm halo}$ relation in the ILLUSTRISTNG run well (*top left panel*). As a reference, the prediction of three popular SAMs — SAG (Cora et al. [28]), SAGE (Croton et al. [31]), and GALACTICUS (Benson [12]) — are shown here as *gray contours* demarcating $\rho_{\rm bin,\,cutoff} = 0.01\,/(\log_{10}{\rm M}_\odot)^2$. Also as a reference, added to Figure 4.5 is the result of

[14]The MULTIDARK-GALAXIES data can be found in the COSMOSIM database at http://www.cosmosim.org/.

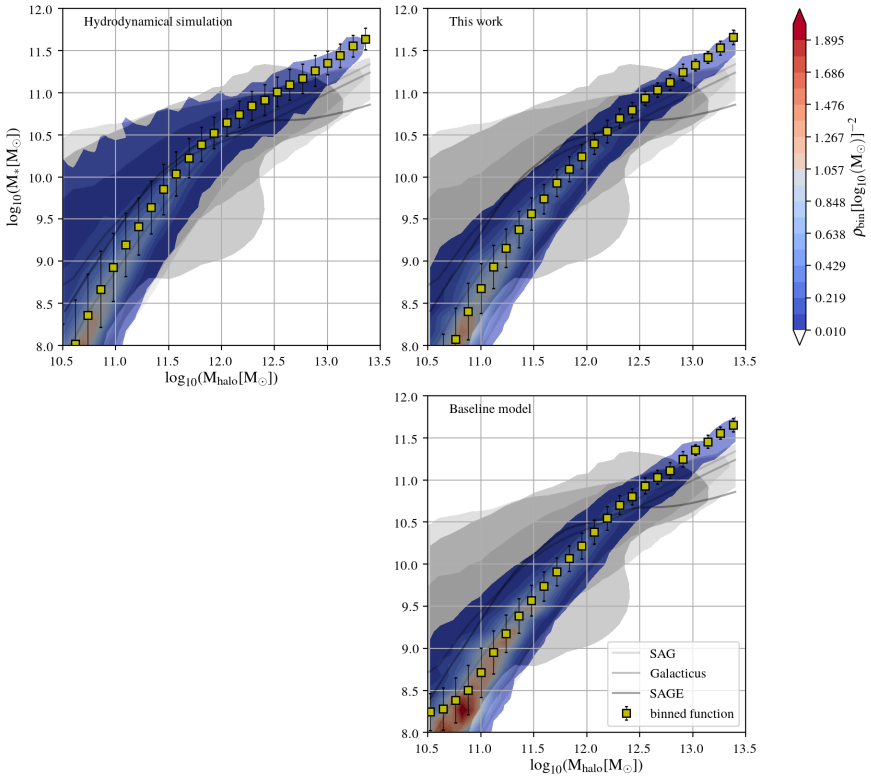

Figure 4.5: Two-dimensional probability distribution of DM halo masses, $M_{\rm halo}$, and predicted stellar masses, $M_\star$ at $z = 0$. Colors indicate $\rho_{\rm bin} = N_{\rm bin}/(N_{\rm tot} S_{\rm bin})$, where $N_{\rm tot}$ is the total number of halos, $N_{\rm bin}$ is the number of halos in each two-dimensional bin, and $S_{\rm bin}$ is the bin area. Yellow squares represent binned averages. The *actual* baryon data in the IllustrisTNG itself is also presented (*top left panel*). Shown in each panel as gray contours are results by three popular SAMs: Sag, Sage, and Galacticus.

the baseline model (*bottom right panel;* see Section 4.2.2 and Table 4.1). Because of the various improvements, our MSSM tends to perform better in the lower-mass end ($M_\star < 10^{9.5}\,{\rm M}_\odot$) than the baseline model does.

As illustrated in Figures 4.4 and 4.5, we find that the MSSM pipeline can be a promising way to transplant the baryon physics of a high-resolution

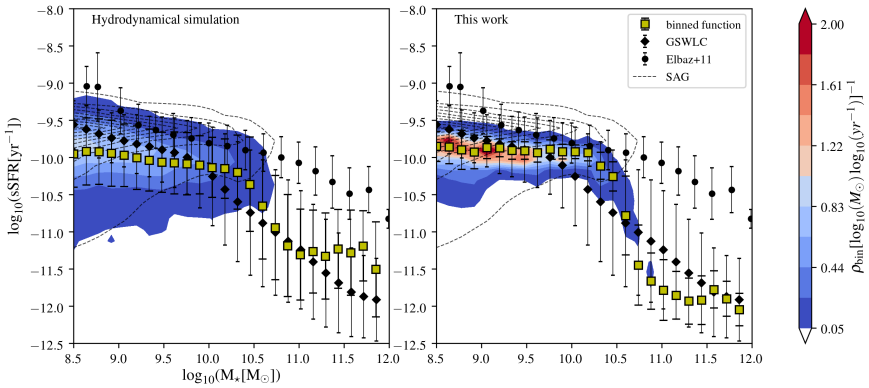

Figure 4.6: Two-dimensional probability distribution of predicted stellar masses, $M_\star$, and predicted specific SFRs at $z = 0$. Colors indicate $\rho_{\rm bin} = N_{\rm bin}/(N_{\rm tot}S_{\rm bin})$, where $N_{\rm tot}$ is the total number of halos, $N_{\rm bin}$ is the number of halos in each two-dimensional bin, and $S_{\rm bin}$ is the bin area. Yellow squares represent binned averages, black diamonds represent *GALEX*-SDSS-*WISE* Legacy Catalog (GSWLC) [153] at $z \sim 0$, and black circles represent a compilation of observations [48] at $z \sim 0$. Our machine trained with ILLUSTRISTNG is applied to a MULTIDARK-PLANCK dataset to predict the PDF (*right panel*). The *actual* baryon data in the ILLUSTRISTNG itself is also presented (*left panel*). Shown in each panel as black dotted contours is the result by a SAM code, SAG.

galaxy-scale hydrodynamic simulation (e.g., ILLUSTRISTNG) onto a larger-volume DM-only simulation (e.g., MULTIDARK-PLANCK). It is also worth noting that our machine can "paint" galaxies and their baryonic properties on a large $(1\ h^{-1}{\rm Gpc})^3$ DM-only run, within a fraction of time required for a high-resolution hydrodynamic calculation — a few tens of minutes (at most) versus a few weeks (at least).

Where the MSSM can be improved

In Figure 4.6, we plot the probability distribution of halos on the plane of predicted stellar masses and predicted specific star formation rates (sSFR). Shown in each panel is the MDPL-SAG catalogue (*black dotted contours;* the outermost contour marks $\rho_{\rm bin,\,cutoff} = 0.05\,/(\log_{10}{\rm M}_\odot \log_{10}{\rm yr}^{-1})$) which

best matches the observational data (*black circles;* [48]) among SAMs; Notice that the ILLUSTRISTNG data itself (*red-blue contours* in the *left panel*) slightly underpredicts the Elbaz's estimation [48] data at a given stellar mass when compared with MDPL-SAG, but better matches the *GALEX*-SDSS-*WISE* Legacy Catalog (*black diamonds*) [153]. The MSSM prediction behaves in a similar way (*red-blue contours* in the *right panel*), which is again exactly what the MSSM is trained to do. However, the two-dimensional distribution of halos is narrower in machine predictions than in the original ILLUSTRISTNG data, as is indicated by the smaller error bars for the binned averages (*yellow squares* in the *right panel*). A similar tendency is found in Figure 4.5 as well, where the halos are distributed in a narrower strip in MSSM predictions but not as much. When only one axis is of a predicted property (e.g., Figure 4.5), the two-dimensional distribution seems broader than when both x- and y-axis are of predicted properties (e.g., Figure 4.6).

The narrower distribution of halos likely implies reduced diversity of galaxies of same stellar masses. We suspect that when the machine is asked to predict baryonic features from DM-related features only, it may have been underfitted due to the inherently limited number of available input features. That is, there are only a very few *important* input features that decides the output, so the diversity of resulting outputs is highly restricted (more discussion in Section 4.4). This is the area where our MSSM pipeline should and can be improved in future studies (see Section 4.5).

4.4 Discussion

In this section, we discuss two topics we find useful to appreciate our MSSM pipeline and its scientific usages.

Relative importance of input features

Since our machine is built with ERT, a RF-type learning algorithm, we can easily find which input feature contributes more than the others (e.g.,

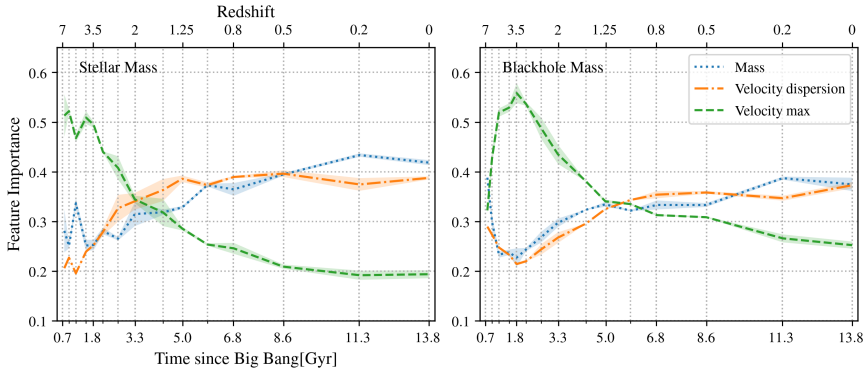

Figure 4.7: Relative importances of input features — halo mass, velocity dispersion, maximum circular velocity — when the machine predicts stellar masses (*left panel*) and central black hole masses (*right panel*) based only on the three DM features of halos in ILLUSTRISTNG. The evolution of the feature importances are plotted as functions of time.

halo mass vs. halo angular momentum) in estimating a particular halo property (e.g., stellar mass). The degree of contribution by each of the input features is termed "feature importance". Feature importance is a relative metric among all input features adopted, and lie in the range of [0, 1]. For example, the feature importances of input parameters P_1, P_2, P_3 could be 0.6, 0.3, 0.1, respectively, which add up to 1.

Figure 4.7 shows how relative importances of input features in the baseline model change over time when predicting two baryonic properties. At high z, the maximum circular velocity is the most responsible in constructing the mappings from inputs to outputs — to both stellar mass (*left panel*) and central black hole mass (*right panel*). However, at lower z, the halo mass and velocity dispersion take over and become more dominant. The trends robustly appear across 15 redshift snapshots from $z = 7$ to 0 we tested, and are highly similar for both mass predictions. At $z = 0$, the halo mass is the most important feature in estimating both properties with features importances $\gtrsim 0.4$.

From feature importances we expect to extract physical insights about how cosmological structures have formed and evolved. We may also use features importances to evaluate how effective a new input feature is compared to preexisting ones. For example, a similar test with our improved MSSM reveals that the three input features shown in Figure 4.7 are still more important than most other newly introduced features in Table 4.1 most of the time. To raise the scientific potential of MSSM, our next goal would be to develop a set of new inputs whose feature importances are comparable to the three existing ones'.

Required training set size to build MSSM

Generally speaking, the size of a training set is one of the deciding factors in the quality of supervised learning. To check whether our TNG100-1 training set is sufficiently large, we measured the machine accuracy with PCC, Equation 4.2, as we increase the size of the training set. In Figure 4.8, we see the effect of the training set size on the accuracy of the baseline model (that uses just three input features — halo mass, velocity dispersion, maximum circular velocity; see Table 4.1). Readers may notice that for all six baryonic properties we estimate, the "learning curves" reach their maximum accuracies with only a surprisingly small number of halos in the training set. For example, for stellar mass and gas mass predictions, $\sim 10^3$ halos are enough to yield reasonably good estimates. For stellar magnitudes (g band) and metallicities, $\sim 10^2$ halos seem sufficient for the machine to reach its maximum potential. From the shapes of learning curves one may argue, for example, that the stellar magnitudes are highly correlated with the three input features (steep ascent to PCC ~ 1 only with $\sim 10^2$ halos), or that SFR is relatively hard to predict no matter how many halos are used in training (steep ascent but only to PCC ~ 0.5).

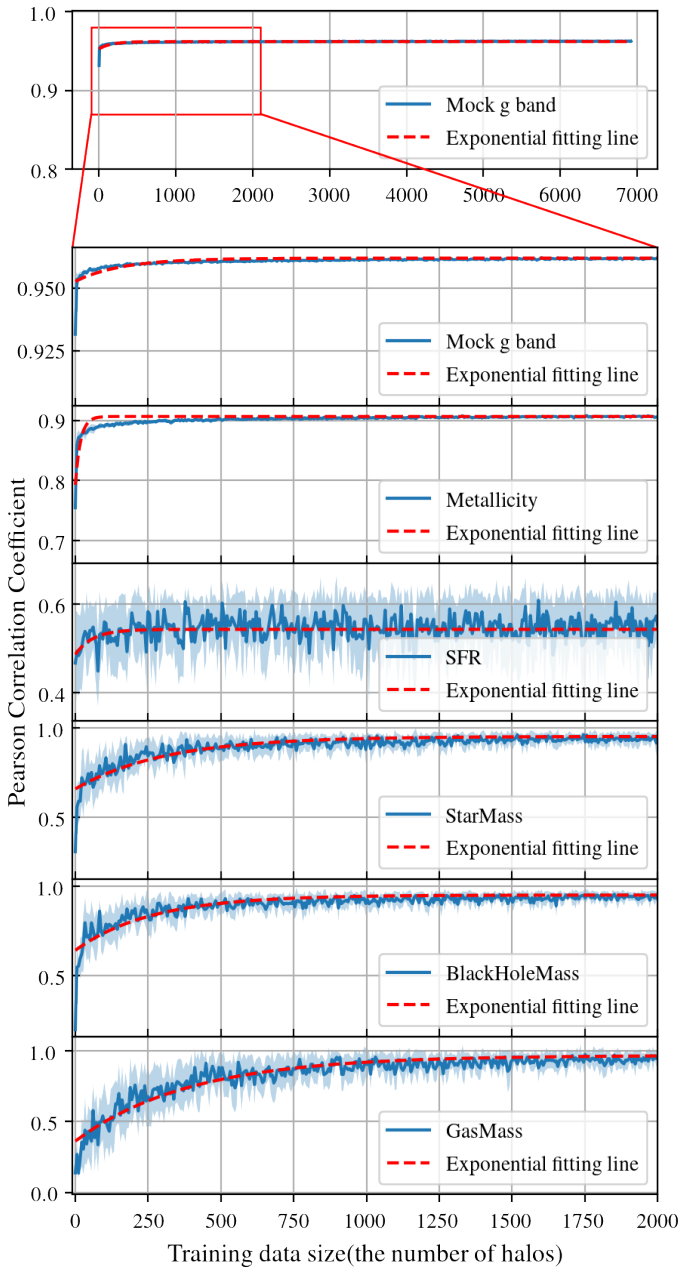

Figure 4.8: Effect of a training set size on the machine accuracy for various baryonic properties (each of *six panels*) based on three DM features of halos.

The baseline model can be well-trained up to its full potential with just $\lesssim 10^3$ halos, at least for the presented machine learning algorithm. Because the $z = 0$ training set from TNG100-1 even after aggressive data pruning (Section 4.2.3) still holds $\sim 4 \times 10^4$ halos, the machine trained with TNG100-1 can be considered to have reached its maximum accuracy.[15] We suspect that if the machine is built with more *important* input features (i.e., not just three features in the baseline model; see Section 4.4), a bigger training set would be needed to converge to the maximum accuracies in the learning curves. Combined with what we see in Sections 4.3.1 and the previous feature importance, our experiments suggest that the machine's accuracy is limited not necessarily by the data size available for training, but more likely by the number of *important* input features.

Effect of Baryons on Dark Matter Halos

In Section 4.3.2, we feed a DM-only simulation data to the machine trained with a hydrodynamic simulation data to generate a galaxy catalogue. For this to work, an implicit assumption is that DM halos from DM-only simulations and the ones from hydrodynamic simulations starting from an identical IC should have an one-to-one match. In hydrodynamic simulation, however, the so-called baryon back-reaction may have an effect on the internal properties of a DM halo such as its shape, profile, and circular velocity (e.g., Duffy et al. 2010; Cui et al. 2012; Martizzi et al. 2012; Sawala et al. 2013; Henson et al. 2017; Chua et al. 2019) and possibly some large-scale properties (e.g., Cui et al. 2017). Internal structure of DM halo can also be affected by sophisticated baryonic physics such as AGN feedback. In this study, however, we consider only the bulk properties of DM halos such as the ones in Table 4.1. For our MSSM to work, one of the crucial

[15] To doubly ensure that our $z = 0$ training set is sufficiently large, we trained a machine with all nine halo catalogues within $z < 0.1$. Using a $\sim$9 times bigger training set did not significantly improve the machine accuracy, as expected by the saturated learning curves in Figure 4.8.

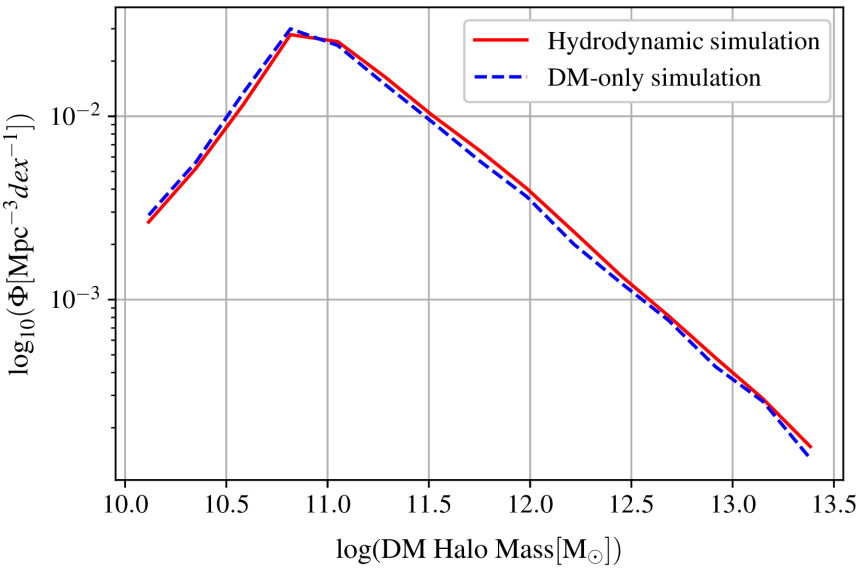

Figure 4.9: Probability distribution functions Φ (PDFs) of DM halo masses for IllustrisTNG (*red solid line*) and IllustrisTNG-Dark simulations (*blue dotted line*). The shift between the two lines is only less than 1%, and can be safely ignored for our purpose when applying our machines.

indicators to inspect would be the DM mass function of halos, not the individual internal structures. Studies have shown that the DM halo mass function of a hydrodynamic simulation including AGN feedback matches well that of a DM-only simulation (e.g., Duffy et al. 2010; Martizzi et al. 2012). Our own comparison of DM halo mass functions from IllustrisTNG and IllustrisTNG-Dark (DM-only run of IllustrisTNG) in Figure 4.9 reveals high resemblance with only a slight shift ($<$1%). For these reasons, we have assumed that DM halos from a DM-only simulation can be used as inputs for a machine trained with a hydrodynamic simulation. Further correction and investigation on this issue remains as future work.

Verifying Stellar Magnitudes As Information Containers

Stellar magnitudes play an important role in the two-stage learning

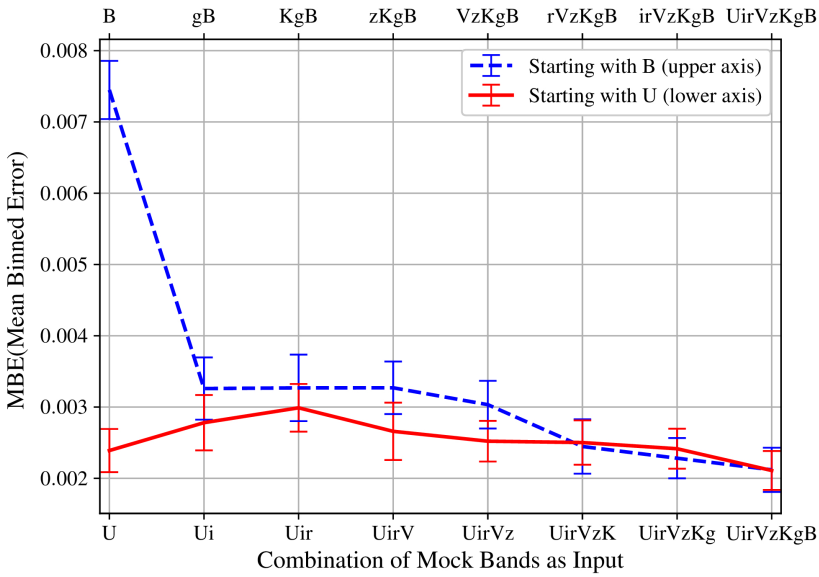

Figure 4.10: Mean binned error (MBE), Eq. (4.4), of stellar mass prediction as a function of how mock band stellar magnitudes are used as an intermediary in the two-stage learning for our MSSM (see Section 4.3.1).[9] Shown on the x-axis are various combinations of mock band magnitudes (e.g., "$zKgB$" means z, K, g, B bands are used as an intermediary in machine training). The *blue dashed line* is for the sequence of combinations shown in the *upper axis*, B to $UirVzKgB$. The *red solid line* is for the sequence of combinations shown in the *lower axis*, U to $UirVzKgB$.

(Section 4.3.1). As discussed, stellar magnitudes are found to be a good intermediary between e.g., DM halo mass and SFR. Typically, star particles in the simulation are convolved with a stellar population synthesis model (e.g., Bruzual & Charlot 2003) and a photometric filter to produce mock band stellar magnitudes. Therefore, one may argue that additional astrophysical information is provided to the machine as we utilize stellar magnitudes as an intermediary.

To better understand how stellar magnitudes and the two-stage learning help our MSSM to achieve better accuracy, here we evaluate if stellar

magnitudes in different bands contain different information. In other words, we check if including more photometric bands improves the MSSM's accuracy. On the x-axis of Figure 4.10, we list combinations of mock band magnitudes used as an intermediary in the two-stage learning when predicting stellar masses. For the *blue dashed line*, we start with just one band, B, and add one more band at a time in the order of g, K, z, V, r, i, U (from left to right on the *upper axis*). This is the ascending order of feature importance among the eight band magnitudes. One can see that as we add more bands, the machine error, MBE, decreases. On the other hand, the *red solid line* is for the reversed order of combinations starting with U (from left to right on the *lower axis*). Since the U band magnitude has the highest feature importance, the MBE is already near its minimum only with the U band. Adding more bands does not significantly improve the machine accuracy.

Our tests reveal that for stellar mass predictions the U band is dominant; for metallicity predictions, the i band is. Because different band magnitudes carry different information about baryonic physics in a galaxy, we expect that including stellar magnitudes in more photometric bands would improve the MSSM's accuracy.

4.5 Summary

Using machine learning techniques, we have developed a pipeline to estimate baryonic properties of a galaxy based purely on DM-related features of its host halo (Section 4.2). Our MSSM pipeline was trained with the ILLUSTRISTNG high-resolution hydrodynamic simulation of a $(75\ h^{-1}\mathrm{Mpc})^3$ volume, so it can establish correlations between DM and baryonic properties (Figure 4.1). Compared to a simpler baseline model similar to prior studies, our machine's accuracy has been significantly improved by several improvements — such as a refined error function with logarithmic scaling in machine training, considering historical and environmental factors of a halo

as inputs, and the two-stage learning with stellar magnitudes as an intermediary. Machine accuracies by each and combinations of these improvements were extensively discussed (Sections 4.2.3 and 4.3.1). For example, the logarithmic scaling in the error function alleviates the inaccuracy in the lower end of the predicted gas masses. The two-stage learning in which predicted stellar magnitudes from one machine is used as an input in the next, is found to be very effective in increasing the prediction accuracy for SFRs.

Once a well-trained machine is in place, in just a few tens of minutes we can rapidly populate a DM-only simulation volume that is large enough to address topics like baryonic acoustic oscillations, with galaxies having basic properties. With our MSSM mimicking ILLUSTRISTNG's galaxy-halo correlation better than previous models, we painted baryonic properties on DM halos in a $(1\ h^{-1}\mathrm{Gpc})^3$ volume of the MULTIDARK-PLANCK DM-only simulation (Section 4.3.2). The resulting MSSM galaxy catalogue is largely compatible with popular SAM catalogues. Furthermore, our MSSM has multiple scientific advantages:

- Within a fraction of time needed for a hydrodynamic simulation, one can efficiently transplant the baryon physics of galaxy-scale hydrodynamic calculations onto a much larger volume. Readers should note that, unlike SAMs, this process does not require any recipes with fine-tuned parameters or human bias.

- The ERT algorithm naturally assesses the relative importances of input features in estimating each baryonic properties (Section 4.4). The feature importance enables us to select important input features easily, and refine the machine with newly added input features with higher importance scores.

- From feature importances, and by comparing the MSSM catalogue with SAMs', one can expect to discover physical insights in structure formation and improve the physics models in SAMs.

Chapter 5

Part II: Calibration of cosmological simulations using simulation-based inference

This chapter is included in the publication submitted to The Astrophysical Journal.

5.1 Introduction

Over the decades, the significant progress of cosmological simulations and observations has greatly improved our understanding of a wide variety of phenomena, such as the formation and evolution of large-scale structure. N-body simulations have successfully simulated formation and evolution of large-scale structure of ΛCDM universe [19, 80, 82, 170]. Furthermore, (magneto-)hydrodynamic simulations that include comprehensive subgrid models such as star formation, stellar winds and active galactic nuclei (AGN) feedback have been performed in a cosmological context and have made significant strides towards reproducing a realistic galaxy population across a range of cosmic epochs. These include ILLUSTRIS [57, 121, 188, 190], cosmo-OWLS [90], MAGNETICUM [65, 144], HORIZON-AGN [44], MASSIVEBLACK-II [77], EAGLE [157], BLUETIDES [51], MUFASA [35], ROMULUS [179], BAHAMAS [108], SIMBA [38], ILLUSTRISTNG [138, 173], HORIZON RUN 5 [91].

Meanwhile, wide-field and deep surveys have identified samples of many thousands of nearby and distant galaxies, respectively [2, 20, 89, 184]. In addition, high-resolution imaging and spectroscopy have enabled investigations into the structure and kinematics of galaxies [40, 192]. These observational breakthroughs have not only enabled access to a plethora of galax-

ies from which to construct global distributions of galaxy properties, such as cosmic star formation history [93, 104] and galaxy stellar-mass functions at different redshifts [8, 94, 111], but also the exploration of many dimensions of galaxy properties, leading to scaling relationships such as the (baryonic) Tully-Fisher relation [109, 110, 182], mass-metallicity relation [53, 55, 98], star-forming sequence [95, 125, 162], size-mass relation [116, 158, 183], and relations between galaxy properties and the mass of the central massive black hole [85, 87, 112].

The remarkable progress of simulations and observations has provided considerable insights into physical processes for galaxy formation and evolution and has played a crucial role in constraining theoretical models. However, simulations and observations have not been reconciled due to several factors, such as various uncertainties in both observation and simulation, limitations of physical models for both observation and simulation, and the absence of delicate and sophisticated comparisons. For instance, cosmological simulations—specifically, the subgrid models—have generally been calibrated against only a handful of observables through by-eye comparisons between simulations and observations, along with educated guesses or simple parameter-space search algorithms [126, 138, 157]. The limits of this conventional calibration approach are as follows: (1) The dimensions of the subgrid parameter space that one can cover is significantly limited when using by-eye comparisons. In addition, as the number of parameters of interest increases, it becomes harder to provide educated guesses due to complex and intertwined relations between physical models and observables. (2) It is challenging to calibrate against numerous observables simultaneously. (3) The accuracy of the calibrated parameters is hard to determine due to the objective nature of the comparison process. (4) The simulation uncertainty, due to such sources as cosmic variance, is generally not taken into consideration. The cosmological simulations suffer from uncertainty that comes from various sources of randomness such as initial conditions [57, 58, 76].

This can lead to appreciable bias in the calibration.

Similarly, semi-analytical models (SAMs), which estimate properties of the galaxy population using parameterized physical models that include a number of free parameters, have rigorously tuned those parameters to reproduce certain observational properties of galaxy population using Bayesian inference together with the Markov chain Monte Carlo method (MCMC) [13, 101, 102]. Bayesian inference is a widely-used method of statistical inference that updates one's knowledge—or belief in the Bayesian sense—of the parameters by making new observations. The majority of the problems of the conventional calibration method for cosmological simulations can be alleviated by the use of Bayesian inference. For instance, the Bayesian inference is conducted through the likelihood functions that can mathematically guarantee the precision of calibration and enables inference from numerous observations simultaneously. Also, the probabilistic nature of the Bayesian inference captures the uncertainty of parameters through Bayes' theorem. However, Bayesian inference usually entails MCMC, which is computationally expensive, for determining the posterior distribution. In contrast to SAMs, hydrodynamical cosmological simulations with subgrid models are too computationally costly to perform hundreds of thousands of simulations for MCMC, rendering such an approach impossible in practice. This is the primary reason, despite all the merits, why calibration of cosmological simulations in the Bayesian framework has never been conducted thus far. In addition, conventional Bayesian inference has its limitations in terms of its need for an explicit (analytic) likelihood. Since the likelihood should be explicitly formulated, commonly-used analytic likelihoods such as Gaussians are only an approximation to the true unknown one.

The simulation-based inference approach (SBI)—also known as likelihood-free inference or simulation-based inference—provides a framework for performing rigorous Bayesian inference in a computationally efficient way, especially for inferences on computationally expensive simulations. In con-

trast to the conventional Bayesian inference that requires an explicit (analytic) formulation for the likelihood, SBI obtains posteriors directly from simulated parameter-observable pairs insert reference to SBI here. Often SBI methods adopt a neural density estimator (NDE) that learns the likelihood—the conditional distribution of observable given the parameters [47, 64, 103, 132, 133]. The NDE removes the need to assume a particular analytical form for the likelihood. On the computational cost side, the likelihood can be evaluated through the trained NDE without performing further simulations. The number of simulations required for the inference is equivalent to the number of simulations for training the NDEs, which is generally thousands of simulations [5]. In conventional inference, although it generally depends on the complexity of the problem, the convergent MCMC typically requires at least 10^5 samples in cosmological applications [52, 181], which is significantly more simulations than the NDE requires. SBI has already started to vigorously be exploited for inference and estimation of physical quantities in astrophysics, for example: inference of the Hubble constant from binary neutron star mergers [59]; constraints on the cosmological parameters from weak lensing [176]; mass estimations of the Milky Way and M31 [96]; inference of strong gravitational lensing parameters [92]; dynamical mass estimation of galaxy clusters [84]; inference of reionization parameters from 21 cm power spectrum and light cones [203, 204];

In this chapter, we adopt the `sbi` package [177], the successor of `delfi` package, which is equipped with various NDEs for SBI, to calibrate cosmological simulations against observations. We also exploit the suite of cosmological simulations of the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project [185] that includes the largest data set designed to train machine learning models and provides more than a thousand simulations for exploring the cosmological and astrophysical parameter space. Despite the large number of simulations in the CAMELS project, it is still not enough to directly employ our SBI tech-

nique, so we build an emulator that is trained on the CAMELS simulations to estimate the target observable taking the cosmological and astrophysical parameters as input. This provides much flexibility on computational cost during inference, which is in line with Elliott et al. [49] that uses an emulator to calibrate a semi-analytic model. Using the emulator as a surrogate to the actual cosmological simulations, we perform SBI using the observed cosmic star formation history [93] and the observed stellar mass functions [94] to infer the parameters that are varied in the CAMELS suite – two cosmological parameters ($\Omega_{\rm m}$ and σ_8) and four astrophysical parameters (related to stellar wind feedback and kinetic black holes feedback).

The structure of this chapter is as follows. In Section 5.2.1, we describe the CAMELS simulations that we use to train emulators and the details of the target observables. In Section 5.2.4, we give a brief review of the SBI including the NDE and our emulator design. In Section 5.3.1, we investigate the performance and convergence of the posterior distributions of the cosmological and astrophysical parameters inferred from an emulated star formation rate density (SFRD) as the target observable. In Section 5.3.2, we perform the inference from the observed SFRD and study how the inferred SFRD matches the observed one. In Section 5.3.3, we investigate the performance and convergence of the posterior distributions from an emulated stellar mass function (SMF) as the target observable and in Section 5.3.4, we perform the inference from observed SMFs and study the discrepancies between those and the inferred ones. In Section 5.4, we discuss the properties of the inferred posteriors and mismatch between inferences and observations with respect to the correlation between the parameter-observable pairs and physical analysis of cosmological simulations. In Section 5.5, we present a summary of the results and findings.

5.2 Methodology

5.2.1 Cosmological simulations: the CAMELS project

Overview

Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS)[1] is a suite of 4,233 cosmological simulations: 2,184 (magneto-)hydrodynamic simulations with the AREPO and GIZMO codes, and 2,049 N-body simulations [185]. Each simulation contains 256^3 dark matter particles of mass $6.49 \times 10^7 (\Omega_m - \Omega_b)/0.251 h^{-1}\mathrm{M}_\odot$ and 256^3 gas cells with an initial mass of $1.27 \times 10^7 h^{-1}\mathrm{M}_\odot$ in a periodic box of comoving volume of $(25 h^{-1}\mathrm{Mpc})^3$, which results in a resolution comparable to but slightly lower than that of ILLUSTRISTNG300-1 [106, 118, 123, 138, 172]. The CAMELS project has been exploring a wide cosmological and astrophysical parameter space for the applications of machine learning in astrophysics. The cosmological and astrophysical parameters of interest are Ω_m, σ_8, A_SN1, A_SN2, A_AGN1, and A_AGN2 (refer to Equations 2.9 to 2.17 for details). The suite of (magneto-)hydrodynamic CAMELS simulations comprises four different sets for each of the AREPO and GIZMO codes, as follows: 1) the LH set consists of 1000 simulations with different initial conditions varying all parameters sampled from a latin hypercube; 2) the 1P set consists of 61 simulations with the same initial condition varying only one parameter at a time; 3) the CV set consists of 27 simulations with fixed cosmology and astrophysics that sample cosmic variance using different initial conditions;

The AREPO and GIZMO simulations are referred to as TNG and SIMBA, respectively. Throughout this chapter, we adopt the TNG suites of the CAMELS simulations unless specified otherwise. The LH set of the TNG suites is exploited to train the emulator (Section 5.2.5). The CV set of the TNG suites is used to model simulation uncertainty. The 1000 sim-

[1] `https://www.camel-simulations.org`

ulations of the LH set are performed with $\Omega_{\rm m} \in [0.1, 0.5]$, $\sigma_8 \in [0.6, 1.0]$, $A_{\rm SN1} \in [0.25, 4.0]$, $A_{\rm SN2} \in [0.5, 2.0]$, $A_{\rm AGN1} \in [0.25, 4.0]$, and $A_{\rm AGN2} \in [0.5, 2.0]$ arranged in a latin hypercube. The 27 simulations of the CV set are performed with $\Omega_{\rm m} = 0.3$, $\sigma_8 = 0.8$, $A_{\rm SN1} = A_{\rm SN2} = A_{\rm AGN1} = A_{\rm AGN2} = 1$ but with different initial conditions. In the meantime, the following cosmological parameters are fixed across all simulations: $\Omega_{\rm b} = 0.049$, $h = 0.6711$, $n_{\rm s} = 0.9624$, $M_\nu = 0.0eV$, $w = -1$ and $\Omega_{\rm K} = 0$. The TNG suite of the CAMELS simulations implements the subgrid physics models of IllustrisTNG [138, 194]. These simulations employ the AREPO code[2] [164, 197] to solve gravity (TreePM) and magneto-hydrodynamics using a Voronoi moving-mesh approach. The IllustrisTNG physics includes various subgrid models: radiative cooling and heating, star-formation, stellar evolution, feedback from galactic winds, the formation and growth of the supermassive black holes (SMBH), and feedback from AGN. The detail implementations of galactic winds and AGN feedback are described in Chapter 2.

Target observables: cosmic star formation rate density & stellar mass function

In this chapter, the target observables from which the cosmological and astrophysical parameters are inferred are the cosmic star formation rate density (SFRD) and the stellar mass functions (SMF). To make an inference from observations, the data structures of simulations must be consistent witht that of observations. To this end, we manipulate the simulation data to fit into observation data. On the observational side, we adopt the SFRD of Leja et al. [93] ranging from $z = 3$ to 0.5 and the five SMFs of Leja et al. [94] at $z = 0.5$, 1, 1.5, 2.0, and 2.5 that are inferred with `Prospector` using galaxies in the 3D-HST and COSMOS-2015 surveys in a way that SFRD and SMFs are consistent as in simulations.

[2] `https://arepo-code.org`

On the simulation side, the SFRD is constructed from the global star formation rate per unit co-moving volume for 21 snapshots (21 redshifts). Here, we match the values of 21 redshifts between simulations and observations. The SMFs are obtained from the stellar mass of the galaxy catalog. Each SMF is binned with 13 bins in the range $[10^{8.9}, 10^{11.4}]\mathrm{M}_{\odot}$.

5.2.2 Resolution effects: convergence & rescaling

The resolution convergence and effects in the TNG simulations have been extensively studied in Weinberger et al. [194, Appendix B], Pillepich et al. [138, Appendix A], and Pillepich et al. [136, Section 3.3]. In general, observables such as the SMF at different resolutions are not fully converged. Figure 5.1 shows that there are appreciable offsets in both SMFs and SFRDs between TNG100-1 and TNG100-2. Note that the fiducial parameters for the suite of the TNG simulations are calibrated against observations for TNG100-1. The CAMELS parameter variations are chosen around the fiducial values of the TNG suite. Hence, based on the approach of Pillepich et al. [136, Appendix A], we re-scale the SFRD and SMFs with a mass-modulated re-scaling factor. Since the resolution of CAMELS simulations are comparable to TNG100-2, we construct a re-scaling factor using the star formation rate-halo mass relation and the stellar mass-halo mass relation from TNG100-1 and TNG100-2. The re-scaled CAMELS simulations in Figure 5.1 demonstrates that the application of the re-scaling reduces the resolution effects at some level. One remark is that we find that the re-scaling depends on the cosmological and astrophysical parameters. However, in this work, we ignore this effect. Further will be discussed in Section 5.4.5.

Details of re-scaling

We adopt the re-scaling method introduced in Pillepich et al. [136, Appendix A]. There, the stellar mass functions (SMF) of TNG300-1 is re-scaled using the ratio between TNG100-2 and TNG100-1. The re-scaling

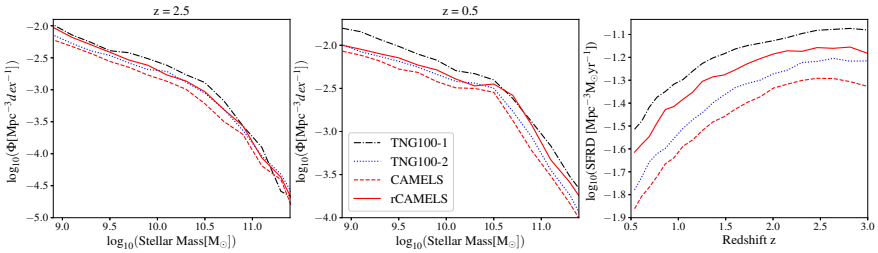

Figure 5.1: *Left* and *Middle*: Stellar mass functions at $z = 2.5$ and 0.5. *Right*: Cosmic star formation rate densities. The difference in TNG100-1 (*black dash-dotted*) and TNG100-2 (*blue dotted*) illustrates the resolution effect of the TNG simulations. The fiducial CAMELS simulation (*red dashed*) is re-scaled with the ratio between TNG100-1 and TNG100-2. The re-sacled CAMELS simulation is shown in *red solid* lines.

is based on the resolutions of TNG100-2 and TNG300-1 being essentially the same, and the SMFs of the two simulations being in a good agreement despite the difference in the simulation volumes. As a result, the re-scaled SMF of TNG300-1 (rTNG300-1) coincides with that of TNG100-1 with high accuracy.

Figure 5.1 shows resolution effects in each observable in comparison between TNG100-1 (*black dash-dotted*) and TNG100-2 (*blue dotted*). The fiducial CAMELS (*red dahsed*), computed as the average of the CV set, is overall slightly lower than TNG100-2 in both SMF and SFRD. This is because the resolution of the CAMELS simulations is lower than TNG100-2. For dark matter particles, the mass and spatial resolution in CAMELS are $\sim 9.67 \times 10^7 \mathrm{M}_\odot/\mathrm{h}$ and $\sim 2\,\mathrm{kpc}$ comoving, whereas TNG100-2 has a mass resolution of $\sim 5.97 \times 10^7 \mathrm{M}_\odot$ and a spatial resolution of $\sim 1.48\,\mathrm{kpc}$ comoving. In this chapter, we ignore the discrepancies between the CAMELS simulations and TNG100-2. We re-scale the SMF by estimating stellar mass

($M_\star$) as a function of each bin of halo mass ($M_{\rm halo}$) as follows:

$$M_{\star,\rm rCAMELS}(M_{\rm halo}) = M_{\star,\rm CAMELS}(M_{\rm halo}) \times \frac{\langle M_{\star,\rm TNG100-1}(M_{\rm halo})\rangle}{\langle M_{\star,\rm TNG100-2}(M_{\rm halo})\rangle}, \quad (5.1)$$

where $\langle\cdot\rangle$ stands for average over all halos in each bin to which $M_{\rm halo}$ belongs. We multiply the stellar mass of each CAMELS halo by the corresponding re-scaling factor (the last fractional term) for that halo mass. The re-scaled SMF of the CAMELS simulations (*red solid* in Figure 5.1) at $z = 2.5$ (*left*) is in a good agreement with TNG100-1 (*black dash-dotted*), whereas the re-scaled SMF at $z = 0.5$ (*center*) deviates appreciably from TNG100-1, especially in the low-mass end. The discrepancy is attributed to the resolution limit that leads to the lower bounds for both dark matter mass ($\sim 10^8 \mathrm{M}_\odot$) and stellar mass ($\sim 10^7 \mathrm{M}_\odot$) of halos. Since the mass-resolution limits of halos cause lack of galaxy population in the vicinity of the limits, the construction of the re-scaling factor becomes unfeasible in the halo-mass range of $[10^8, 10^9]\mathrm{M}_\odot$, which largely affects the low-mass end of SMFs. This results in the significant discrepancy between the SMFs of rescaled CAMELS and TNG100-1 in low-mass end at $z = 0.5$. In this chapter, we have not employed any post-processing for the zero-stellar mass galaxies whose stellar mass is not resolved due to the mass-resolution limit.

Similarly, we re-scale the SFRD by multiplying the SFR of the CAMELS simulations by the re-scaling factor as a function of bin of halo mass as follows:

$$\mathrm{SFR}_{\rm rCAMELS}(M_{\rm halo}) = \mathrm{SFR}_{\rm CAMELS}(M_{\rm halo}) \times \frac{\langle \mathrm{SFR}_{\rm TNG100-1}(M_{\rm halo})\rangle}{\langle \mathrm{SFR}_{\rm TNG100-2}(M_{\rm halo})\rangle}, \quad (5.2)$$

where $\mathrm{d}M_\star/\mathrm{d}t$ is a star formation rate of galaxy. The re-scaled SFRD in Figure 5.1 (*right, red solid*) has a consistent offset from TNG100-1 (*black*

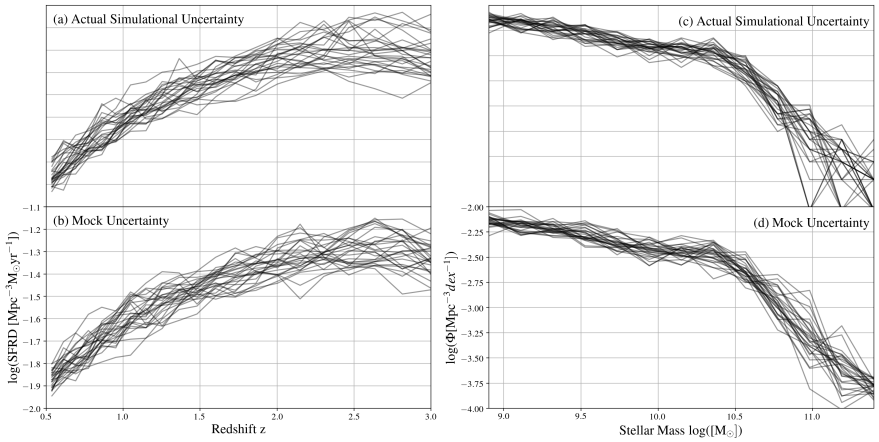

Figure 5.2: (a) Cosmic star formation rate densities from 27 simulations of the CV set; (b) Mean of the 27 cosmic star formation rate densities from the CV set + Mock uncertainty; (c) Stellar mass functions at $z = 0.5$ from 27 simulations of the CV set; (d) Mean of the 27 stellar mass functions at $z = 0.5$ from the CV set + Mock uncertainty Each panel consists of 27 different curves.

dash-dotted), which is simply attributed to the difference between TNG100-2 (*blue dotted*) and the CAMELS (*red dashed*).

5.2.3 Uncertainties in simulation

In this section, we focus on *simulation uncertainty* that is modeled as mock uncertainty and added to emulators in Section 5.2.5 The *simulation uncertainty* is the intrinsic uncertainty of cosmological simulations. In cosmological simulations, randomness in the initial conditions that correspond to the density fluctuations of the early universe leads to *cosmic variance*. The minute position differences of the initial conditions owing to random seeds manifest as differences in the large-scale structure that directs to the galaxy populations. On an observational side, *cosmic variance* can be attributed to the limited volume of the surveys. Meanwhile, the *butterfly effect* stems from the chaos-like behaviors of cosmological simulations. In

dynamical systems, we can quantify chaotic or stochastic behaviors using a quantity called Lyapunov exponent λ. If the positive Lyapunov exponent, the minute perturbations evolve exponentially with Lyapunov timescale λ^{-1} and manifest as macroscopic differences. The chaos-like behavior of the galactic dynamical systems amplifies minute fluctuations, seeded by randomness in the numerical computations such as stochasticity in subgrid models and floating errors, into appreciable differences in later times [58]. Also, Keller et al. [76] studied stochasiticity of galaxy propertices in terms of using the particle-base code GASOLINE and the grid-base code RAMSES.

Here, we quantify the simulation uncertainty using the CV set and model the mock uncertainty. Figure 5.2 illustrates 27 SFRDs (*top left*, (a)) and 27 SMF at $z = 0.5$ (*top right*, (c)) from 27 cosmological simulations in the CV set, which represents the simulation uncertainty (refer to Section 5.2.1 for details of the CV set). We measure the standard deviations of 27 SFRDs and 27 SMFs as $\sigma_{\rm sim,sfr} = 0.057$ dex and $\sigma_{\rm sim,smf} = 0.111$ dex, respectively.

Implementation of the mock uncertainty

We model the mock uncertainty of the SFRD and SMF using a modified six-dimensional Gaussian noise in the form

$$\mathbf{Z}(\boldsymbol{x}) = \mathbf{C}^{\rm T} \frac{\exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\rm T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}{\sqrt{(2\pi)^6|\boldsymbol{\Sigma}|}} \tag{5.3}$$

with $\mathbf{C} = \boldsymbol{I}$ for the SFRD and $[0.3,\ 0.34,\ 0.38,\ 0.43,\ 0.47,\ 0.51,\ 0.56,\ 0.6,$

$0.7,\ 0.9,\ 1.5,\ 1.7,\ 0.5]$ for the SMF, $\boldsymbol{\mu} = \mathbf{0}$, and $\boldsymbol{\Sigma} = \sigma e^{\gamma \boldsymbol{\Gamma}}$ where

$$\boldsymbol{\Gamma} = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 1 & 2 & 3 & 4 \\ 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 4 & 3 & 2 & 1 & 0 & 1 \\ 5 & 4 & 3 & 2 & 1 & 0 \end{bmatrix}.$$

Here, (σ, γ) are $(0.02, 0.2)$ for the SFRD and $(0.02, 2)$ for the SMF. Note that the parameters are empirical. The parameters of the mock uncertainty are tuned such that the mock uncertainty can have a visually similar form and similar standard deviations to the simulation uncertainty. In Figure 5.2, panels (b) and (d) exhibit the 27 SFRDs and 27 SMFs with mock uncertainty. Here, 27 SFRDs and 27 SMFs are generated by adding the mock uncertainty to the mean of 27 SFRDs and 27 SMFs from the CV set, respectively. The mock uncertainty-implemented SFRDs and SMFs are visually in a good agreement with SFRDs and SMFs from the simulation uncertainty. The standard deviations of the mock uncertainty of SFRD and SMF are $\sigma_{\mathrm{mock,sfr}} = 0.061$ dex and $\sigma_{\mathrm{mock,smf}} = 0.096$ dex, which approximates to that of the simulation uncertainty. Note that in principle, the mock uncertainty depends on the cosmological and astrophysical parameters. Nevertheless, we model the mock uncertainty as if the simulation uncertainty is consistent over the entire parameter space since it is computationally impossible to perform simulations across parameter space to obtain the simulation uncertainty as a function of parameters.

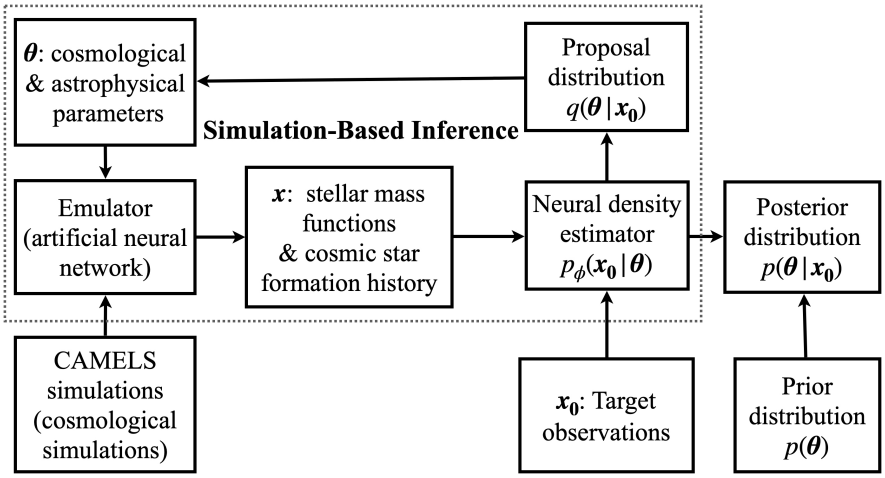

Figure 5.3: Diagram of the pipeline for this chapter.

5.2.4 Simulation-based inference with emulator

Simulation-based inference (SBI) aims at identifying the regions of parameter space of forward models[3] to match observations [29, 177]. In Bayesian terminology, SBI retrieves the posterior distribution over the parameters given observation. In contrast to the conventional Bayesian inference, SBI requires no assumption or ansatz for likelihood $p(\boldsymbol{x}|\boldsymbol{\theta})$ so that it can also be applicable when analytical formulation for the likelihood is not accessible. Here, $\boldsymbol{x}$ and $\boldsymbol{\theta}$ generally stand for observable and relevant parameters, respectively.

In this chapter, we adopt `sbi`[4], a package designed to retrieve the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$ [177]. Instead of analytical probability distributions such as the Gaussian distribution, `sbi` employs a neural network to output a probability distribution, called a neural density estimator (NDE). The NDE is a neural network that takes data points $\boldsymbol{x}$ as input and returns a

[3] A forward model numerically estimates output taking parameters as input, but the corresponding inverse model does not exists.

[4] https://github.com/mackelab/sbi

conditional probability distribution $\tilde{p}_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\boldsymbol{x})$ over the simulation parameters such that $\int \tilde{p}_{\boldsymbol{\phi}}(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{x} = 1$ where $\boldsymbol{\phi}$ represents neural network parameters. Please refer to Figure 5.3 for the pipeline of this chapter.

The `sbi` package also aims at training the NDE with the least number of evaluations of forward models such as cosmological simulations, which are in general computationally expensive, by making an inference concentrating on the relatively small parameter space around the parameters of interest. To this end, of importance is the choice proposal $q(\boldsymbol{\theta})$ from which we draw the sets of parameters for the new simulations in the next iteration. In general, the choice of the optimal proposal for a specific problem is an open question. Since we are interested in the high probable regions of parameter space, it might be natural to adopt the current approximate posterior density as a proposal density [47, 134]. On the other hand, Alsing et al. [4] adopts the geometric mean of prior and the current approximate posterior density in the context of sequential Approximate Bayesian Computation (ABC). It might increase the probability of exploring parameter space more broadly beyond the posterior density. Please refer to Section 3.2 for details of simulation-based inference and the NDE.

5.2.5 Emulator: surrogate for cosmological simulations

The number of simulations required to retrieve the posterior density is highly correlated with dimensions and complexity of a problem. Nevertheless, SBI generally requires more than thousands of simulations [34, 47, 64, 70, and Figures 5.5 and 5.11 in this chapter], which exceeds the total number of the CAMELS simulations. We, thus, circumvent this issue by adopting an emulator as a surrogate simulation. The emulator is constructed upon a fully-connected neural network that is faster than hydrodynamic simulations by several of orders of magnitude. We split the LH set of the CAMELS simulations into *training* (750), *test* (150), and *validation* (150) sets. Two

independent neural *emulators* are trained on the training sets to estimate the SFRD and SMFs, respectively, as a function of six cosmological and astrophysical parameters: ($\Omega_{\rm m}$, σ_8, $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$). We use `Optuna` [3], an automatic hyper-parameter optimization tool, to train and optimize the emulators. The hyper-parameters subject to optimization includes learning rate, weight decay, the number of layers and the number of neurons. During training, both inputs (cosmological and astrophysical parameters) and outputs (the SFRD and the SMFs) are normalized, using linear scaling[5] and z-score[6], respectively. We measure the accuracy of the emulators with the mean square error (MSE) and the Pearson correlation coefficient. The MSE for emulators are 0.0007 dex (SFRD) and 0.0011 dex (SMF). The Pearson coefficient correlations are 0.98 (SFRD) and 0.94 (SMF).

Connection between emulator and cosmological hydrodynamic simulation

The emulator can be the best option for a surrogate simulation with reasonable accuracy in terms of computational cost. However, aside from accuracy, the imperative missing piece in an emulator is the realization of physical uncertainty such as cosmic variance and butterfly effects (refer to Section 5.2.3 for details of simulation uncertainty). The cosmological simulations generate different observable $g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta}))$ with the same set of parameters $\boldsymbol{\theta}$ depending on its initial density fluctuation $\delta(\lambda, \boldsymbol{\theta})$ where λ stands for any types of randomness (e.g., in case of cosmic variance, λ is a random seed for the initial condition).

The *simulation uncertainty* plays a significant role in a probabilistic inference such as SBI, especially in quantifying uncertainty in the inferred

[5] $x' = (x - x_{\rm min})/(x_{\rm max} - x_{\rm min})$ where x' is the normalized input, and $x_{\rm min}$ and $x_{\rm max}$ are the minimum and maximum of the inputs, namely the edge values of the parameter ranges (refer to Section 5.2.1).

[6] $x' = (x - \mu)/\sigma$ where x' is the normalized output and μ and σ are the mean and the standard deviation of the outputs, respectively.

parameters. In case of an emulator without any uncertainties, the parameters $\boldsymbol{\theta}$ and observable $\boldsymbol{x}$ are highly correlated via the emulators: $\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{x}$. Then, the posterior density $p(\boldsymbol{\theta}|\boldsymbol{x})$ can be given in the form of a Dirac delta function $\delta(\boldsymbol{f}(\boldsymbol{\theta}) - \boldsymbol{x})$. However, with uncertainty in observable, no longer can the posterior density be written with a Dirac delta function. The uncertainty degrades the correlations between parameters and observables, propagates to the posterior density, and ends up having a non-negligible amount of uncertainties in the inferred parameters. Thus, implementation of *simulation uncertainty* in the emulators is indispensable for quantifying uncertainty and estimating error in the inferred parameters given observation.

Prior to the implementation of the *simulation uncertainty* in the emulator, we first investigate how the emulator marginalizes the *simulation uncertainty* during training. To this end, we estimate how much the emulators $f(\boldsymbol{\theta})$ deviated from the uncertainty-marginalized ideal simulations $\bar{g}(\boldsymbol{\theta})$ that ideally marginalize *simulation uncertainty*; The deviation $\Delta(\boldsymbol{\theta})$ = $f(\boldsymbol{\theta}) - \bar{g}(\boldsymbol{\theta})$. Although the deviation itself is unobtainable due to the technical limitations, we can approximate the mean and variance of the deviations by taking an average over the ensembles of the thousand simulations in the LH set. Here, two assumptions are made: (1) The mean of the infinite number of simulation ensembles converges to the uncertainty-marginalized ideal simulation: $\bar{g}(\boldsymbol{\theta}) = \langle g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta}))\rangle_{\lambda}$. (2) The *simulation uncertainty* is constant across every possible simulated Universe: $\sigma[g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta}))] = \sigma_{\rm sim}$ where $\sigma_{\rm sim}$ is a constant [7]. The details of the derivations and the assumptions are further described in Appendix 5.6.2.

As mentioned above, we use two metrics: the bias $b_{\rm LH} = \langle \Delta(\boldsymbol{\theta})\rangle_{\boldsymbol{\theta}\in \rm LH\,set}$ and the variance $\hat{\sigma}_{\rm LH} = \left\langle \Delta(\boldsymbol{\theta})^2 \right\rangle_{\boldsymbol{\theta}\in \rm LH\,set}$. The bias $b_{\rm LH}$ measures the mean of the deviations of emulated observable from the unceratinty-marginalized ideal observable over the LH set that the emulators are trained on. The vari-

[7] $\sigma[g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta}))] \equiv \left\langle (g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta})) - \bar{g}(\boldsymbol{\theta}))^2 \right\rangle_{\lambda}$
where $\bar{g}(\boldsymbol{\theta}) = \langle g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta}))\rangle_{\lambda}$.

ance $\hat{\sigma}_{\rm LH}$ measures the variance of the deviations of emulated observables from uncertainty-marginalized ideal observables over the LH set. If $b_{\rm LH} = 0$ and $\hat{\sigma}_{\rm LH} = 0$, the emulator perfectly marginalizes the *simulation uncertainty* and emulated prediction follows the uncertainty-marginalized ideal simulation. In this work, the emulators have the bias of 0.003 and the variance of 0.066 over the LH set whereas the variance of the *simulation uncertainty* that is estimated from the simulations of the CV set is 0.057 (refer to Appendix 5.6.2). Provided that the variances from emulators and simulations are comparable, it is less likely for the emulators to marginalize the *simulation uncertainty*. Each emulated universe rather corresponds to one of the realizations of the cosmological simulations with unknown random seeds. Without proper marginalization, the implementation of uncertainty in the emulator leads to a greater uncertainty in the parameters as well as the observables. In addition, this weakens the connection between simulations and emulators to a large extent in terms of physical interpretation of the parameters. Hence, in this work, we treat an emulator as ground truth or the mean of simulation uncertainty and implement a mock *simulation uncertainty* on top of the emulators. We model the *simulation uncertainty* using the multivariate Gaussian noise with minor modifications. Then, the *mock simulation uncertainty*—hereafter *mock uncertainty*—is added to the emulators manually (refer to Section 5.2.3 for details of implementation).

5.3 Results

5.3.1 Inference from emulated star formation rate density

In this section, we study the performance and properties of SBI on the cosmic star formation rate density (SFRD) using an *emulated* SFRD rather than an observed SFRD. To this end, we use the emulator to predict SFRDs as function of the cosmological and astrophysical parameters and adopt an

emulated SFRD as a target observation from which the cosmological and astrophysical parameters can be inferred. As discussed in Section 5.2.5, the emulators are adopted as surrogates for the hydrodynamic simulations and considered to be the ground truth throughout this chapter unless it is specified otherwise.

In contrast to deterministic approaches, probabilistic inference such as Bayesian inference and SBI have a significant flexibility in that the inferred posterior distributions can take on versatile structures of probability distributions depending on the nature of the problems. The probabilistic distributions have three main beneficial aspects in our application: (i) the variance of the posteriors can be interpreted as the error bars of inferred parameters; (ii) the posterior can have a appreciable volume of parameter space that reproduces the same target observable within some accuracy; (iii) we can measure variances and uncertainties of each parameter under the presence of the uncertainty of the observations. Lastly, in Section 5.3.2, we apply our SBI machinery to a measurement of the observed SFRD and study how well the inferred posterior distribution can match the observations.

Performance of inference

We first investigate the accuracy of our ILI and how stably the inferred posterior density converges without the mock uncertainty. The performance of ILI with the mock uncertainty will be discussed later this section. Using the suite of CAMELS simulations, we train an emulator that takes as input six cosmological and astrophysical parameters $\boldsymbol{\theta}$ and predicts the cosmic star formation rate density (SFRD) $\boldsymbol{x}$ (for details, refer to Section 5.2.5). Then, we perform SBI to retrieve the posterior density over six cosmological and astrophysical parameters $\boldsymbol{\theta}$ given an *emulated* target SFRD $\boldsymbol{x}_0$. Here, the *emulated* target SFRD $\boldsymbol{x}_0$ is generated by the emulator with $\boldsymbol{\theta}_0$ that is one of the data points in the LH set.

Shown in Figure 5.4 are two-dimensional projections of the inferred

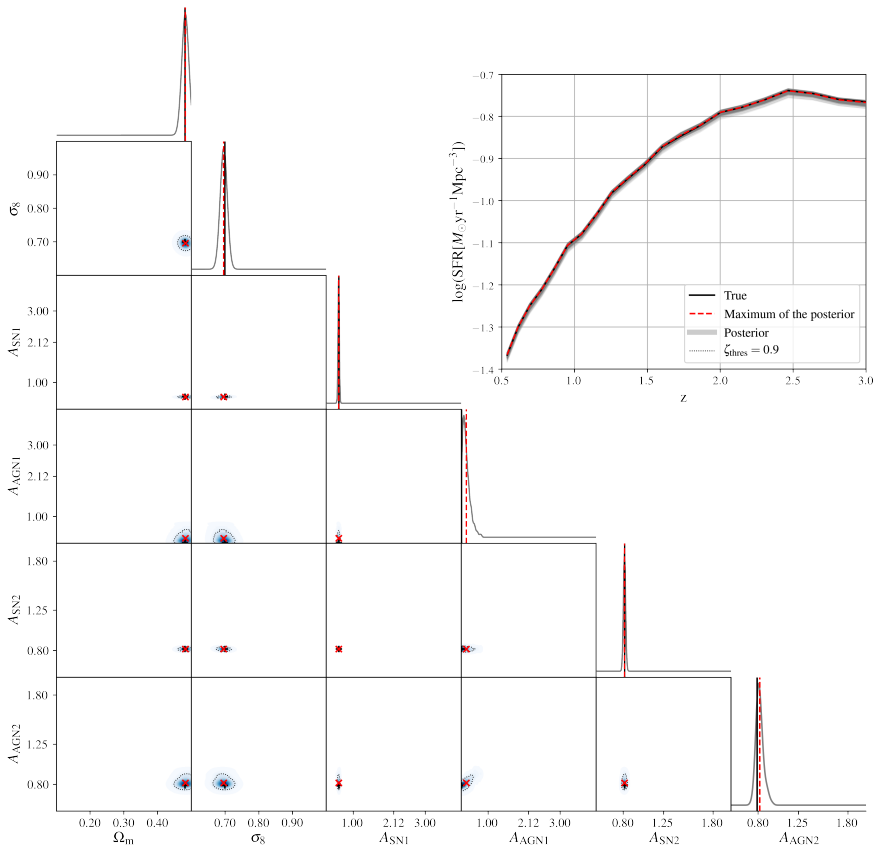

Figure 5.4: *Top right*: *Emulated* cosmic star formation rate densities (SFRDs) from the inferred posterior (*grey*), the maximum of the posterior (*red dashed*), and the target SFRD (*black solid*). *Bottom left*: Two-dimensional distribution of the inferred posterior. The *black* and *red* crosshairs represent the values of target and maximum of the posterior, respectively. The marginal distributions are obtained by kernel density estimation. The *black solid* and *red dashed* vertical lines indicate the true values and the maximum values of the inferred posterior, respectively.

posterior (*bottom left*) and the SFRD plot (*top right*) that includes the corresponding SFRDs drawn from the posterior as well as the target SFRD. In this example, a total of 40000 emulations (400 epochs) is used to retrieve the posterior density $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$ (*grey*) given the SFRD $\boldsymbol{x}_0$ (*black solid*). The loca-

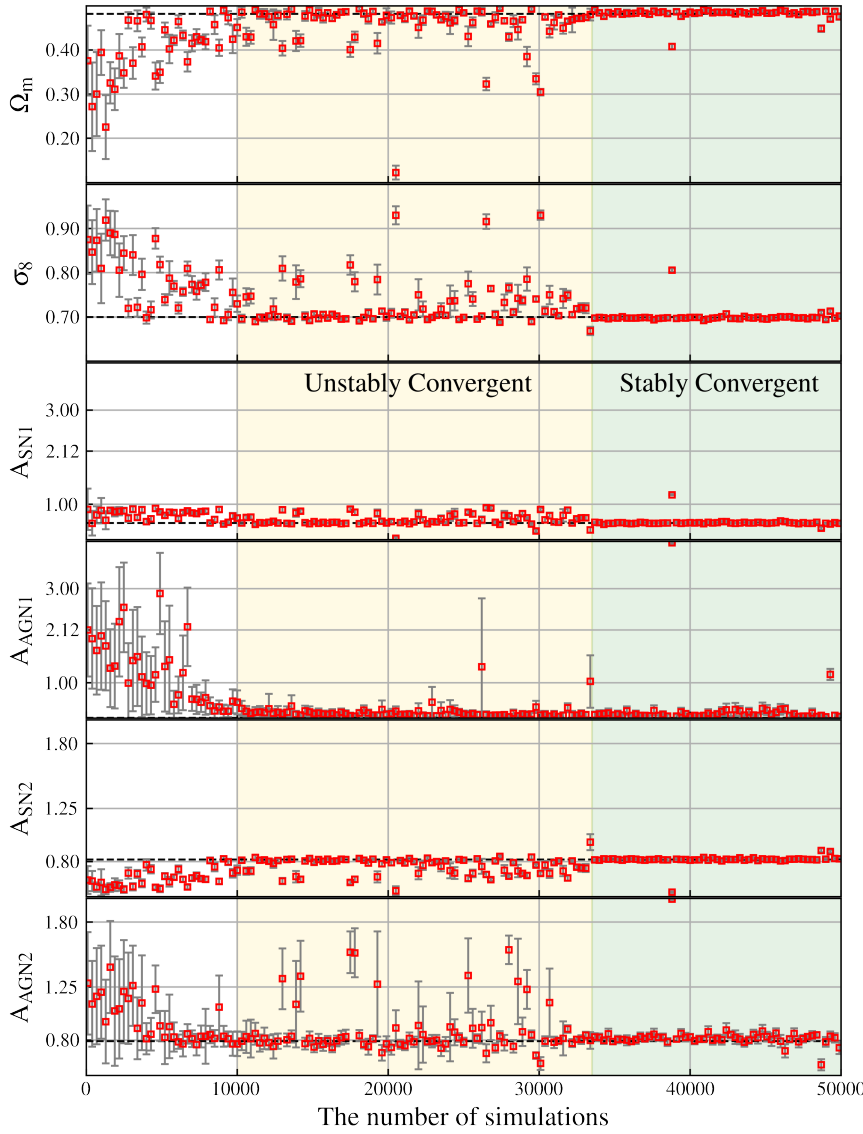

Figure 5.5: The convergence of each parameter as a function of the number of simulations used in the SBI on an emulated SFRD. The *red squares* and *grey error bars* present the maximum and the standard deviation of the posterior density. The *black dotted* lines show the true values.

tion of the maximum of the posterior density[8] (*red dashed* or *red cross-hair*) matches the true values (*black solid* or *black cross-hair*) almost perfectly for all six parameters with the relative errors $\boldsymbol{\delta}_{(\Omega_\mathrm{m},\sigma_8,A_\mathrm{SN1},A_\mathrm{AGN1},A_\mathrm{SN2},A_\mathrm{AGN2})}$= (0.20, 0.49, 1.36, 49.8, 0.25, 3.65)%. We estimate the standard deviations $\boldsymbol{\sigma}$ and the coefficients of variation $\mathbf{c}_\mathrm{v}$[9] of each parameter and find: $\boldsymbol{\sigma}_{(\Omega_\mathrm{m},\sigma_8,A_\mathrm{SN1},A_\mathrm{AGN1},A_\mathrm{SN2},A_\mathrm{AGN2})}$=(0.004, 0.003, 0.014, 0.105, 0.007, 0.040) and $\mathbf{c}_{\mathrm{v}(\Omega_\mathrm{m},\sigma_8,A_\mathrm{SN1},A_\mathrm{AGN1},A_\mathrm{SN2},A_\mathrm{AGN2})}$ = (0.79, 0.42, 2.33, 40.5, 0.76, 5.01)%. The values of the standard deviations imply that the inferred posterior density has very small variances, in that compared to the volume of parameter space covered by the prior of $\mathcal{O}(10)$, the volume of the posterior density approximates to $\mathcal{O}(10^{-12})$. In the *top right* panel, the SFRD from the maximum of the posterior (*red dashed*) coincides with the true SFRD (*black line*) with a relative error of 0.17%. The SFRDs from the full posterior (*grey region*) have an exceedingly narrow distribution with the standard deviation of 0.003 dex.

Figure 5.5 illustrates the convergence of the SBI on the SFRD and its stability. Each panel shows maxima of the posterior (*red square*) and standard deviations (*grey errorbar*) of each parameter as a function of the (cumulative) number of simulations used for training so far (i.e. the number of epochs $\times$ the number of new simulations per epoch). The panel for A_SN1 (*third row*) shows that the A_SN1 parameter converges to the truth almost right after the beginning. The rest of the parameters (Ω_m, σ_8, A_AGN1, A_SN2, A_AGN2) can seemingly come to convergence after 10000 simulations. However, the convergence can be divided into two different stages: the unstably convergent stage in $[10000, 33500]$ simulations (*yellow region*) and the stably convergent stage in $[33500, 50000]$ simulations (*green region*). In the unstably convergent stage (*yellow region*), the inferred parameters jump

[8]The set of parameters, giving the maximum value of the posterior density, is drawn from the MCMC samples. i.e. $\boldsymbol{\theta}_\mathrm{max} =_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\boldsymbol{x}_0)$.

[9]The coefficient of variation, also known as relative standard deviation, is a standardized measure of dispersion of a probability distribution or frequency distribution. It is usually defined as $c_\mathrm{v} = \frac{\sigma}{\mu} \times 100$, but in this chapter we adopt a definition using the maximum of posterior θ_max instead of the mean μ, namely $c_\mathrm{v} = \frac{\sigma}{\theta_\mathrm{max}} \times 100$.

around occasionally without a particular period (unstably) but rapidly return to the truth in the next epoch (convergent). The average relative errors[10], standard deviations, and the coefficients of variation over $[10000, 33500]$ simulations are as follows (also see Table 5.1): $\bar{\boldsymbol{\delta}}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ $= (7.9 \pm 12.5,\ 6.1 \pm 8.8,\ 18.8 \pm 19.4,\ 52.1 \pm 71.6,\ 8.5 \pm 8.0,\ 17.8 \pm 30.4)\%$, $\boldsymbol{\sigma}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})} = (0.016,\ 0.014,\ 0.043,\ 0.138,\ 0.026,\ 0.117)$, and $\mathbf{c}_{{\rm v}(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ =(3.2, 2.0, 7.1, 53.0, 3.2, 14.8)%, respectively. All three measures above are an order of magnitude greater than that of the converged posterior in Figure 5.4. However, the mean of the relative errors of all the SFRDs including not only the maximum of posteriors but also the posteriors themselves, at the unstably convergent stage, $\bar{\boldsymbol{\delta}}_{\rm sfrd}$ approximates to 0.95%, which is clearly greater than that of one single convergent SFRD (0.17%) but in terms of the value itself, it seems acceptable. The most interesting feature of the stage is that although the inferred parameters are jumping around, the corresponding SFRDs are relatively well converging to the truth. This is attributed to the weak correlation between the parameters and the SFRD (see Section 5.4.1 for a discussion). More importantly, this implicitly indicates the possibility of a multi-modal distribution that can reproduce the same observable from different sets of the parameters, as further discussed in the following section.

On the other hand, the posterior distributions are stably convergent with relatively small variances after $\sim$ 30000 simulations (*green region*). All six maxima (*red square*) for the six parameters stably converge to the truth (*black dotted line*) with the average relative errors[10] and their standard deviations over $[33500, 50000]$ simulations (see Table 5.1): $\bar{\boldsymbol{\delta}}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},}$ $_{A_{\rm SN2},A_{\rm AGN2})} = (0.9\pm0.9,\ 1.0\pm3.7,\ 6.3\pm29.8,\ 52.1\pm95.7,\ 1.6\pm5.5,\ 8.2\pm 23.2)\%$. The variances of the posterior density (*grey errorbar*) are conver-

[10]The average relative error of a parameter θ is defined as $\bar{\delta}_\theta \equiv \frac{1}{N}\sum_n \left(\theta_{\rm truth} -_\theta p^{(n)}(\theta|\boldsymbol{x}_0)\right)/\theta_{\rm truth} \times 100$ where n stands for the number of simulations (epoch) in Figure 5.5 and $N = \sum_n 1$. $\boldsymbol{x}_0$ is the target observation.

Table 5.1: The average relative errors $\bar{\boldsymbol{\delta}}$ (*first* and *second rows*), the standard deviations $\boldsymbol{\sigma}$ (*third* and *fourth rows*), and the coefficients of variation $\mathbf{c}_{\rm v}$ (*fifth* and *sixth rows*) of the unstably convergent stage (*U.C.*) and the stably convergent stage (*S.C.*), respectively.

		$\Omega_{\rm m}$	σ_8	$A_{\rm SN1}$	$A_{\rm AGN1}$	$A_{\rm SN2}$	$A_{\rm AGN2}$
$\bar{\boldsymbol{\delta}}$	*U.C.*	7.9%	6.1%	18.8%	52.1%	8.5%	17.8%
	S.C.	0.9%	1.0%	6.3%	52.1%	1.6%	8.2%
$\boldsymbol{\sigma}$	*U.C.*	0.016	0.014	0.043	0.138	0.026	0.117
	S.C.	0.004	0.003	0.015	0.072	0.009	0.036
$\mathbf{c}_{\rm v}$	*U.C.*	3.2%	2.0%	7.1%	53.0%	3.2%	14.8%
	S.C.	0.8%	0.4%	2.5%	27.5%	1.1%	4.6%

gent as well, with average standard deviations of $\boldsymbol{\sigma}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})} = (0.004,\ 0.003,\ 0.015,\ 0.072,\ 0.009,\ 0.036)$ and average coefficients of variation of $\mathbf{c}_{{\rm v}(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ =(0.8, 0.4, 2.5, 27.5, 1.1, 4.6)% over [33500, 50000] simulations. These values are an order of magnitude lower than the values at the unstably convergent stage, which can be a clear sign of transition. Also, in comparison to the values for the posterior in Figure 5.4, the average relative errors of stably convergent posteriors are somewhat greater but the standard deviations and coefficients of variation are comparable. That is, the peaks or the maxima of posteriors have appreciable scatters compared to the truth at the stably convergent stage, whereas the widths of the posteriors are consistent. As in the variances, the convergence properties depend not only on the number of simulations (epoch) but also differ by parameter. In the convergence of $A_{\rm AGN1}$, there is no evident transition from unstably convergent to stably convergent in terms of the relative errors, standard deviations, or the coefficients of variation. On the other hand, the stellar feedback parameters ($A_{\rm SN1}$ and $A_{\rm SN2}$) converges rapidly as soon as the training begins, leading to seamlessly smooth transition to the unstably convergent stage. However, the transition to the stably convergent stage takes place drastically both visually and quantitatively. In

the meantime, the AGN feedback parameters ($A_{\rm AGN1}$ and $A_{\rm AGN2}$) have notable scatter even in the stably convergent stage (*green region*) compared to the other four parameters. Hence, we can conclude that the strengths of correlation between each parameter and the SFRD are in the following order: $A_{\rm SN1} \simeq A_{\rm SN2} > \Omega_{\rm m} \simeq \sigma_8 > A_{\rm AGN1} \simeq A_{\rm AGN2}$. An indication for such correlations can be also found in Villaescusa-Navarro et al. [185, Figure 11 and Figure 12].

Thus far, we have not implemented any uncertainties in the inferences. Therefore, assuming that the emulators are injective (one-to-one function), a negligible variance is expected for all six parameters; i.e. $p(\boldsymbol{x}, \boldsymbol{\theta}) = \delta(\boldsymbol{x} - f(\boldsymbol{\theta}))$ and $\boldsymbol{\theta}$ is unique. However, a tiny amount of variance exists in both the posterior and the inferred SFRDs (*grey* in Figure 5.4). This can be attributed to (1) physical degeneracy and (2) inaccuracy of the NDE. A certain amount of errors of the NDE or machine learning in general is inevitable. In the SBI, the error of the NDE propagates to the posterior density proportionally to the degree of correlations between each parameter and observable. In both Figures 5.4 and 5.5, $\Omega_{\rm m}$, σ_8, $A_{\rm SN1}$, and $A_{\rm SN2}$ show high convergence and precision for both maxima and variances whereas $A_{\rm AGN1}$ and $A_{\rm AGN2}$ have larger variances on average. The magnitude of the variance indicates how strongly the observable can constrain each parameters, or how intimately each parameter and observable correlate with each other. The greater the variance, the weaker the correlation is. Due to the weak correlations, $A_{\rm AGN1}$ and $A_{\rm AGN2}$ require more simulations to converge stably and tend to have a larger variance than other parameters. Nevertheless, the relative errors of both parameters and SFRDs are less than 1% on average with a total of 34000 simulations.

Bimodality in the posterior distribution

Here, we present one of the bimodal posterior distributions that can be found at the unstably convergent stage (*yellow region*) in Figure 5.5. Figure

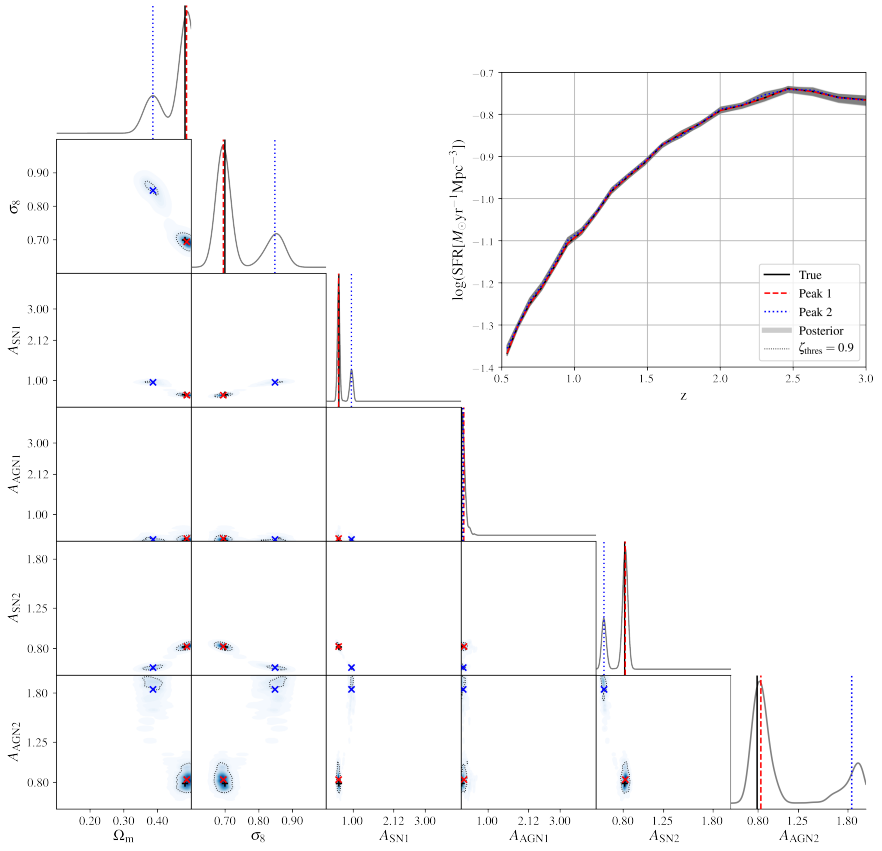

Figure 5.6: *Top right*: *Emulated* star formation rate densities (SFRD) from the inferred posterior (*grey*), two peaks of the posterior (*red dashed* and *blue dotted*), and the target SFRD (*black solid*). *Bottom left*: 2D contour projections of the inferred posterior.

5.6 illustrates two-dimensional projections of the inferred posterior (*bottom left*) as well as the corresponding SFRDs from the posterior and the target SFRD (*top right*). One of the most intriguing feature in Figure 5.6 is clearly the *bimodality* (*red* and *blue*) in the posterior density. Not only do the two peaks exist in the posterior density, but also they reproduce the target observation within a barely appreciable margin of error.

We use k-means clustering to divide the posterior density into two uni-

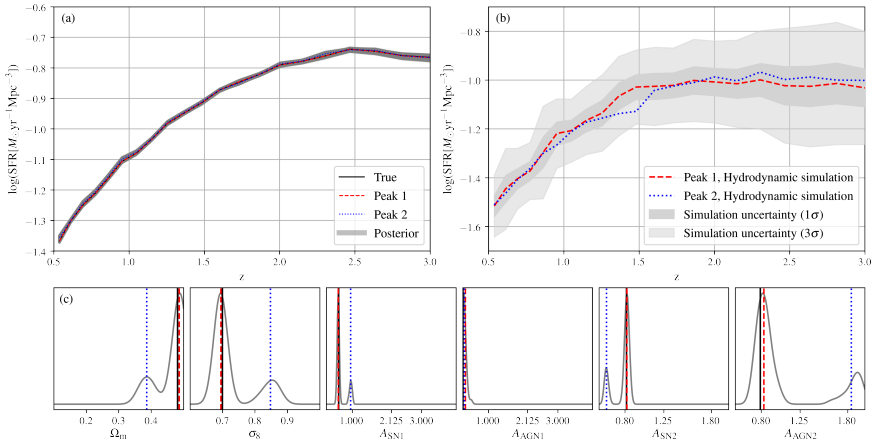

Figure 5.7: (a) Cosmic star formation rate density from the emulator, representing the posterior and its two peaks; (b) Cosmic star formation rate density from cosmological simulations using the points in parameter space that are the peaks of the posterior; (c) One-dimensional projections of the inferred posterior (*grey*) based on Figure 5.6. The two peaks are drawn with the *red dashed* and *blue dotted* lines consistently across Figure 5.6 and panels (c). However, the marginals of the *degenerate* posterior distribution in panel (c) are reconstructed with a modified probability density, namely that which is limited to only the region within $\zeta_{\rm degen} = 0.9$ in Figure 5.6. The *grey* and *light grey* regions in panel (b) indicate the σ and 3σ confidence regions of the simulation uncertainty.

modal distributions. The k-means clustering partitions samples into k clusters in which each sample belongs to the cluster with the nearest mean. Each peak is defined by the center of each cluster. The relative errors of the two SFRDs with respect to the true SFRD are 0.35% for the peak 1 and 0.98% for peak 2. The peak 2 has slightly larger errors than peak 1 but still less than 1%. Furthermore, peak 1 (*red dashed*) accurately coincides with the true (*black solid*) with the relative errors $\boldsymbol{\delta}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ $= (0.45, 0.36, 0.55, 30.8, 0.0041, 5.7)\%$. We measure the variances with respect to each peak by using the result of the k-means clustering. The standard deviation $\boldsymbol{\sigma}$ and coefficients of variation $\mathbf{c}_{\rm v}$ of the peak 1-cluster with

respect to peak 1 are (0.015, 0.014, 0.042, 0.090, 0.032, 0.087) and (3.1, 2.0, 7.0, 26.5, 3.9, 10.5)%. The standard deviation $\boldsymbol{\sigma}$ and coefficients of variation $\mathbf{c}_{\mathrm{v}}$ of the peak 2-cluster with respect to peak 2 are (0.018, 0.025, 0.046, 0.039, 0.020, 0.148) and (4.8, 3.0, 4.9, 13.1, 3.4, 8.1)%. There is no marked difference between the variances of peak 1 and peak 2. The level of deviation is slightly higher than that of the stably convergent posterior in Figure 5.4 but similar to the unstably convergent ones shown in Figure 5.5.

Having two strong peaks in the posterior distribution is the result of a physical *degeneracy*, namely a situation where more than a single set of parameters reproduces the same observable. To study degeneracy, we propose a definition of degeneracy in a mathematically consistent way using a given posterior distribution. The set of degenerate points in parameter space, Θ_{degen}, is defined such that it satisfies $\int_{\boldsymbol{\theta}\in\Theta_{\mathrm{degen}}} p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} = \zeta_{\mathrm{thres}}$ where $p(\boldsymbol{\theta} \in \Theta_{\mathrm{degen}}) > p(\boldsymbol{\theta} \notin \Theta_{\mathrm{degen}})$, assuming that the inferred posterior distribution $p(\boldsymbol{\theta})$ is normalized. Here, ζ_{thres} is a free parameter, and the degenerate-parameter set Θ_{degen} collects parameters according to their probability density in a descending order until the integration of the probability over the degenerate set becomes equal to ζ_{degen} (refer to Section 5.6.1 for a precise and detailed definition). In this chapter, we decide to set ζ_{degen} to 0.9. Also, every two-dimensional projection of posterior distributions in this paper includes a contour line for $\zeta_{\mathrm{degen}} = 0.9$.

Shown in the panel (c) of Figure 5.7 are the marginals of the *degenerate* posterior distribution that is reconstructed with the probability density only enclosed within $\zeta_{\mathrm{degen}} = 0.9$ through the Gaussian kernel density estimation. On the other hand, two peaks (*red dashed* and *blue dotted*) are identical to the ones in Figure 5.6. Figure 5.6 illustrates details of the posterior distribution with $\zeta_{\mathrm{thres}} = 0.9$ (*black dotted contour*) in the two-dimensional projection plot (*bottom left*). The two emulated SFRDs that correspond to two peaks of the posterior distribution (*red dashed* and *blue dotted*) in panel (a) are approximately on top of each other. The distribution of SFRDs (*grey*)

is also sufficiently concentrated with a variance of 0.003 that is similar to that of the SFRDs from the stably convergent posterior in Figure 5.4. Hence, we can conclude that the two peaks are degenerate in terms of the emulated SFRDs.

However, not necessarily does the presence of degeneracy in the emulated SFRD demand that degeneracy exist in the actual cosmological simulations as well. The panel (b) in Figure 5.7 demonstrates the *simulated* SFRDs performed with each set of parameters from the two peaks (*red dashed* and *blue dotted*). Here, we investigate whether two sets of parameters (*red dashed* and *blue dotted*) are also degenerate in the simulations given the simulation uncertainty (refer to Section 5.2.3). We use $\pm 1\sigma$ (*grey*) and $\pm 3\sigma$ (*light grey*) regions that correspond to 68.1% and 99.7% confidence levels assuming that the simulation uncertainty follows the Gaussian distribution, respectively. The standard deviations of the Gaussian distribution are directly calculated from the suite of simulations in the CV set along redshift (refer to Section 5.2.3). In panel (b), *blue dotted* SFRD lies within the 1σ region (*grey*) of the *red dashed* SFRD. Thus, two simulated SFRDs are highly likely to be degenerate sharing the same parameters, while having discrepancies originating from the simulation uncertainty.

Above all, the emulator result demonstrates the clear signs of degeneracy in the SFRD having two approximately identical SFRDs from two different sets of parameters. On the simulation side, cosmological simulations with the two sets of parameters also produce two SFRDs that are close to each other, though simulation uncertainty leads to notable differences in the simulated SFRDs, which are nevertheless not statistically significant. The SFRD lies approximately inside 64.8% (1σ) confidence regions and is completely enclosed within 99.7% (3σ) confidence regions of the simulation uncertainty. This might be strong evidence that both emulator and cosmological simulation have degeneracy in the SFRD, in spite of the emulator not being a perfect representation of the simulations.

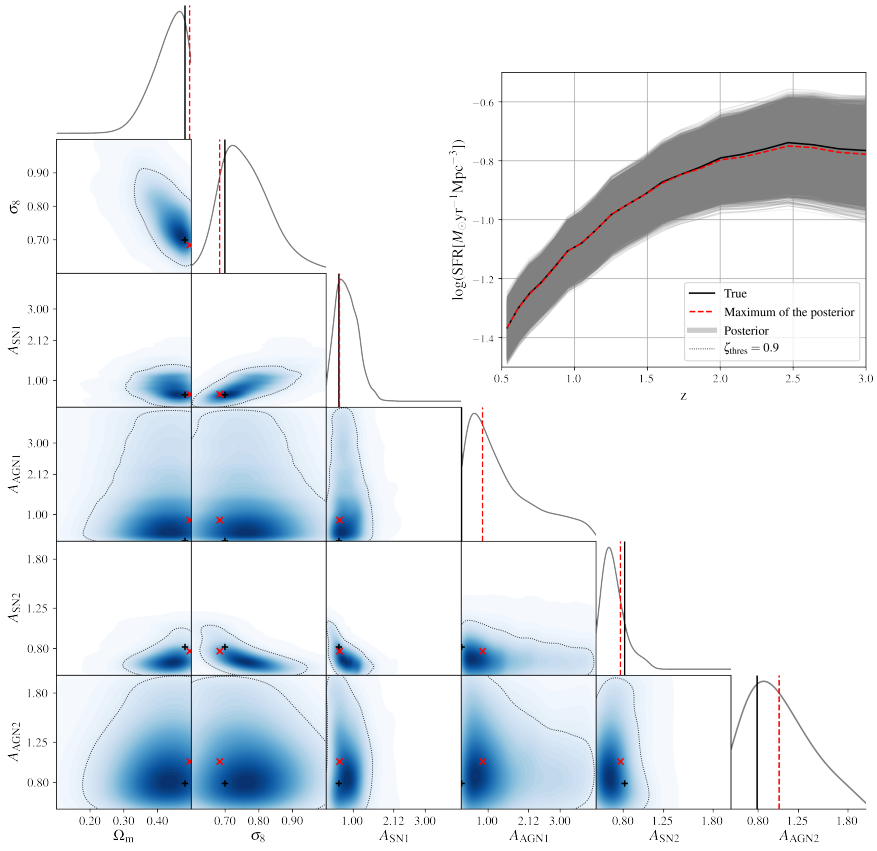

Figure 5.8: *Bottom left*: 2D projections of the inferred posteriors (*blue*) from the emulated SFRD with the mock uncertainty. *Top right*: Cosmic star formation rate density from the inferred parameters (*grey*) and the truth (*black solid*). The *red* vertical lines indicate the locations of the maximum value of the inferred posterior in the six dimensional parameter space.

Response to mock uncertainty

Thus far, we performed SBI without including any uncertainty in the forward model, namely the emulator, but such uncertainties do exist in full cosmological simulations, as discussed in Section 5.2.3. In this section, we include the *simulation uncertainty* that originates from various sources of randomness in the cosmological simulations using the mock uncertainty.

Shown in Figure 5.8 are two-dimensional projections of the posterior density inferred from the emulated SFRD (*bottom left*) with the mock uncertainty. The mock uncertainty that we impose is modelled to have the same standard deviation as the simulation uncertainty (refer to Section 5.2.3). In comparison to the inferences without the mock uncertainty, the posterior densities inferred with the uncertainty cover notable regions of parameter space with the standard deviations $\boldsymbol{\sigma}_{(\Omega_\mathrm{m},\sigma_8,A_\mathrm{SN1},A_\mathrm{AGN1},A_\mathrm{SN2},A_\mathrm{AGN2})}$ of (0.073, 0.100, 0.353, 1.158, 0.145, 0.358) and the coefficient of variation $\mathbf{c}_\mathrm{v}$ of (16.0, 14.7, 60.7, 450.6, 17.8, 47.0)% for the SFRD. In comparison to Figure 5.6, the bimodal peaks are merged to a single oval or banana-shape distribution (*bottom left*). Although the deviations of the inferred A_AGN1 and A_AGN2 become notable, the relative error of SFRD is only 0.38 dex. The variance of the inferred SFRDs (0.057 dex) is comparable to the variance of the mock uncertainty (0.061 dex). This demonstrates that the inferred posterior densities and the corresponding observables successfully reproduce the mock uncertainty in terms of the variances. Thus, concerning the uncertainty propagation from observation to parameters, the variance of inferred parameters can be reliable. On the other hand, the inclusion of the mock uncertainty has led to an increase in the inaccuracy of the posterior as well as the relative errors of the inferred observables to the truth. We can relate it to the stochasticity of sampling of *mock uncertainty*. In every iteration of the SBI, additional training data $(\boldsymbol{x}, \boldsymbol{\theta})$ is generated from the proposal density. Here, the observable $\boldsymbol{x}$ is emulated as a function of the sampled parameters together with the emulator uncertainty $Z(\eta)$. Since the sampling size is finite, the mean of observables $\langle \boldsymbol{x} + Z(\eta) \rangle_\mathrm{samples}$ cannot be the same as the ideal (theoretical) mean $\langle \boldsymbol{x} + Z(\eta) \rangle = \langle \boldsymbol{x} \rangle$, leading to a bias in the sampled data. Notice that for an infinite number of samples (ideal case), $\langle Z(\eta) \rangle = 0$ since $Z(\eta)$ is the Gaussian noise that has a mean of zero. The bias in the newly generated training data is highly likely to result in the inaccuracy of the inferred posterior density.

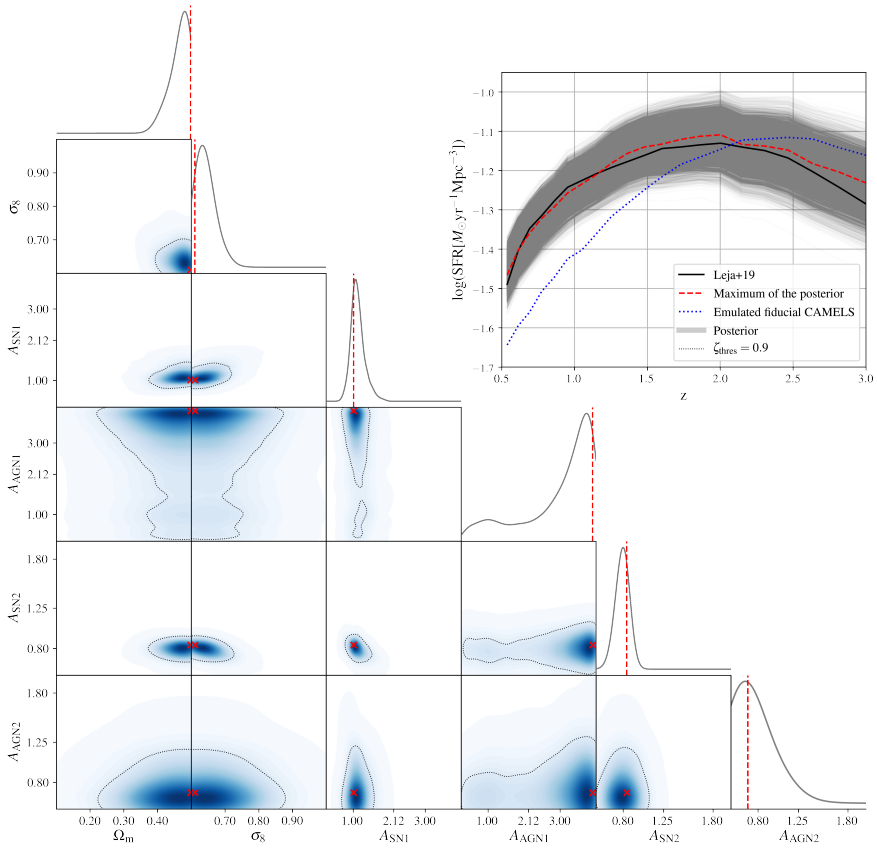

Figure 5.9: *Bottom left*: 2D projections of the inferred posteriors (*grey*) from the observed SFRD [93]. The *red dashed* vertical lines indicate the maximum values of the inferred posterior in the six dimensional parameter space. *Top right*: The inferred cosmic star formation rate density (*red dashd*) and the observed SFRD (*black solid*). The *blue dotted* curve indicates the emulated fiducial SFRD of the CAMELS simulations.

5.3.2 Inference from observed star formation rate density

We now apply our framework to actual observational data. We first, however, must consider some fundamental differences between observations and the cosmological simulations that our emulators are trained on.

For instance, the simulations consistently evolve the baryons and dark matters with which halos and galaxies are made up across cosmic time. In other words, the total mass and implemented physics model of the simulated universes is self-consistent at all times. However, observational data is significantly subject to a survey size and physical models for a particular observable. To make a fair comparison between observations and emulators, we take into consideration the consistency between observed cosmic star formation rate density (SFRD) and stellar mass function (SMF) which is discussed in Sections 5.3.4.

The cosmological simulations, by nature, can guarantee consistency between the evolution of star formation rate and stellar mass. That is, the SFRD $f(z) = \mathrm{d}M_\star(z)/\mathrm{d}z + \dot{M}_{\mathrm{return}}$, where $M_\star(z) = \int_M \phi(M, z)$, $\phi(M, z)$ is the galaxy stellar mass function, and $\dot{M}_{\mathrm{return}}$ is the rate of mass return from evolving stellar populations. The cumulative quantities such as the stellar mass should necessarily coincide with the integrated instantaneous quantities such as the star formation rate. On the observational side, the consistency depends on e.g. the modeling for each observable and related observables such as the SFRD and the SMF at different times are not guaranteed by construction to be consistent with one another. To circumvent this, we adopt SFRD and SMFs from Leja et al. [93] and Leja et al. [94] that aim to resolve the issue of consistency in observations (more details in Leja et al. [93, Section 5.1]).

We first infer the posterior density for the cosmological and astrophysical parameters from the actual observed SFRD [93]. We perform the SBI with an identical setup to the previous sections that includes the mock uncertainty. Figure 5.9 illustrates the two dimensional projections of the posterior density inferred from the observed SFRD. Note that we have not included the uncertainty of the observation data; rather, only the mock uncertainty is adopted. The observed SFRD (*black solid*) lies totally within the region of the inferred SFRDs (*grey*), whereas the SFRD from the maximum of

the posterior (*red dashed*) cannot match it precisely with a relative error of 4.1%. The standard deviations and the coefficients of variation are (0.040, 0.039, 0.224, 1.475, 0.078, 0.181) and (8.3, 5.6, 37.1, 566.9, 9.7, 23.0)%, respectively. The discrepancy between the inferred and observed SFRDs is now notable compared to the previous inference from the emulated SFRD. The relative error is an order of magnitude greater than the case of inference from the emulated SFRD which is 0.38% (e.g. see Figure 5.8).

Nevertheless, it can be thought of as a successful inference given that the emulated fiducial SFRD (*blue dotted* curve in Figure 5.9) shows entirely different trends from the target observation over all redshifts, with an average deviation of ~ 0.1 dex. In addition, the fiducial SFRD has its peak at $z \sim 2.5$, whereas the peak of the [93] SFRD is located at $z \sim 2.0$. Here, the emulated fiducial SFRD (*blue dotted*) is an SFRD that is generated by the emulator from the fiducial parameters, which is close to the simulated SFRD of the CV set (see Figure 5.1 in Section 5.2.2). Despite the huge discrepancy between the fiducial SFRD and the target observation, the inferred SFRD follows the observation relatively well with a mean deviation less than 0.01 dex and even matching the observed peak precisely.

In the case of inference from observations, the intrinsic limit of the emulators should also be accounted for. The most significant difference between the inferences from the emulated SFRD and observed SFRD is that any emulated SFRD can be predicted precisely by the emulator for sure, whereas we cannot ensure whether there exists a data point in parameter space that can reproduce the observation perfectly. Mathematically speaking, the emulator (and simulation) prediction is limited by the set of all SFRDs and SMFs it may produce: the image of the emulator. The image of an emulator might not coincide with the set of all of the physically possible SFRDs, including observed SFRDs: the codomain of the SFRDs. In other words, the emulators and simulations might not be able to reproduce every possible universe. The necessary condition for the successful preci-

sion inference is that the observed SFRD be a member of the image of the emulator. Concerning the limit of the emulator and the simulations that the emulator is trained on, there are fundamental issues: e.g., (1) the limited dimension of parameter space (domain); (2) the gap between the emulators and the simulations; (3) the limits of the physical models in the simulations. The above will be discussed in Section 5.4.4.

5.3.3 Inference from emulated stellar mass functions

We now turn to inference on stellar mass functions (SMFs) as the target observable. We study the dependence of the properties of the SBI and the inferred posteriors on the choice of observable with a comparison to the cosmic star formation rate density (SFRD). The latter covers the evolutionary history of the universe ranging from $z = 3$ to $z = 0.5$ whereas a single SMF contains information of only one epoch. To be consistent, we concatenate five SMFs at $z = 0.5$, 1.0, 1.5, 2.0, and 2.5 and each SMF is binned with 13 bins in the mass range $[10^{8.9}, 10^{11.4}]\mathrm{M}_{\odot}$ (refer to Section 5.2.1). Hereafter, "SMFs" denotes the five concatenated SMFs from the five different redshifts throughout the paper unless specified otherwise. In Section 5.3.3, We first investigate the performance and convergence of the inference on the SMFs compared to that of the SFRDs. In Section 5.3.4, we perform SBI from observed five concatenated SMFs, and lastly, we study the SBI from one *individual* observed SMF (at a single redshift) at a time.

Performance of inference

We first investigate the performance of the SBI and how stably the inferred posterior converges in terms of the stellar mass functions (SMFs). As in Section 5.3.1, we train emulators that take as input six cosmological and astrophysical parameters $\boldsymbol{\theta}$ and predict the SMFs $\boldsymbol{x}$ (for details, refer to Section 5.2.5). Using the emulator, we perform the SBI to retrieve the posterior density for six cosmological and astrophysical parameters $\boldsymbol{\theta}$ given

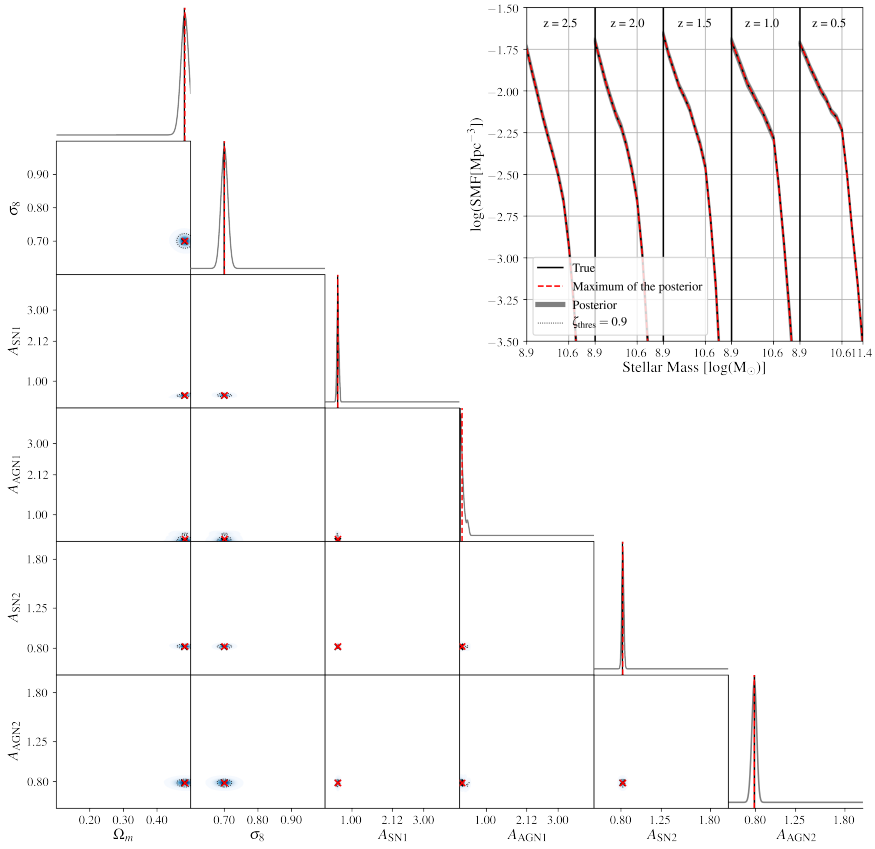

Figure 5.10: *Top right*: *Emulated* stellar mass functions (SMFs) from five different redshifts from the inferred posterior (*grey*), the maximum of the posterior (*red dashed*), and the target SMFs (*black solid*). *Bottom left*: Two-dimensional projections of the inferred posterior. The *black* and *red* cross-hairs represent the values of target and mean of the posterior, respectively. The marginal distributions are obtained by the kernel density estimation. The *black solid* vertical lines indicate the true values. The *red dashed* vertical lines indicate the mean values of the inferred posterior.

the *emulated* SMFs $\boldsymbol{x}_0$.

Shown in Figure 5.10 are two-dimensional projections of the inferred posterior (*bottom left*) and *emulated* SMFs from the posterior along with the target SMFs (*top right*). A total of 6000 simulations (30 epochs) are used to

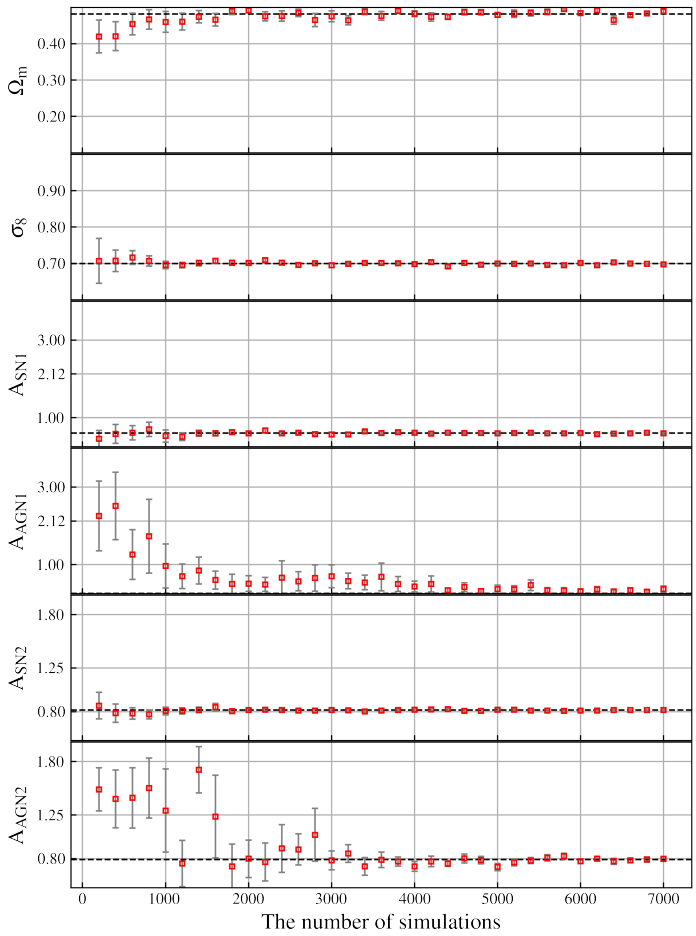

Figure 5.11: The convergence of each parameter as a function of the number of simulations for SBI from emulated stellar mass functions (SMFs). The *red square* and *grey error bar* present the maximum and the standard deviation of the posterior density. The *black dotted* lines show the true values.

retrieve the posterior density $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$ given the SMFs $\boldsymbol{x}_0$. Here, the target SMFs $\boldsymbol{x}_0$ is generated by the emulator with the parameters that we used to generate the emulated SFRD in Section 5.3.1. The maximum of the posterior density[8] (*red dashed* or *red cross-hair*) match the true values (*black solid* or

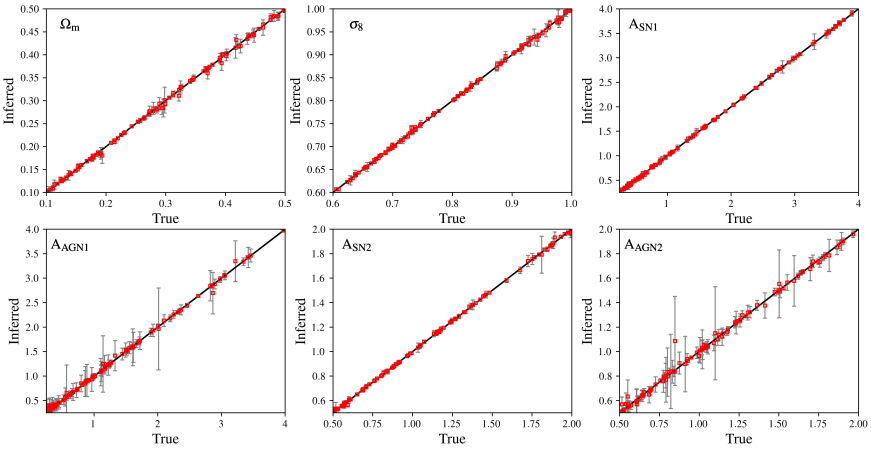

Figure 5.12: Comparison between true values and its inferred values from posteriors for the stellar mass functions without uncertainty. Each *red square* represents the maximum value of the inferred posterior against its true value. Each *grey bar* represents the standard deviation of the inferred posterior. The black solid line represents inferred equals true.

black cross-hair) nearly perfectly for all six parameters with the relative errors $\boldsymbol{\delta}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$=(0.15, 0.01, 0.08, 4.9, 0.72, 0.12)% and with the standard deviations of $\boldsymbol{\sigma}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$=(0.006, 0.002, 0.024, 0.076, 0.006, 0.018) and the coefficients of variations of $\mathbf{c}_{\rm v(}{}_{\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$=(0.9, 0.2, 3.7, 17.7, 0.6, 2.3)%. In the *top right* panel, the SMFs from the maximum of the posterior (*red dashed*) coincide with the true SMFs (*black solid*) with a relative error of 0.4%. The SMFs from the full posterior (*grey region*) have an exceedingly narrow distribution with the mean standard deviation of 0.007 dex. In comparison to the SFRD in Section 5.3.1, the relative errors of SMFs are slightly better except $A_{\rm SN2}$, whereas the standard deviations of the SMFs are comparable to that of the SFRD. In general, the error of inference can be related to the correlation between the parameters and the observable. We discuss why $A_{\rm SN2}$ can be more precisely predicted from the SFRDs in Section 5.4.1, where we show how

surprisingly the errors presented here are well explained by the correlations.

Figure 5.11 illustrates the convergence and its stability for the inference from the emulated SMFs. Each panel shows maxima (*red square*) and variances (*grey errorbar*) of the posterior for a given parameter as a function of the (cumulative) number of simulations used for training so far (i.e. the number of epochs $\times$ the number of new simulations per epoch). All six maxima (*red square*) stably converge to the truth (*black dotted line*) with the average relative errors[10] over $[4800, 8000]$ simulations: $\boldsymbol{\delta}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ $= (1.4 \pm 1.1,\, 0.3 \pm 0.2,\, 1.2\pm,\, 1.1\ 34.1 \pm 17.8,\, 0.6 \pm 0.4,\, 1.9 \pm 1.3)\%$. The variances of the posterior density (*grey errorbar*) are convergent with the average standard deviation $\boldsymbol{\sigma}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})} = (0.007,\, 0.002,\, 0.022,\, 0.076,\, 0.006,\, 0.028)$ and the coefficient of variation $\mathbf{c}_{\rm v(\Omega_{\rm m},\sigma_8,A_{\rm SN1},}$ $_{A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ =(0.2, 0.1, 2.6, 4.9, 0.7, 20.2)% over [4800, 8000] simulations. Hence, the convergence of the inference requires at least ~ 4000 simulations.

In contrast to the SFRD, the unstably convergent stage is absent in the SMFs. To be stably convergent, the SMFs require only 4000 simulations whereas ~ 40000 simulations are needed for the SFRD, an order of magnitude difference. Moreover, the average variances at the stably convergent stage, $\bar{\sigma}_{\rm smf} \sim \mathcal{O}(1)$ for the SMFs and $\bar{\sigma}_{\rm sfrd} \sim \mathcal{O}(10)$ for the SFRD, demonstrate that the SMFs converge with less fluctuations. Hence, we conclude that the SMFs converge far more rapidly and stably to the truth compared to the SFRD. Furthermore, the NDE of the SFRD has more hidden units ($N_{\rm hid,sfrd} = 250$) than that of the SMFs ($N_{\rm hid,smf} = 100$). Given that the more neurons the easier it converges, we can conclude that the SMFs-parameters pairs are more apt to be mapped than the SFRD-parameters pairs. This indicates that the degree of correlation between the SMFs and the parameters is stronger than between the SFRD and the parameters (as discussed further in Section 5.4.1).

We also investigate the accuracy of the inference for various SMFs over

the parameter space. Figure 5.12 shows the maxima (*red square*) and variances (*grey errorbar*) of 100 posterior densities inferred from 100 *emulated* SMFs. Here, 100 SMFs are generated by the emulators with 100 randomly-sampled sets of parameters from the LH set (refer to Section 5.2.1). We consistently perform the SBI with the total of 4800 simulations over all 100 SMFs without any further convergence tests. Due to computational cost, the minimal (necessary but perhaps not sufficient) number of simulations for the convergence is adopted[11], deduced from the previous convergence test in Figure 5.11. Most maxima (*red square*) are on top of the ideal prediction line (*black solid*) or are indistinguishably close to it. The average relative errors[10] of the maximum are (1.1 ± 1.2, 0.2 ± 0.2, 1.3 ± 2.2, 4.0 ± 8.5, 0.6 ± 0.8, 1.8 ± 3.4)%. These values are in line with the relative errors from the convergence test. Thus, the SBI can be performed on the SMFs stably with relatively small, constant errors regardless of the choice of parameters.

No response to uncertainty

We perform SBI on the SMFs with the mock uncertainty (for details of mock uncertainty for the SMFs, refer to Sections 5.2.5 and 5.2.3). However, unlike the uncertainty case of the SFRD which shows appreciable variances in the inferred posterior, the posterior density inferred from the SMFs *with* the mock uncertainty is *identical* to the posterior density inferred *without* the mock uncertainty. Note that the figure for the *with-uncertainty* case is not presented here since it is *so identical* to the *without-uncertainty* case that we *cannot* tell any difference both visually and quantitatively. We find that the mock uncertainty that mimics the simulation uncertainty cannot have any impact on inference from the SMFs in contrast to the SFRD. Hence, in Section 5.4.3, we study how the inferred posteriors response to various types of uncertainties and which type of uncertainty is suitable for the SBI

[11] For the same reason, we cannot produce this plot for the SFRD since the SFRD requires significantly more simulations than the SMFs and owing to numerous hidden units for the SFRD, the computational time is at least 10 times greater than that of the SMFs.

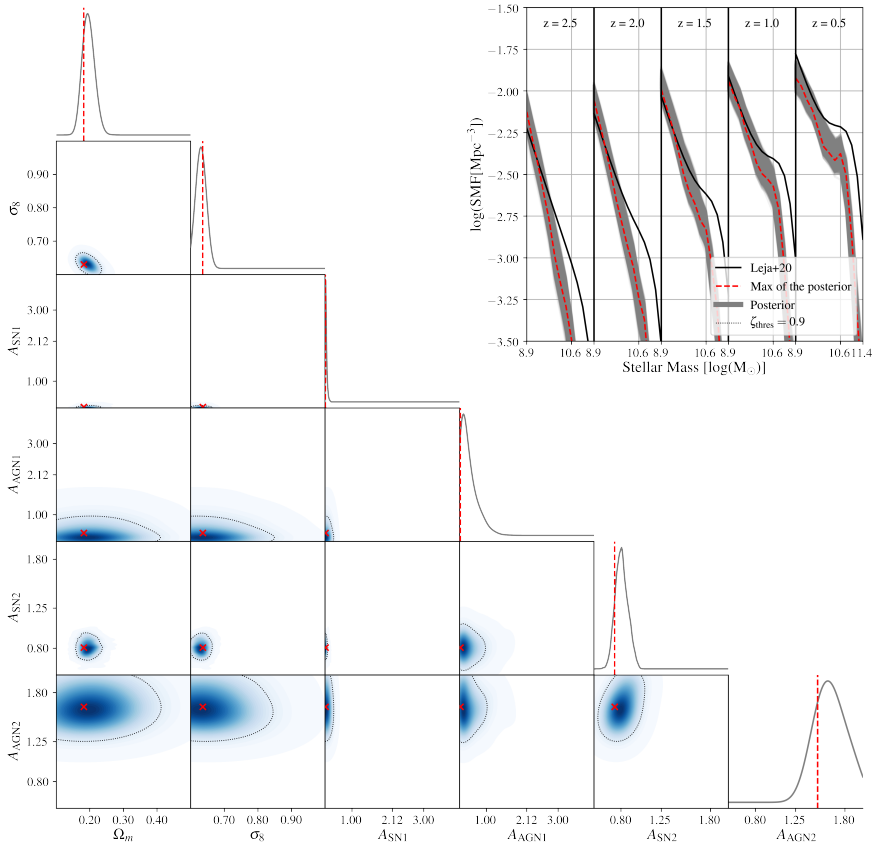

Figure 5.13: *Bottom left*: 2D projections of the inferred posteriors for the observed stellar mass functions (SMFs) from five different redshifts (z =2.5, 2.0, 1.5, 1.0, and 0.5) [94]. *Top right*: the SMFs from the inferred parameter (*grey* and *red*) and the observed SMFs (*black*). The *red* vertical lines indicate the maximum values of the inferred posterior in the six dimensional parameter space.

in detail.

5.3.4 Inference from observed stellar mass functions

Five concatenated stellar mass functions

Figure 5.13 shows two-dimensional projections of the posterior density

inferred from the observed SMFs [94] (*bottom left*) and the corresponding SMFs (*top right*). The discrepancy between the observation (*black solid*) and the maximum of posterior (*red dashed*) is considerable with a mean relative error of 41.1 % over all mass ranges. The discrepancy of the high mass ends is dominant for the errors with the relative error for stellar masses $\gtrsim 10^{11}\mathrm{M}_{\odot}$ being $\sim 80.9\%$ versus 13.7% at the lower mass ends. The high mass ends ($\gtrsim 8 \times 10^{9}\mathrm{M}_{\odot}$) of the observed SMF are located completely outside the region of the SMFs from the full posterior (*grey*) across all five redshifts. The population of the massive galaxies in the observed SMF is far greater than that in the inferred SMFs.

At the low mass ends, even though the observed SMFs lie within the (*grey*) region, the emulated and observed SMFs have different, distinct characteristics: (1) the slope of SMFs at high redshift and (2) the evolutionary rate of the SMFs. At $z = 2.5$ and $z = 2.0$, the slopes of the inferred SMFs are notably steeper than the observed SMFs. The observed SMFs do not show the appreciable difference in slope across the redshifts, whereas the slopes of inferred SMFs become less steep as redshift decreases. Also, there is a significant difference in the evolutionary rate of SMFs. The evolution of the inferred SMFs is barely notable, but the apparent growth can be seen in the observation. For example, the differences between SMFs at $M_{\star} = 10^{8.9}\mathrm{M}_{\odot}$ at $z = 2.5$ and $z = 0.5$ are 0.20 dex for the inference and 0.47 dex for the observation.

On the simulation side, the significant mismatch between the inference and observation can be attributed to the following: (1) the limited volume of simulations; (2) low resolution of simulations and the failure of re-scaling; (3) limitations of the physical models of simulations; (4) inaccuracy of the emulators. The size of simulation is directly linked to the level of simulation uncertainty that we fail to realize in the SMFs. Since we cannot account for the simulation uncertainty, the inference results are suffering from sampling bias. Also, the problem of lack of massive galaxies are prevalent in a

Table 5.2: The maximum of the posteriors, $\boldsymbol{\theta}_{\max}$, (*first-fourth rows*) in Figures 5.13 and 5.14 and their standard deviations σ (*fifth-eighth rows*) with respect to the maximum of the posteriors. 'ALL' indicates that the parameters are inferred from all five redshifts simultaneously (*red cross-hair* in Figure 5.13).

		$\Omega_{\rm m}$	σ_8	$A_{\rm SN1}$	$A_{\rm AGN1}$	$A_{\rm SN2}$	$A_{\rm AGN2}$
$\boldsymbol{\theta}_{\max}$	ALL	0.18	0.64	0.26	0.27	0.73	1.50
	$z = 0.5$	0.22	0.97	0.41	1.50	0.51	0.52
	$z = 1.5$	0.10	1.00	0.47	0.25	0.58	1.63
	$z = 2.5$	0.11	1.00	0.72	2.77	0.83	0.61
σ	ALL	0.02	0.01	0.03	0.31	0.10	0.22
	$z = 0.5$	0.02	0.04	0.11	1.32	0.18	0.71
	$z = 1.5$	0.01	0.03	0.04	1.38	0.04	0.35
	$z = 2.5$	0.05	0.03	0.11	1.02	0.06	0.82

low-resolution simulations (Appendix A. in Pillepich et al. [136, 138] and Section 5.2.2 in this paper). Even though we apply re-scaling to alleviate the resolution effects, the rescaled SMFs are still subject to resolution convergence (refer to Section 5.2.2). Inside the parameter space that we adopt in this chapter, there might not exist a set of parameters that can reproduce the observations simultaneously across redshifts. Lastly, since we are using the emulators as a surrogate to cosmological simulations, the inference results are also limited by the accuracy of emulators. Further will be discussed in the next part and Section 5.4.4. Note that technical and physical limits of observation can also be crucial factors for the mismatch but will not be discussed in this paper.

Three individual stellar mass functions

To isolate the problem, we perform the inferences from one SMF at a time for each redshift ($z = 0.5$, 1.5, and 2.5) separately. Figure 5.14 illustrates the three posterior densities inferred from the observed SMFs (*thick black solid*) at $z = 0.5$ (*sky blue*), $z = 1.5$ (*violet*), $z = 2.5$ (*red*) and the corresponding SMFs (*top right*). The separate inferences lead to signif-

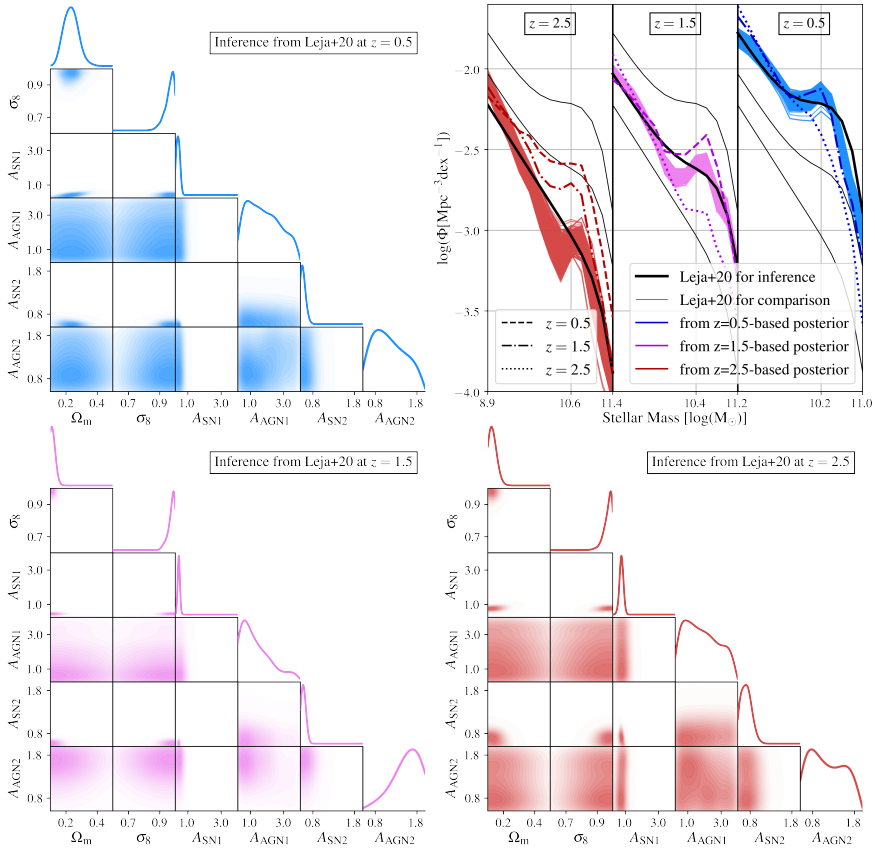

Figure 5.14: Two-dimensional projections of the inferred posteriors from the individual stellar mass functions (SMFs; Leja et al. [94]) at the different redshifts: $z = 0.5$ (*top left sky blue*), $z = 1.5$ (*bottom left, violet*), and $z = 2.5$ (*bottom right, red*). The *top right* panel shows three inference results from the target observed SMFs (*thick black solid*) from different redshifts $z = 2.5$ (*left*), 1.5 (*middle*), and 0.5 (*right*). Here, *thin* black solid lines are the observed SMFs that are not involved in the inferences at a particular redshift. The *dashed*, *dotdashed*, and *dotted* lines are the emulated SMFs for $z = 0.5$, 1.5, and 2.5 from the maximum of the posteriors inferred from the observed SMFs at $z = 0.5$ (*dark blue*), 1.5 (*dark violet*), and 2.5 (*dark red*).

icantly higher accuracy than the previous inference, with the relative errors of SMFs of 17.7% for $z = 2.5$, 10.3% for $z = 1.5$, and 10.3% $z = 0.5$ (the

relative error of the previous inference is 31.4%). In comparison to the previous inference (Figure 5.13), the locus of the posterior densities are notably different, especially in σ_8 and $A_{\rm SN1}$ (refer to Table 5.2). The standard deviations of the AGN parameters in Figure 5.14 are notably larger than those from the inference based on the five concatenated SMFs in Figure 5.13, whereas the standard deviation of the inferred SMFs are similar. This might indicate that the impact of the AGN parameters on the SMFs is trivial. In general, the impact of the AGN parameters on the SMFs can be negligible in this analysis, given the extensive variances that cover almost entire ranges of parameter space. The average variances of $A_{\rm AGN1}$ and $A_{\rm AGN2}$ over the three redshifts are 1.23 and 0.62, whereas the variances of other parameters are less than 0.01 on average.

σ_8 from the separate inferences show considerably higher values than the previous inference, whereas $\Omega_{\rm m}$ show relatively small changes. In both emulators and simulations, the AGN parameters only have a minor effect in the stellar populations of galaxies (Sections 5.4.1). As a compensatory action, the emulated universes have exploited the cosmological parameters, especially σ_8, to control the populations of massive galaxies in the context of structure formation. Meanwhile, to sustain the density of low mass galaxies, stronger supernova feedback can be inevitable, which we can see (the increases in $A_{\rm SN1}$, even though it is not significant) in Figure 5.14.

In the *top right* panel consisting of three subpanels, the emulated SMF for $z = z_X$ obtained by the maximum of the posteriors inferred from the observed SMF at $z = z_Y$ is shown in "Y X" line style where $X_{z_X=(0.5,1.5,2.5)} =$ (dashed, dash-dotted, dotted) and $Y_{z_Y=(0.5,1.5,2.5)}$ =(dark blue, dark violet, dark red). For example, the emulated SMF at $z = 0.5$ from the maximum of the $z = 1.5$-based posterior (*bottom left, violet*) is drawn with a "dark violet dashed" line. Each subpanel exhibits the discrepancies of the slope and evolution of SMFs across different redshift (different line styles). Notice that although the separate inferences can predict observations with much im-

proved accuracy, the discrepancy cannot be resolved. Further discussion on the discrepancies between inferred SMFs and observed SMFs is described in Section 5.4.4 in detail.

The mismatch between inference and observations significantly highlights the intrinsic limits of emulators and simulations: (1) the limited dimension of parameter space (domain); (2) simulation volume and resolutions; (3) inaccuracy of the emulators; (4) the limits of the physical models in the simulations.

5.4 Discussion

5.4.1 Correlation between observables and parameters

The correlation plays a significant role in mapping between domain (input) and codomain (output), such as in machine learning and inference. What a correlation between two sets measures is how dependent the two sets are on each other. In machine learning, the degree of correlation can have a decisive impact on the precision of a machine. As input and output are more correlated, the sensitivity of output to input increases. This results in a higher error of the objective function (or loss function) in response to a minute change in input. Thus, given the same amount of error tolerance, the accuracy of a machine increases as the correlation increases. In a similar context, the variance of posterior is an indirect indicator of the correlation between parameters and observable. The less correlated they are, the larger the variances and vice versa. We have already encountered several trails of the correlation between parameters and observables in the inference results. One of the most evident examples is that the SFRD requires $\mathcal{O}(10^4)$ simulations to reach the stable convergence, whereas only $\mathcal{O}(10^3)$ simulations are sufficient for the SMFs (see Sections 5.3.1 and 5.3.3 for the SFRD and the SMFs, respectively). In other words, with the same size of training data—i.e. the same number of simulations, the SMF would have attained a more

accurate posterior with a smaller variance than the SFRD would;

For a quantitative analysis, We measure the correlation between parameters and observables using the mutual information (MI), which is a fundamental measure for the inter-dependence or relationship between two variables. In contrast to linear correlation coefficients, such as the Pearson correlation coefficient, the mutual information captures non-linear statistical dependencies [78]. The mutual information is defined by

$$I(X,Y) = D_{\mathrm{KL}}(P_{XY} \| P_X \otimes P_Y) \tag{5.4}$$

where $D_{\mathrm{KL}}(\cdot\|\cdot)$ is the Kullback-Leibler divergence[12], and P_{XY} and P_X are joint distributions and marginal distributions. Here, $\otimes$ denotes the outer product that spans the probability space from X and Y to (X, Y). $I(X, Y)$ quantifies a statistical distance between the joint probability and product of marginals as per property of the Kullback-Leibler divergence. The mutual information is zero if and only if $X,\ Y$ are independent.

However, the estimation of the mutual information is challenging and only tractable for discrete variables or when probability distributions are known [131]. Thus, we adopt the mutual information regression function in the `sklearn` package [88, 135, 152]. Using the package, we estimate the mutual information for (1) simulated observable and parameters from the LH set, (2) emulated observable and parameters from the LH set. Here, the observables and the parameters are normalized to reduce the effects of the difference in magnitude of values in the same way used to train an emulator (Section 5.2.5).

Table 5.3 shows the estimated mutual information of each observable-parameter pair. In the simulations, the mutual information of the SMFs

[12] A Kullback-Leibler divergence, also known as relative entropy, is a measure of how one probability distribution $Q(x)$ is different from the other reference probability distribution $P(x)$ and it is defined as follows: $D_{\mathrm{KL}}(P\|Q) \equiv \sum_{x\in\mathcal{X}} P(x)\log(P(x)/Q(x))$ or $\int_{-\infty}^{\infty} p(x)\log(p(x)/q(x))\mathrm{d}x$.

Table 5.3: The mutual information between the parameters and obseravbles (*first-fourth* rows). The relative errors $\boldsymbol{\delta}$ of the convergent SFRD and SMFs from Sections 5.3.1 and 5.3.3, respectively (*fifth* and *sixth* rows). '*Sim*' and '*Emu*' stands for "simulated" and "emulated". The values of the MI are multiplied by 100 for convenience and the unit for the relative error is percentage.

	$\Omega_{\rm m}$	σ_8	$A_{\rm SN1}$	$A_{\rm AGN1}$	$A_{\rm SN2}$	$A_{\rm AGN2}$
SimSFR	1.19	0.62	0.54	0.29	2.11	1.43
SimSMF	12.5	9.27	36.9	1.17	4.69	0.41
EmuSFR	11.4	6.82	30.0	0.7	17.1	0.85
EmuSMF	14.6	12.8	41.4	1.49	5.79	0.30
$\boldsymbol{\delta}_{\rm sfr}$ (%)	0.20	0.49	1.36	49.8	0.25	3.65
$\boldsymbol{\delta}_{\rm smf}$ (%)	0.15	0.01	0.08	4.9	0.72	0.12

(*SimSMF*) is significantly higher than that of the SFRD (*SimSFR*) over all parameters except $A_{\rm AGN2}$. Interestingly, the gap has been drastically reduced in the emulators (*EmuSMF* and *EmuSFR*). The increase of mutual information in the emulators (especially in the SFRD) can be attributed to the training of the emulators. During training, the emulator (neural network) keeps updating itself to minimize the loss function or maximize mutual information (correlation) between input and output [54, 60]. As a result, input and *emulated* output pairs can attain overall more correlation via the emulator than input and *simulated* output pairs. Note that the correlation of the *emulated* pairs should depend on the precision of training. Nonetheless, the relative magnitudes among each parameter-observable pair remain the same. For instance, the mutual information of SMF-parameter pairs is still higher than that of SFRD-parameter pairs except for $A_{\rm SN2}$ and $A_{\rm AGN2}$.

In addition to the gap between simulations and the emulator, the MI among the parameters shows considerable differences in magnitude. $A_{\rm SN1}$ is found to be the most relevant parameter to both SFRD and SMFs, whereas the AGN feedback parameters ($A_{\rm AGN1}$ and $A_{\rm AGN2}$) have MI that is one or two orders of magnitude lower than other parameters. That is, $A_{\rm SN1}$ can

precisely be inferred or estimated in terms of inference and machine learning. On the other hand, not only the AGN parameters can be hardly constrained but also the AGN physics themselves might have a negligible impact on the SFRD and the SMFs in the TNG suite of the CAMELS simulations that have low resolution and limited volume. Such impact of relative magnitudes among parameter-SFRD pairs can also be found in Villaescusa-Navarro et al. [185, Figure 9] which demonstrates that the SFRD is only sensitive to $\Omega_{\rm m}$, σ_8, $A_{\rm SN1}$, and $A_{\rm SN2}$, but not $A_{\rm AGN1}$ and $A_{\rm AGN2}$.

The values of the MI are also in line with the performance of inference with respect to the parameters. For instance, the MI between $A_{\rm SN2}$ and the SFRD, which is an exceptional case, is larger than that between $A_{\rm SN2}$ the SMFs, which is in line with the relative error of $A_{\rm SN2}$ in the SFRD (*fifth* rows) being also smaller than that of the SMFs (*sixth* row). The ratio among the MIs of each parameter is inversely proportional to that of the relative error and/or the variance for all inferences presented in this chapter. In addition, the convergence depends on the degree of the correlation of the pairs, implying that more training data—i.e. simulations—are required to converge in the SFRD (compare Figures 5.5 and 5.11). Lastly, it is also in agreement with previous work. The CAMELS introduction paper [185] builds a fully-connected neural network to predict cosmological and astrophysical parameters taking the SFRD as input. With the neural network, Ω_m, σ_8, $A_{\rm SN1}$, and $A_{\rm SN2}$ are predicted relatively well compared to $A_{\rm ANG1}$ and $A_{\rm ANG2}$ (Villaescusa-Navarro et al. [185, Figure 11]).

We now focus on the MI of the SMF-parameter pair with respect to stellar mass. Figure 5.15 specifically shows the MI between each parameter and the SMF at $z = 0.5$ (*solid*) and 2.5 (*dotted*) as a function of stellar mass. At $z = 0.5$, $A_{\rm SN1}$ is dominant over other parameters especially for $M_\star \lesssim 10^{10.8}{\rm M}_\odot$, while $\Omega_{\rm m}$ also has appreciable MIs. In the higher mass end ($M_\star \gtrsim 10^{10.8}{\rm M}_\odot$), $A_{\rm SN2}$ is most effective, followed by σ_8 and $A_{\rm SN1}$, whereas the MIs of $A_{\rm AGN1}$ and $A_{\rm AGN2}$ are negligible across the

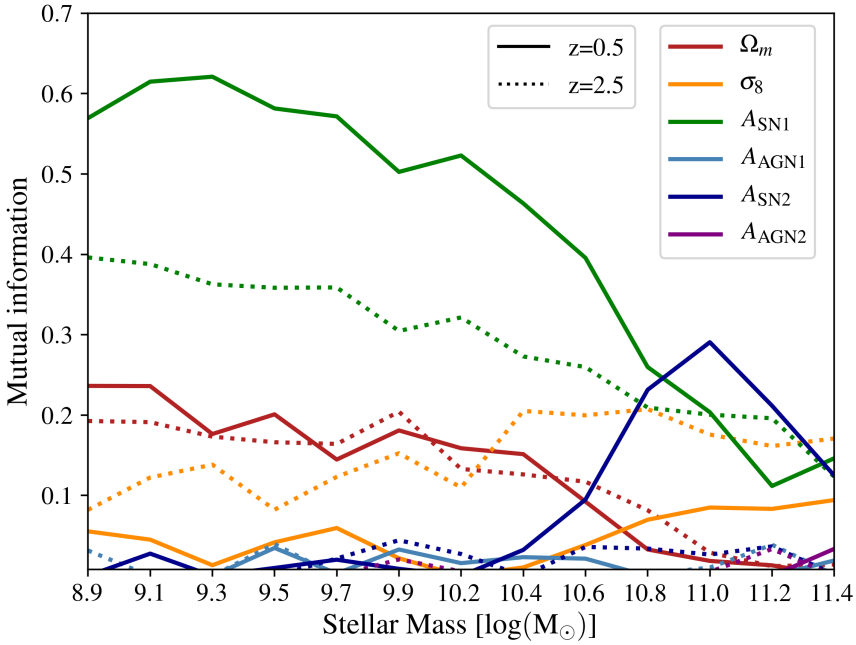

Figure 5.15: Mutual information of the stellar mass function (SMF) at $z = 0.5$ (*solid*) and 2.5 (*dotted*) with respect to stellar mass. Each colored line indicates mutual information between one parameter and the SMF.

entire stellar-mass range. Interestingly, although the average MI of $\Omega_{\rm m}$ is higher than $A_{\rm SN2}$, $A_{\rm SN2}$ can be more practically effective than $\Omega_{\rm m}$ since $A_{\rm SN2}$ dominates the higher mass end of the SMF. In the case of the higher redshift ($z = 2.5$), the order of the relative magnitudes of the MIs among parameters has changed. $A_{\rm SN1}$ is still the most dominant parameter in the low mass end, but the gap with $\Omega_{\rm m}$ is smaller. The MI of σ_8 is drastically larger overall and dominates the high mass end, whereas $A_{\rm SN2}$ becomes negligible. Compared to $z = 0.5$ (*solid* lines), $\Omega_{\rm m}$ (*red dotted*) remains approximately the same and σ_8 (*yellow dotted*) is significantly increased, while the $A_{\rm SN1}$ (*green dotted*) and $A_{\rm SN2}$ (*blue dotted*) are appreciably decreased, which, overall, leads to the increase in the ratio of MI of the cosmological parameters to the astrophysical parameters. This can be indicative that cosmology

is more influential at early times than late times compared to astrophysical phenomena.

The MI can be an indirect but crucial measure for the degree of the relative impact of parameters on observables. We can extract physical insights from Figure 5.15 as follows: (1) The impact of cosmology ($\Omega_{\rm m}$ and σ_8) diminishes as time goes by in terms of the SMF. (2) The energy budget of the supernova-driven winds ($A_{\rm SN1}$) is of paramount importance in the galaxies of stellar mass $\lesssim 10^{10.8}{\rm M}_\odot$. (3) For massive galaxies ($M_\star \gtrsim 10^{10.8}{\rm M}_\odot$), the stellar wind velocity ($A_{\rm SN2}$) has more effect than its energy budget ($A_{\rm SN1}$). (4) The portion of dark matter ($\Omega_{\rm m}$) in the universe has more impact on lower stellar-mass galaxies than higher-mass galaxies, (5) whereas the density fluctuation (σ_8) is more effective in massive galaxies. (6) The kinetic feedback of the black holes ($A_{\rm AGN1}$ and $A_{\rm AGN2}$) are negligibly weakly related with the SMF regardless of both redshift and stellar mass of the galaxies. Note that these results are subject to the resolution effect and the implementation of subgrid physics. Above all, the importance of AGN feedback in massive galaxies has been reiterated by many authors [17, 30, 43, 50, 139, 151]. In contrast, both inference results and MIs demonstrate the insignificance of the AGN physics overall. Thus, we discuss why the AGN parameters become fruitless parameters in the following section.

AGN feedback weakly correlated with observables

We have seen the weak correlation between the observables and AGN feedback parameters from both the variances of the posteriors and the mutual information. In this section, we investigate the impact of the AGN parameters on the properties of their host galaxies. The AGN parameters in this work control the energy budget of the low-accretion mode (kinetic feedback) of black holes ($A_{\rm AGN1}$) and its burstiness ($A_{\rm AGN2}$) (refer to Section 5.2.1). The kinetic wind from the black holes can play a more crucial role in quenching massive galaxies than thermal feedback can, since

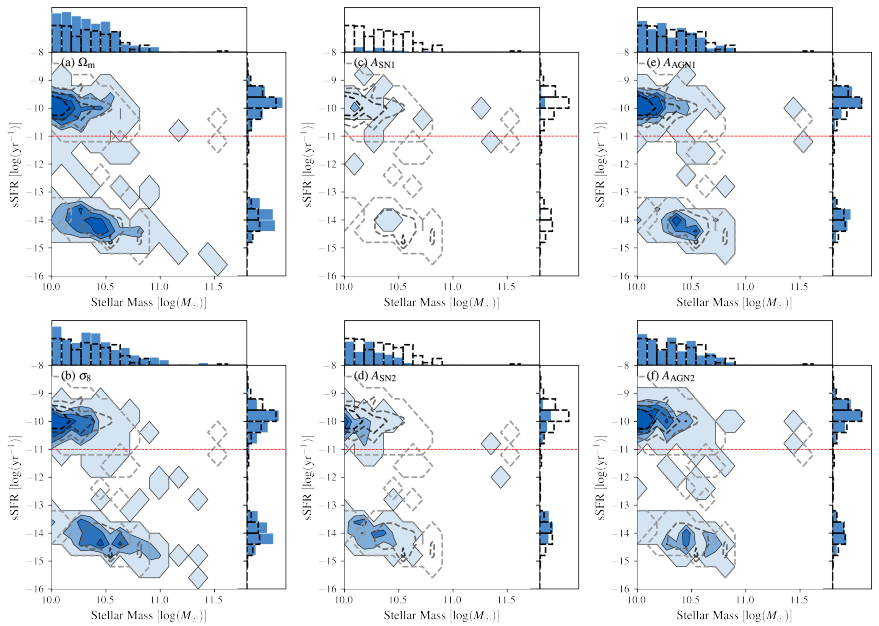

Figure 5.16: Two-dimensional specific star formation rate (sSFR)-stellar mass contour plots and one-dimensional histograms of sSFR and stellar mass for six extreme parameters. Six *grey scale dashed*-contours and pairs of *black dashed* bars are from the fiducial parameter set. Six *blue-filled* contours and pairs of *blue* bars are from $\Omega_{\rm m} = 0.5$ (*top left*), $\sigma_8 = 1.0$ (*top middle*), $A_{\rm SN1} = 4.0$ (*top right*), $A_{\rm SN2} = 2.0$ (*bottom left*), $A_{\rm AGN1} = 4.0$ (*bottom middle*), and $A_{\rm AGN2} = 2.0$ (*bottom right*). Each panel illustrates the extreme case of each parameter. Here, the other parameters are fixed as the fiducial values (refer to Section 5.2.1). The *red* horizontal lines indicate the threshold below or above which a galaxy is quiescent or star-forming.

the kinetically-injected energy is less vulnerable to gas (over-)cooling, especially for dense gas where the cooling time is short. For instance, the black hole-driven winds not only efficiently heat up and eject the gas within the host galaxy, which ends up becoming quiescent, but also affect the gas beyond the galaxy by increasing gas entropy and the cooling time of the CGM [205]. Furthermore, Terrazas et al. [178] find that the galaxy can be quenched when the cumulative kinetic feedback energy of the central black

hole exceeds the gravitational binding energy of the gas within the galaxy. In this regard, the lack of correlation between the AGN parameters and the SMFs in CAMELS, even at the higher mass end, seems puzzling.

Hence, we study how the changes in kinetic winds ($A_{\rm AGN1}$ and $A_{\rm AGN2}$) affect the star formation rate (SFR) in the higher mass end ($M_\star > 10^{10}{\rm M}_\odot$) and compare them with the impact of the other parameters ($\Omega_{\rm m}$, σ_8, $A_{\rm SN1}$, and $A_{\rm SN2}$). First, we define as "quiescent" those galaxies whose specific star formation rate (sSFR) is lower than $10^{-11}{\rm yr}^{-1}$. Second, although the time-averaged SFR over tens or hundreds of Myr is a more common choice for studying SFR as an observable, we adopt the instantaneous SFR since the results can remain qualitatively the same [178]. Lastly, due to the mass resolution limits, there exist galaxies whose SFR is zero. To keep them in the analysis, we randomly assign their SFRs in the range $[10^{-4}, 10^{-3}]{\rm M}_\odot{\rm yr}^{-1}$, following Donnari et al. [41, 42], Terrazas et al. [178].

Shown in Figure 5.16 are sSFR-$M_\star$ diagrams at $z = 0.5$ with respect to the six extreme cases. We vary the values of six parameters to the extreme level from the fiducial case but only one at a time, using the 1P set of CAMELS [186, see Section 3.3.2]. Note that all the simulations presented in this section are performed with the same initial conditions to minimize the effects of simulation uncertainty through cosmic variance. The fiducial sSFR-$M_\star$ relation is indicated with *black dashed* contours and histograms identically in the six panels. Note that all color map and axis limits are consistent across the six panels. In the case of cosmology ($\Omega_{\rm m}$ and σ_8) in the *left* panels (a) and (b), the increase in either $\Omega_{\rm m}$ or σ_8 leads to increased abundance of massive galaxies. Interestingly, the increase in the quiescent galaxies that are located below the threshold line (red dashed) is notable while the population of star-forming galaxies is similar to the fiducial case (*black dashed*). The fractions of the quiescent galaxies of the $\Omega_{\rm m}$ and σ_8 cases are significantly increased from the fiducial case (35%) to cosmology variation cases (50% and 49%, respectively).

On the other hand, the star formation-driven winds ($A_{\rm SN1}$ and $A_{\rm SN2}$) in the *middle* panels (c) and (d) suppress the formation of galaxies with high stellar mass to a large degree, especially $A_{\rm SN1}$. $A_{\rm SN1}$ has a significant impact on both star-forming and quiescent galaxies, leaving exceedingly small number of galaxies in the stellar mass $\geq 10^{10}{\rm M}_{\odot}$ compared to the fiducial case. However, the quiescent galaxy fraction has decreased to 13% from the fiducial case of 35%. Meanwhile, the strong stellar-wind velocity ($A_{\rm SN2}$) also suppresses the massive population of galaxies, but contrarily to $A_{\rm SN1}$, the quiescent galaxy fraction has increased to 56%.

In contrast, the black hole winds have the most minor impact on both the high-mass population and SFR in the massive galaxies. The increase in the energy budget of the black hole winds ($A_{\rm AGN1}$) can suppress the star-forming galaxies with a quiescent-galaxy fraction of 41%, whereas the change in burstiness ($A_{\rm AGN2}$) hardly affects both star formation and population of the massive galaxies, with a quiescent-galaxy fraction of 35%, as in the fiducial run (35%). This result is approximately in agreement with both Figure 5.15 (*solid lines*) and Table 5.3 (*SmSMF*) that show mutual information between each parameter and the SMF at z=0.5. The mutual information of the stellar wind feedback ($A_{\rm SN1}$ and $A_{\rm SN2}$) is dominantly high, followed by $\Omega_{\rm m}$ and σ_8, whereas the kinetic black hole feedback ($A_{\rm AGN1}$ and $A_{\rm AGN2}$) has exceedingly small mutual information across the entire mass range.

It is known that the black hole physics, especially kinetic feedback, is more responsible to the population of massive galaxies than stellar physics such as supernova [43, 127, 174, 178]. This argument can still be valid, but our analysis introduces a slightly different point of view on it in the context of the fraction of quiescent galaxies. The black-hole kinetic feedback can be a primary mechanism for suppression of star formation in massive galaxies. For instance, cosmological simulations without black hole kinetic feedback (only with thermal feedback) hardly quench the star formation in massive galaxies [178, see Figure 1]. However, we find that the black hole kinetic

feedback has a negligible impact on sSFR-$M_\star$ compared to cosmological parameters and stellar wind feedback. One possible explanation is that the implemented AGN feedback ($A_{\rm AGN1}$ and $A_{\rm AGN2}$) might be sufficiently effective, even in small value of AGN parameters, so that the changes in AGN feedback can hardly have a critical impact on the fraction of the quiescent galaxies compared to the other physics.

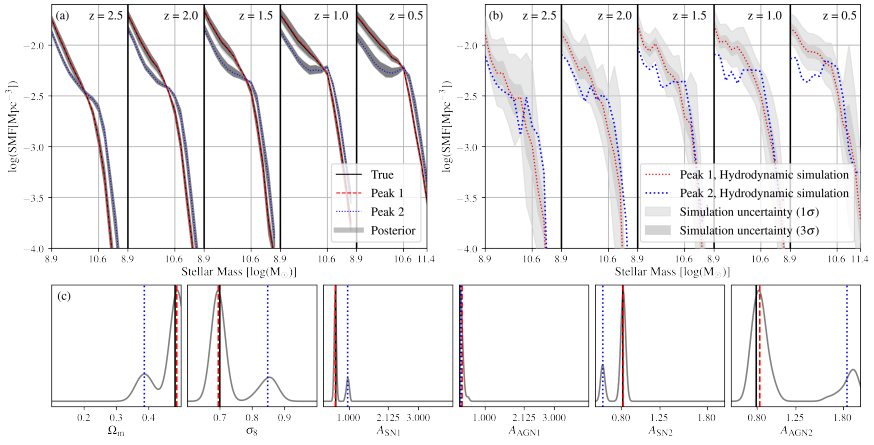

Figure 5.17: (a) Stellar mass functions from the emulator; (b) Stellar mass functions from the cosmological simulation; (c) One-dimensional projections of the inferred posterior based on Figure 5.6. The two peaks in Figure 5.6 are drawn with the *red dashed* and *blue dotted* lines consistently. Note that there exists an offset between the emulated SMFs and simulated SMFs. This is because the emulated SMFs are re-scaled, whereas the simulated SMFs come right from the cosmological simulations.

5.4.2 Degeneracy broken with stellar mass functions

We have seen the bimodality of the inferred posterior density and degeneracy in SFRD in Section 5.3.1 and Figures 5.6 and 5.7. In this section, we discuss how degeneracy in the SFRD can be broken with both the emulated SMF and the simulated SMF. Figure 5.17 illustrates the marginal distributions (*bottom*) that are exactly the same marginals as in Figure 5.7,

derived by inference on the SFRD, and the corresponding *emulated* SMFs (*top left*) and the *simulated* SMFs (*top right*). In contrast to the case of the SFRD where two *emulated* SFRDs from two peaks (*red dashed* and *blue dotted*) are nearly on top of each other, two *emulated* SMFs from two peaks are appreciably separate in panel (a). The standard deviation of SMFs (*grey*) in panel (a), which is 0.081, is an order of magnitude higher than that of the SMFs from the convergent posterior in Figure 5.10 that is only 0.007. In addition, the relative error of the SMFs from peak 2 (*blue dotted*) with respect to the SMFs from peak 1 (*red dashed*) is 59%, which is far larger than 0.79% from the SFRDs in Figure 5.7. Therefore, we can conclude that the SMFs can break the degeneracy of the SFRDs in terms of the emulators.

Also, we perform new simulations with the parameter sets of the peak 1 (*red dashed*) and peak 2 (*blue dotted*) to study whether the SMFs can break the degeneracy even under the simulation uncertainty. In panel (b) of Figure 5.17, the *red dashed* and *blue dotted* lines indicate the simulated SMFs from the two peaks. Similarly to the emulated SMFs, the two simulated SMFs do not coincide with each other. However, since the simulation is affected by the simulation uncertainty, such as cosmic variance and butterfly effect (refer to Section 5.2.3), we include the confidence regions (*grey* in panel (b)) of the simulation uncertainty. We use $\pm 1\sigma$ (*grey*) and $\pm 3\sigma$ (*light grey*) regions that correspond to 68.1% and 99.7% confidence levels for the Gaussian distribution, respectively. The standard deviations of the SMF are directly calculated from the simulations in the CV set as a function of stellar mass. The confidence regions are drawn with respect to the *red dashed* curves. Panel (b) (*top right*) demonstrates that the low mass ends of the *blue dotted* SMFs noticeably fall outside the 3σ regions at $z \geq 1.5$. By the definition of 3σ regions that corresponds to 99.7% confidence level, the *blue dotted* SMFs have *only 0.3%* chance that it shares the same origin with the *red dashed* SMFs and is differed by the simulation uncertainty, namely degenerate with the *red dashed* SMFs.

Both emulated and simulated results support two main conclusions: (1) There exists degeneracy in the SFRD (refer to Section 5.3.1); (2) The SMFs can break the degeneracy in the SFRD (see Figure 5.17). Taken together with the correlation analysis in Section 5.4.1, these conclusions further support the notion that the higher the correlation between observable and parameters, the stronger the observable constrains the parameters; here, the SMF is shown to be a stronger constraint than the SFRD.

5.4.3 Uncertainty in simulation-based inference

We have briefly discussed that the mock uncertainty applied to the SMFs has negligible impact on the variance of posteriors in Section 5.3.3 unlike the case of the SFRD in Figure 5.8. To elucidate the origin and implications of these results, and what type of uncertainty should be adopted for SBI, here we perform tests of SBI using various types of uncertainties. We adopt six different uncertainties as follows: (a) *Mock uncertainty* (that which we have used throughout this paper) is modelled to mimic the simulation uncertainty from the CV set in Section 5.2.3; (b) *Uniform uncertainty* is made of the univariate Gaussian distribution. Random variables are drawn from the Gaussian and added to the SMF uniformly with respect to stellar mass, leading to overall shifts in normalization; (c) *Asymmetric Uncertainty* is simply modelled as the modulus of the mock uncertainty such that the uncertainty only goes in the positive direction.(d) For *Gaussian* $\Omega_{\rm m}$ *uncertainty*, given a set of parameters $(\Omega^0_{\rm m}, \sigma^0_8, A^0_{\rm SN1}, A^0_{\rm AGN1}, A^0_{\rm SN2}, A^0_{\rm AGN2})$, we generate the SMFs out of $(\Omega'_{\rm m},\ \sigma^0_8,\ A^0_{\rm SN1},\ A^0_{\rm AGN1},\ A^0_{\rm SN2},\ A^0_{\rm AGN2})$ where $\Omega'_{\rm m} \sim \mathcal{N}(\Omega^0_{\rm m},\ \sigma_{\Omega_{\rm m}})$. Here, $\sigma_{\Omega_{\rm m}}$ is set to 0.04; (e) We model *Gaussian* $\Omega_{\rm m}$ *uncertainty with white noise* by adding Gaussian white noise directly to the SMFs with the *Gaussian* $\Omega_{\rm m}$ uncertainty; (f) *Gaussian* $\Omega_{\rm m}$ *and* σ_8 *uncertainty* is modelled similarly to *Gaussian* $\Omega_{\rm m}$ except that we additionally vary σ_8 as $\sigma'_8 \sim \mathcal{N}(\sigma^0_8, \sigma_{\sigma_8})$ with σ_{σ_8} of 0.04. The visual description of above

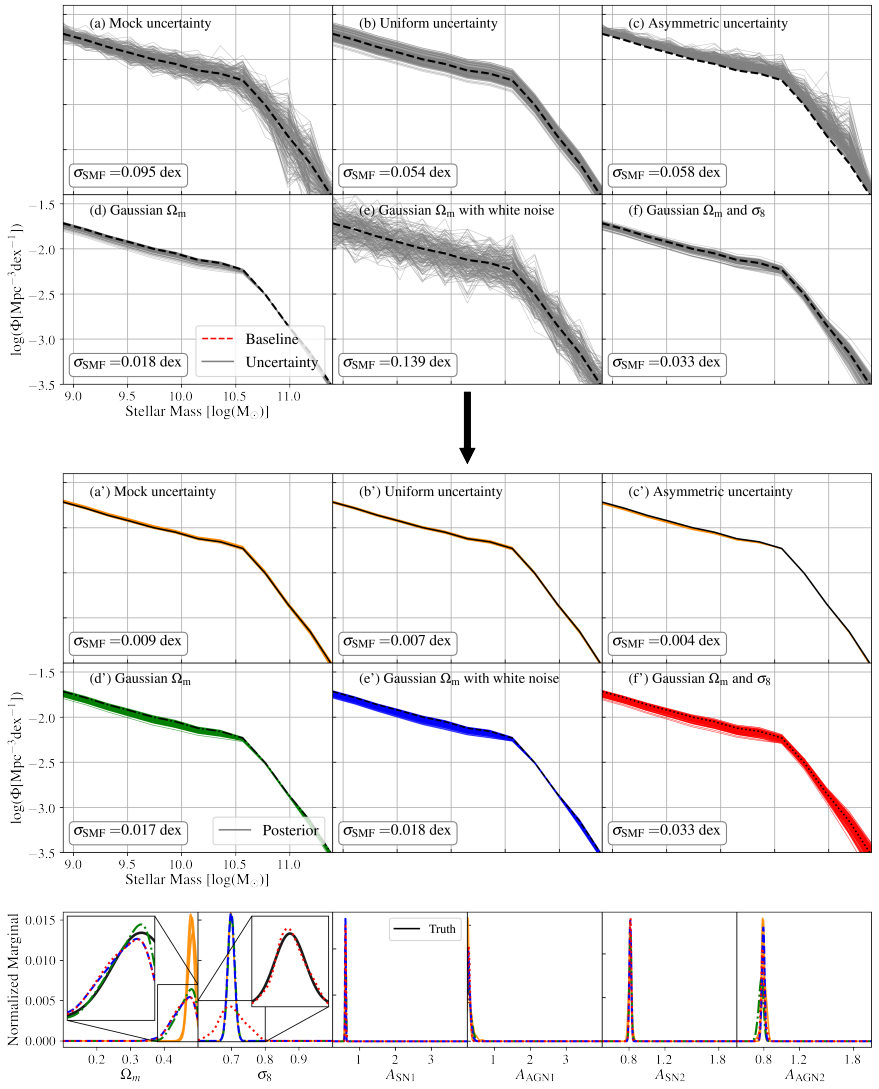

Figure 5.18: The *upper* six panels show the stellar mass functions (SMF) at $z = 0.5$ with six different types of uncertainties. The *middle* six panels show the inferred SMFs at $z = 0.5$ with six different types of uncertainties. The *lower* six panels show six marginal of the inferred posteriors from the SMFs with six different types of uncertainties.

six uncertainties is shown in the *top* panels of Figure 5.18. The standard deviations of six uncertainties are 0.095, 0.054, 0.058, 0.018, 0.139 and 0.033 dex from (a) to (f), respectively.

We perform SBI from the emulated SMF at $z = 0.5$ that is generated with the same parameters that are used in Section 5.3.3, together with the six uncertainties, respectively. Figure 5.18 shows the marginals of the inferred posteriors (*bottom*) and the corresponding SMFs (*middle*). The corresponding posterior (*middle*) inferred from each implemented uncertainty (*top*) (a, b, c, d, e, and f) in shown in the panels (a', b', c', d', e', and f'), respectively. Also, the inferred SMFs (*middle*) and marginal densities (*bottom*) are consistently color-coded as follow: *orange dotted*—(a', b', and c'), *green solid*—(d'), *blue dotdash*—(e'), and *red dashed*—(f'). We color (a', b', and c') with the same color since the implemented uncertainties (a, b, and c) share the same property that the uncertainty injected to the SMFs is *uncorrelated* with the parameters. The uncertainty is uncorrelated with parameters if there does not exist $\boldsymbol{\theta}'$ satisfying $f(\boldsymbol{\theta}') = f(\boldsymbol{\theta}) + \mathbf{Z}(\eta)$ for all θ where f is a function that takes parameters $\boldsymbol{\theta}$ as input and outputs a stellar mass function and $\mathbf{Z}(\eta)$ is an implemented uncertainty. We will discuss it further at the end of this section in detail.

Compared to the implemented uncertainties (a, b, and c), the inferred SMFs (a', b', and c') have exceedingly small variances. The standard deviations $\bar{\boldsymbol{\sigma}}_{\rm smf}$ of (a', b', and c') are dex, whereas that of (a, b, and c) are dex, which is an order of magnitude smaller in the inferred SMFs. The marginal distributions (a', b', and c') (*yellow*) are nearly on top of each other and have the standard deviations $\boldsymbol{\sigma}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$ = (0.005, 0.001, 0.006, 0.084, 0.006, 0.010), (0.007, 0.002, 0.012, 0.056, 0.008, 0.023), and (0.002, 0.001, 0.006, 0.027, 0.004, 0.011), respectively. The standard deviations of SMFs are 0.009, 0.007, and 0.004 dex (a', b', and c'), which is an order of magnitude smaller than that of the implemented uncertainty (0.095, 0.054, and 0.058 dex). Most importantly, the standard deviations of both the

SMFs and the marginals is close to that of the inference from the SMFs *without uncertainty* in Figure 5.10 of Section 5.3.3, which are $\bar{\boldsymbol{\sigma}}_{\rm smf} = 0.007$ dex and $\boldsymbol{\sigma}_{(\Omega_{\rm m},\sigma_8,A_{\rm SN1},A_{\rm AGN1},A_{\rm SN2},A_{\rm AGN2})}$=(0.006, 0.002, 0.024, 0.076, 0.006, 0.018). This can lead us to the proposition that *uncorrelated uncertainty* has merely no impact on inference.

In the case of the uncertainties directly injected to $\Omega_{\rm m}$ and σ_8, the inferred SMFs (d', f') and the implemented uncertainty (d, f) have approximately the same level of the standard deviations as $\bar{\boldsymbol{\sigma}}_{\rm smf}$ of (d, f) = (0.018, 0.033) and $\bar{\boldsymbol{\sigma}}_{\rm smf}$ of (d', f') = (0.017 0.033). On the other hand, the errors on the inferred SMFs (e') are completely different from the implemented uncertainty (e) that includes white noise in SMFs. The inferred SMFs (e') shows no sign of the white noise but are very similar to the SMFs with the implemented uncertainty (d) that has no white noise. Shown in the zoom-in panels are the analytic Gaussian distributions (*black solid*) that are used to generate the uncertainties (d, e, f). In $\Omega_{\rm m}$, the three inferred marginals—*green dotdashed, blue dashed, red dotted*—are in a relatively good agreement with the analytic lines (*black solid*). Notice that the marginal for (e') is not affected by the white noise. Due to the range limit of $\Omega_{\rm m}$, the inferred marginals are skewed such that the probability density near $\Omega_{\rm m} = 0.5$ drops sharply compared to the analytic line, whereas in the σ_8 panel, the marginal (f) (*red dotted*) precisely matches the analytic line (*black solid*).

The implemented uncertainties (d, e, f) are the *correlated* uncertainty since the uncertainty is directly injected to the parameter such that there exists $\boldsymbol{\theta}'$ satisfying $f(\boldsymbol{\theta}') = f(\boldsymbol{\theta} + \mathbf{Z}(\eta))$ for $\boldsymbol{\theta}' \in \mathbb{R}$. Meanwhile, the uncertainty (e) is the *partially correlated uncertainty* since it includes the *uncorrelated* part as well. In contrast to *uncorrelated* uncertainty (a, b, and c), the variance of *correlated* uncertainty (d and f) can successfully be propagated to the posteriors (d' and f'). Moreover, in the case of *partially correlated* uncertainty (e), the uncorrelated part is cancelled out, leaving only variance of the *correlated uncertainty* in the posterior distribution. This is a strong

evidence that *uncorrelated uncertainty* has no impact on inference and highlights the importance of well-establish mock uncertainty that can reproduce the correlation between the simulation uncertainty and the parameters. From another point of view, if the simulation uncertainty is not correlated with the parameters, performing SBI using cosmological simulations can naturally eliminate the effects of simulations. Note that it is not yet proven that the simulation uncertainty is correlated with the parameters.

5.4.4 Physical limits of simulations

Mismatch between inferred SMFs and observed SMFs

We have seen the significant mismatch between the inferred SMFs and the observed SMFs in Section 5.3.4. There are four primary issues in the SMFs inferred from the observed SMFs shown in Figures 5.13 and 5.14: (1) The population of massive galaxies ($M_\star \gtrsim 10^{10}\mathrm{M}_\odot$) in the inferred SMFs are located far below the observed SMFs across all redshifts; (2) The inferred SMFs have a 'shoulder' at $M_\star = [10^{10.5}, 10^{11}]\mathrm{M}_\odot$, which does not appear in the observations. (3) The evolution of the emulated SMFs with respect to redshift $d\phi(M, z)/dz$ in the low-mass end is smaller than that of the observed SMFs; (4) the SMF-stellar mass slope changes in time are appreciably different between simulations and observations.

The mismatch in the high mass end has been a crucial problem in the inference from the five concatenated observed SMFs in Section 5.3.4. The difference between the inferred SMFs and the observed SMFs is approximately 0.3 dex on average. Concerning physical models in cosmological simulations, there can be two physical mechanisms that can control the populations of massive galaxies: (1) large-scale structure formation and evolution and (2) astrophysical feedback. In the context of structure formation, a sufficient amount of matter (Ω_m) and large density fluctuations (σ_8) can lead to a high abundance of massive halos and galaxies. On the other hand,

it is well known that the AGN feedback is a dominant factor in quenching star formation in massive galaxies (for a seminal review, refer to Kormendy and Ho [86]), whereas the stellar feedback is more effective in low-mass galaxies. However, since the AGN feedback is ineffective as we have seen in Section 5.4.1, the dependence of the massive-galaxy population on cosmology should considerably increase. The compensation of cosmology can be found in Section 5.3.4: In short, the inferences from the individual SMFs in Figure 5.14 have recovered the population of massive galaxies by having significant larger σ_8 compared to the posterior inferred from the concatenated SMFs in Figure 5.13. That being said, the inferences cannot find a set of parameters that can reproduce the evolution of SMFs as observations.

In addition, Figure 5.14 exhibits several notably different properties for observations and emulators. The emulated SMFs tend to have a 'shoulder' at $M_\star = [10^{10.5}, 10^{11}]\mathrm{M}_\odot$, which does not appear in the observed SMFs at all. Although we can find this property in the actual hydrodynamic simulation, the re-scaling might enhance such properties (refer to Section 5.2.2). Figure 5.1 demonstrates that the re-scaled SMFs (*red*) have more notable shoulders than the SMFs without re-scaling (*red dashed*). Thus, such shoulders can be attributed to the re-scaling and/or resolution effects.

The rate of evolution of the SMFs in cosmological simulations is relatively low compared to the rate of evolution in observations. We define the rate of evolution of the SMFs as $R_{\mathrm{smf}} \equiv \Delta\bar{f}(z)/\Delta z$ where $\bar{f}(z) \equiv \int_{10^{8.9}\mathrm{M}_\odot}^{10^{9.5}\mathrm{M}_\odot} f(z, M_\star)\mathrm{d}\log(M_\star) / \int_{10^{8.9}\mathrm{M}_\odot}^{10^{9.5}\mathrm{M}_\odot} \mathrm{d}\log(M_\star)$ and $f = f(z, M_\star)$ is a stellar mass function at redshift z. For $\Delta z = 0.5 - 2.5$, $R_{\mathrm{smf,obs}} \sim -0.23$ dex per unit redshift for the observations and $R_{\mathrm{smf}} \sim -0.18$ dex per unit redshift for the inferred SMFs in Figure 5.13. We find that the average $R_{\mathrm{smf,sim}}$ over 1000 simulations in the LH set is ~ -0.12 dex/redshift. Overall, $R_{\mathrm{smf,sim}}$ is biased toward lower value than $R_{\mathrm{smf,obs}}$.

Lastly, the slope of the SMFs with respect to stellar mass at the low-mass end hardly changes across redshift in observation, whereas the slope

of simulations tends to decrease as time goes by. We define the slope as $f'(z) = \Delta f / \Delta M_\star|_z = \left(f(10^{8.9}\mathrm{M}_\odot, \mathrm{z}) - \mathrm{f}(10^{9.5}\mathrm{M}_\odot, \mathrm{z})\right) / \left(8.9 - 9.5\ \log(\mathrm{M}_\odot)\right)$ where f is a stellar mass function. The changes of the slope with respect to redshift can be written as $\Delta f' = f'(z = 0.5) - f'(z = 2.5)$. Then, $\Delta f'_{\mathrm{obs}} = 0.031$ whereas $\Delta f'_{\mathrm{sim}} = 0.241$. In addition to the discrepancy between $R_{\mathrm{smf,sim}}$ and $R_{\mathrm{smf,obs}}$, the difference in $\Delta f'$ can be strong evidence for the limits of physical models in hydrodynamic simulations and/or the limited dimensions of parameter space. We anticipate that this problem can be at least partially alleviated by introducing additional dimensions in the parameter space.

An extra parameter: star formation timescale

In this section, we investigate whether the parameters can be successfully inferred from a simulated observable that comes from higher-dimensional parameter space that has an *extra* dimension. Thus far, we have performed SBI on six parameters from observables—SFRD or SMFs—that are obtained by a specific set of six parameters, or actual observations. Here, we perform SBI on the same six parameters from SMFs that are simulated from the fiducal six parameters while varying the extra subgrid model parameter, star formation timescale. The two simulations are performed with longer and shorter star formation timescale than the fiducial run (hereafter, LT and ST denote the longer and shorter star formation timescale simulations, respectively). In LT snd ST, the same random seed as the fiducial run is adopted to minimize the effect of simulation uncertainty through cosmic variance. Lastly, to generate emulator counterpart of SMFs for LT (ST), we calculate the ratios of the SMFs from LT (ST) to the simulated fiducial SMFs, and then apply them to the emulated SMFs of the fiducial parameters to obtain emulated SMFs for LT (ST).

Figure 5.19 shows the marginal posterior densities and SMFs inferred from the emulated SMFs from LT and ST (*top* and *middle* panels, respec-

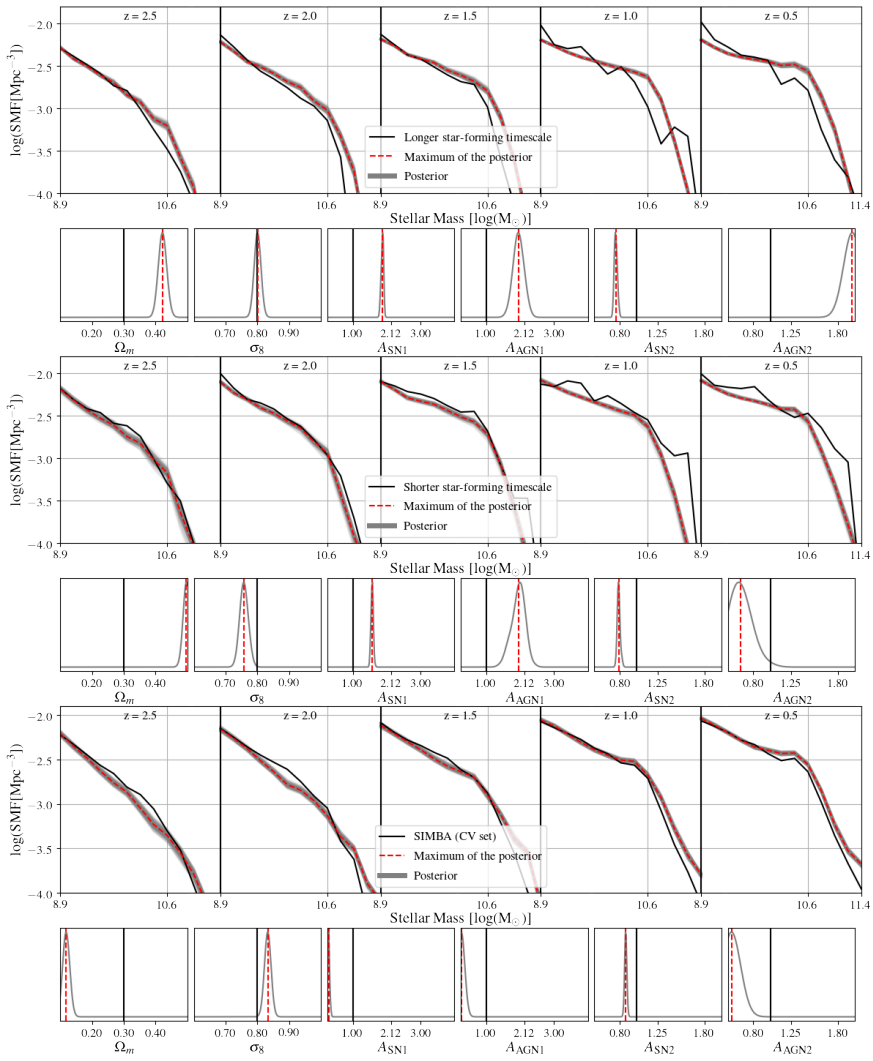

Figure 5.19: *Top*: Stellar Mass functions and the marginal posterior densities that are inferred from the SMFs with the fiducial six parameters and a longer star-formation timescale. *Middle*: Same, but with a shorter star-formation timescale. *Bottom*: Stellar Mass functions and the marginal posterior densities that are inferred from the average SMFs out of the CV set of the SIMBA, rather than IllustrisTNG, suite.

tively). In the case of LT, the peaks of the marginal densities of the inferred posteriors (*red dashed*) considerably deviate from the fiducial parameters (*black solid*), except for σ_8. The low-mass end of the inferred SMFs (*red dashd*) at $z = 2.5, 2.0,$ and 1.5 is in a relatively good agreement with the SMFs from LT (*black solid*), whereas at $z = 1.0$ and 0.5, the inferred SMFs and the SMFs of LT show a complete mismatch. In addition, the LT has a high evolutionary rate of the SMFs and tends to have steeper slopes as redshift decreases, which the inferred SMFs fail to match as in the inference from observation (see Figure 5.13). Similar to LT, the inferred SMFs are in good agreement with the SMFs of ST at $z = 2.5, 2.0$ and 1.5, whereas the high-mass end of the inferred SMFs at low redshifts is clearly lacking compared to the SMFs of ST, which we have also seen in the inference from the observation in Figure 5.13.

We have found that the SMFs from the seven-dimensional parameter space including an extra parameter—star formation timescale—are not necessarily reproduced by points in the conventional six-dimensional parameter space that we have used throughout this work. LT and ST have intrinsic properties that cannot be reproduced from points in the six-dimensional parameter space (e.g., higher rates of evolution in the SMFs, slope evolution in the SMFs, and large populations of massive galaxies). The results indicate that the introduction of extra dimensions or parameters has the potential to resolve the problem of the significant mismatch between the five concatenated observed SMFs and observation.

Inference from SIMBA

In addition to the extra parameter, we also perform SBI from the SMFs that are obtained by the fiducial simulation from the CV set of the SIMBA suite (refer to Section 5.2.1 or Villaescusa-Navarro et al. [185]). Likewise, we re-scale the SIMBA SMFs to obtain the emulated SIBMA SMFs in the same way as we have in the previous part. Note that although we make

inference from a SIMBA-derived observable, we use the same emulators that we have used so far, which are trained on the TNG suite of the CAMELS simulations.

The *bottom* row of Figure 5.19 illustrates the inferred SMFs (*red dashed* and *grey*) and the emulated SIMBA SMFs (*black solid*). The inferred SMFs from the maximum of the posteriors (*red dashed*) seem to match the SIMBA SMFs relatively well even though the relative error is $\sim 21\%$. Unlike LT and ST in the previous part, the inferred SMFs reasonably follow the trends of the SMFs, such as slopes and rates of the evolution of SMFs. Still, the difference between inferred parameters and fiducial parameters has not been narrowed. Such deviations between the inferred and fiducial parameters can be attributed to the differences of physical model between TNG and SIMBA.

5.4.5 Caveats & physical interpretation of inference

This section discusses how the inferred parameters can be interpreted in an astrophysical sense. We defer the physical interpretation of the inference to the last section because the inferred parameters contain very little meaningful physics at this time. Prior to the physical interpretation, it is imperative to understand the key factors that influence inferences on parameters: e.g. emulator, simulation uncertainty, resolution convergence, and limited parameter space.

First, we have employed the emulators for computational efficiency, paying the price of discrepancies between simulations and emulators. The discrepancies inevitably propagate to the inferred posterior and lead to deviations from the posterior that would have been inferred if actual cosmological simulations were employed in the inference. Assuming that we can replace the emulators with actual cosmological simulations, the next questions shall be: "Will the posterior inferred from the actual cosmological simulations contain robust information on physics?" Or, more specifically, "Will

the cosmological parameters inferred from the cosmological simulations, not the emulators, be physical and comparable to the pre-existing estimations?" To answer these questions, resolution convergence and the limited dimension of parameter space, along with simulation uncertainty, should be taken into consideration.

Resolution effect

The resolution effects in hydrodynamic simulations have been discussed in many papers [6, 24, 69, 99, 100, 117, 136, 138, 145, 161, and also see Section 5.2.2 of this chapter]. In general, it is believed that the hydrodynamic simulations are sensitive to (spatial and mass) resolution. On the other hand, pure hydrodynamic simulations themselves without subgrid models, such as radiative cooling, are highly likely to be convergent at some level (refer to Hubber et al. [69] for resolution convergence of both SPH and AMR codes). This implicitly indicates that the resolution effect is likely to be attributed to response of subgrid models to the different resolutions. In addition, cosmological hydrodynamic simulations with subgrid models can become convergent once it reaches a certain resolution and beyond [24, 99, 100]. The above arguments suggest that resolution effects can generally be incorporated into the subgrid models—astrophysical parameters.

We turn our attention to the resolution effect in this section. As discussed in Section 5.2.2, we re-scale the obserables—SFRD and SMF—to minimize the the resolution effect according to Pillepich et al. [136]. However, aside from the inaccuracy of re-scaling, there is a problem with re-scaling. We find that the resolution effect has a significant dependence on the position in parameter space, which is in line with the above arguments. In principle, the re-scaling must be a function of parameters $\boldsymbol{\theta}$, whereas we construct the re-scaling function only based on the fiducial parameters in this chapter. Due to computational cost, it is not realistically possible to obtain the full re-scaling relation as a function of parameters by running

high-resolution simulations over the entire parameter space [66]. Hence, the inference results near the fiducial parameters can be physically comparable to TNG100-1, which is the target resolution of the re-scaling, but other regions of the parameter space may be difficult to grasp.

Limited parameter space

In this work, we have seen the limits of the six-dimensional model parameter space: e.g., weak correlation between AGN parameters and observables, failure of inference from the observed SMF, and potential of the extended parameter space. The limits of parameter space can affect not only the performance or accuracy of inference but also the physical interpretation of the inferred posterior. In addition, the limitations of the subgrid model and the limited dimensions of the parameter space can lead to overfitting or over-fine-tuning of the parameters to the target observable, even if the inferred physics is not realistic. For instance, since our emulators (or simulations) are insensitive to the AGN feedback, the cosmological and stellar feedback parameters should be adjusted to control the population of the high-mass end in Section 5.3.4. The parameter space of the TNG universe or the SIMBA universe is likely to be insufficient to describe the real universe. Thus, we will ultimately move towards extending the parameter space beyond the current TNG universe so that we can resolve the problems of inference failure and secure the physical meaning of the posteriors.

5.5 Summary

In order to calibrate cosmological simulations against observations, we have employed simulation-based inference (SBI) that enables rigorous Bayesian inference in a computationally efficient way by adopting neural density estimators (NDE) that evaluate the likelihood instead of an explicit analytic likelihood in conventional Bayesian inference (Sections 3.2 and 5.2.4). In addition, for computational efficiency, we have adopted emulators

that are trained on $\sim$ 1000 cosmological simulations from the CAMELS project (specifically, those based on the IllustrisTNG framework) to predict simulated observables, taking as input the cosmological and astrophysical parameters, and used these emulators as surrogates to the cosmological simulations (Section 5.2.5). Using the emulators, we have conducted SBI on the cosmological and astrophysical parameters ($\Omega_{\rm m}$, σ_8, stellar wind feedback, and kinetic black hole feedback) from the cosmic star formation rate density (SFRD) and stellar mass functions at different redshifts (SMFs) and retrieved 6-dimensional posterior distributions of the parameters.

We summarize our results as follows:

- The posteriors inferred from emulated SFRD and SMFs converge to the true values well with relative errors of less than 1% in either SFRD or SMFs (Sections 5.3.1 and 5.3.3, respectively). However, the SFRD requires an order of magnitude more training data to converge than the SMFs do, having two convergent stages—an unstably convergent stage and a stably convergent stage (Figure 5.5).

- In the unstably convergent stage, there exists a bimodal posterior distribution that has two *degenerate* peaks (see 5.6). However, the *degeneracy* in the SFRD, which is also confirmed with new cosmological simulations (namely, it is not an artefact of the emulator), is broken with the SMFs in both the emulator and cosmological simulations (Section 5.4.2). This indicates that the SMFs provide stronger constraints for the parameters.

- In the inferences with the mock uncertainty that we add to the emulators to mimic the simulation uncertainty (Section 5.2.3), the posterior inferred from the emulated SFRD has successfully captured the variance of the mock uncertainty in the parameters, whereas the SMFs cannot capture any variances in the posterior. In Section 5.4.3, we

show that uncorrelated uncertainties have a negligible impact on SBI. Future work will require to build a robust model that can precisely reproduce the simulation uncertainty (note that the actual simulation uncertainty is also uncorrelated with the parameters). In addition, an emulator or theoretical model that can marginalize the simulation uncertainty will be another direction to future work.

- Employing SBI on the Leja et al. [93] observed SFRD, we find a set of parameters that matches the target observable with a relatively high accuracy, whereas the similarly inferred SMFs show significant discrepancies with the target observed SMFs (Section 5.3.4). Such discrepancies could potentially originate from the intrinsic physical limits of cosmological simulations with the TNG framework (Section 5.4.4). Resolving this issue will require extending the parameter space and/or updating the subgrid models of cosmological simulations.

- Using mutual information, we measure the correlation between parameter-observable pairs and we find that the performance of inference for each observable largely depends on the correlation between the parameters and observable (Section 5.4.1).

- In both correlations and inferences, we find that the AGN parameters (black hole kinetic feedback) are most weakly correlated with both SFRD and SMFs. This can be attributed to the relatively insignificant impact of the black hole kinetic feedback parameters, within the range varied in CAMELS, on both formation of galaxies with high stellar mass and star formation in massive galaxies compared to cosmological parameters and stellar wind feedback (see Figure 5.16).

- In this work, we have refrained from conducting physical interpretation of the inferred parameters because the inference result is sensitive to emulator accuracy, resolution effects, simulation uncertainty,

and inaccuracy from limited parameter space (Sections 5.4.5).

- This work is only a cornerstone of calibrating cosmological simulation against observation and provides considerable insights into future directions. In future work, we will focus on resolution convergence, simulation uncertainty, the extension of parameter space as well as the number of target observables, and inference (only) with cosmological simulations without emulator bias.

5.6 Appendix

5.6.1 Definition of degeneracy

We have encountered degeneracy in Section 5.3.1 that discusses the bimodality of the SFRD. In this section, we probabilistically discuss the relation between uncertainty and degeneracy and mathematically define degeneracy for our purposes. In terms of parameter-observable pair, the degeneracy originates from *indistinguishability* among these pairs. Given an arbitrary observation $\boldsymbol{x}_0$, if there exists a set of parameters $\boldsymbol{\Theta}_{\rm degen}$ such that $\boldsymbol{f}(\boldsymbol{\theta})|_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\rm degen}} = \boldsymbol{x}_0$ where $\boldsymbol{f}(\boldsymbol{\theta})$ is e.g. a theoretical model, a simulation, and a fast approximation method that predict observable $\boldsymbol{x}$ as a function of parameters $\boldsymbol{\theta}$, then one *cannot pinpoint* the parameters $\boldsymbol{\theta}$ from which the given observation $\boldsymbol{x}_0$ comes. In this case, the pairs $\{(\boldsymbol{\theta}, \boldsymbol{x}_0)|\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\rm degen}\}$ are said to be degenerate with respect to observation $\boldsymbol{x}_0$. This can be usually seen in quantum systems such as the spin triplet state under no magnetic field. In consideration of an arbitrary uncertainty $\boldsymbol{Z}(\eta)$ in observable[13] , the pairs can be written in either $(\boldsymbol{\theta}, \boldsymbol{x} + \boldsymbol{Z}(\eta))$ or $(\boldsymbol{\theta} + \boldsymbol{\epsilon}(\eta), \boldsymbol{x})$ where η is a random seed (see Section 5.2.3). One can notice that $\boldsymbol{Z}$ and $\boldsymbol{\epsilon}$ are basically random variables that require a probabilistic treatment. In the following section, degeneracy will be discussed in a probabilistic manner.

The ideal probabilistic inference naturally traces the propagation of uncertainty in observation onto each parameter. Nevertheless, the consistent, robust confinement for the posterior density is necessary to define a finite region of the degenerate parameter space since an arbitrary inferred posterior density is generally well-defined over the entire parameter space.

We define degeneracy as follows.
Given an arbitrary probability density $p(\boldsymbol{\theta})$ and $\zeta_{\rm thres}$, there exist $p_{\rm thres}$ and

[13] For example, the cosmological simulations $\boldsymbol{f}$ reproduce different output $\boldsymbol{x}$ depending on the initial conditions η even with the same set of physical and free parameters $\boldsymbol{\theta}$. i.e. $\boldsymbol{f}(\boldsymbol{\theta}) = \boldsymbol{x} + \mathbf{Z}(\eta)$ where $Z(\eta)$ stands for cosmic variance.

$\boldsymbol{\Theta}_{\text{degen}}$ satisfying $\boldsymbol{\Theta}_{\text{degen}} = \{\boldsymbol{\theta} \,|\, p(\boldsymbol{\theta}) \geq p_{\text{thres}}\}$ such that

$$\text{(continuous)} \qquad \frac{\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\text{degen}}} p(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}\in\mathcal{V}} p(\boldsymbol{\theta}) d\boldsymbol{\theta}} = \zeta_{\text{thres}} \tag{5.5}$$

or

$$\text{(discrete)} \qquad \frac{\sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\text{degen}}} p(\boldsymbol{\theta}) \Delta\boldsymbol{\theta}}{\sum_{\boldsymbol{\theta}\in\mathcal{V}} p(\boldsymbol{\theta}) \Delta\boldsymbol{\theta}} = \zeta_{\text{thres}} \tag{5.6}$$

where $0 < \zeta_{\text{thres}} < 1$ and $\mathcal{V}$ stands for the entire parameter space. Here, the parameters $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}_{\text{degen}}$ are said to be degenerate. Note that the threshold value ζ_{thres} is a free parameter.

In case of a high dimensional problem with an intractable probability distribution, it is practically impossible to numerically integrate the probability distribution $p(\boldsymbol{\theta})$ over $\mathcal{V}$ or $\boldsymbol{\Theta}_{\text{degen}}$ even if $\boldsymbol{\Theta}_{\text{degen}}$ is known. In low dimensional problems, simple quadrature methods can work. For example, we can evaluate a function on a fixed grid of points, then apply the trapezoid rule. However, in high dimensions, the number of grid points grows exponentially. [14] Hence, MCMC methods are widely adopted to integrate high dimensional functions.

We utilize the MCMC sampling to estimate the degenerate parameter space. The MCMC provides millions of parameter sets such that the ratio of the number of parameter sets in each bin, $N(\Delta\boldsymbol{\theta}_{\text{bin}})$, to the total number of the samplings, N_{total}, represents the approximate probability of that bin. In other words, $p(\boldsymbol{\theta})\Delta\boldsymbol{\theta}_{\text{bin}} \simeq N(\Delta\boldsymbol{\theta}_{\text{bin}})/N_{total}$ where $\Delta\boldsymbol{\theta}_{\text{bin}}$ is the size of each bin. We can rewrite the left-hand side of Eq. 5.6 in the form

$$\frac{\sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\text{degen}}} p(\boldsymbol{\theta}) \Delta\boldsymbol{\theta}}{\sum_{\boldsymbol{\theta}\in\mathcal{V}} p(\boldsymbol{\theta}) \Delta\boldsymbol{\theta}} = \frac{\sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\text{degen}}} N(\Delta\boldsymbol{\theta}_{\text{bin}})}{N_{\text{total}}}. \tag{5.7}$$

[14] In our case, six dimensional parameter space requires $(10^2)^6 = 10^{12}$ grid points where we space each axis with 100 grid points, which means that 10^{12} arithmetic calculations at least are needed.

Then,

$$\sum_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{\rm degen}} N(\Delta\boldsymbol{\theta}_{\rm bin}) = \zeta_{\rm thres} N_{\rm total}. \tag{5.8}$$

Together with $\boldsymbol{\Theta}_{\rm degen} = \{\boldsymbol{\theta} \,|\, p(\boldsymbol{\theta}) \geq p_{\rm thres}\}$, we can write $\boldsymbol{\Theta}_{\rm degen} = _{D\in\mathcal{D}} p_{\rm thres}; \mathcal{D} = \{D | \tilde{p}(\boldsymbol{\theta}) > p_{\rm thres} \forall \boldsymbol{\theta}; \boldsymbol{\theta} \in D, n(D) = \zeta_{\rm thres} N_{\rm total}, D \subset \boldsymbol{\Theta}_{\rm MCMC}\}$ where $\boldsymbol{\Theta}_{\rm MCMC}$, $\tilde{p}$ and $n(\cdot)$ are the set of sampled parameters from MCMC, an arbitrary surrogate posterior function, such as the NDE or kernel density estimation, and the cardinality of a set. The set of degenerate points in parameter space, $\boldsymbol{\Theta}_{\rm degen}$, is a subset of $\boldsymbol{\Theta}_{\rm MCMC}$ such that the sum of probability of the parameters in the subset is equal to $\zeta_{\rm thres}$ (refer to Eq. 5.6). In practice, we identify $\boldsymbol{\Theta}_{\rm degen}$ among $\boldsymbol{\Theta}_{\rm MCMC}$ as follows: (1) Estimate $\tilde{p}(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_{\rm MCMC}$; (2) Pair $\boldsymbol{\theta}$ and $\tilde{p}(\boldsymbol{\theta})$ into $(\boldsymbol{\theta}, \tilde{p}(\boldsymbol{\theta}))$; (3) Sort $(\boldsymbol{\theta}, \tilde{p}(\boldsymbol{\theta}))$ in a descending order along $\tilde{p}(\boldsymbol{\theta})$; (4) Identify the first $\zeta_{\rm thres} N_{\rm total}$ number of $\boldsymbol{\theta}$ of the sorted pairs as $\boldsymbol{\Theta}_{\rm degen}$.

5.6.2 Marginalization of simulation uncertainty in emulator

This section discusses how emulators marginalize the simulation uncertainty in the LH set that the emulators are trained on. The LH set, consisting of 1000 cosmological simulations, suffers from the simulation uncertainty since the initial conditions of the 1000 simulations are all different, namely it is dominated by cosmic variance, since the scale of our box size is $25\,{\rm Mpc}/h$. The simulation uncertainty plays a role as an intrinsic noise in the simulated data in training the emulators, which potentially degrades the accuracy of emulators. Thus, we study whether the emulators can marginalize the noise during training or how much the emulator suffers from the simulation uncertainty. If the emulators can marginalize the simulation uncertainty completely, then the emulator prediction should be equivalent to the mean of the results of a hypothetical suite with a large number of

simulations for each point in parameter space—hereafter the uncertainty-marginalized ideal simulation.

We first describe the simulated observable g (SFRD or SMF) as follows:

$$g = g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta})), \tag{5.9}$$

where $\boldsymbol{\theta}$ is a set of parameters and $\delta(\lambda, \boldsymbol{\theta})$ describes the initial conditions with a random seed λ for cosmological simulations. Here, the marginalization of simulation uncertainty—i.e., the emulator prediction is equivalent to the uncertainty-marginalized ideal simulation—can be written as $\bar{g}(\boldsymbol{\theta}) = \langle g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta})) \rangle_\lambda$ by averaging g over the sources of randomness λ. The physical analogy of the mean of the simulation uncertainty is the mean of cosmic variance by performing the infinite-volume simulation (only if we ignore the effects of the long-wave limit in power spectrum in cosmic variance).

The emulator prediction f can be written with respect to the ideally-marginalized prediction—the uncertainty-marginalized ideal simulation—in the form

$$f = f(\boldsymbol{\theta}) = \bar{g}(\boldsymbol{\theta}) + \epsilon(\boldsymbol{\theta}) + m(\boldsymbol{\theta}). \tag{5.10}$$

Here, we separate the inaccuracy of the emulator into the training error $\epsilon(\boldsymbol{\theta})$ and the marginalization error $m(\boldsymbol{\theta})$. Then, the bias of the emulator with respect to simulations in the LH set, $b_{\rm LH}$, can be written as

$$\begin{aligned} b_{\rm LH} &= \langle g(\boldsymbol{\theta}, \delta(\lambda_{\boldsymbol{\theta}}, \boldsymbol{\theta})) - f(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_{\rm LH}} \\ &= \langle g(\boldsymbol{\theta}, \delta(\lambda_{\boldsymbol{\theta}}, \boldsymbol{\theta})) \rangle_{\boldsymbol{\theta}_{\rm LH}} - \langle \bar{g}(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_{\rm LH}} - \langle \epsilon(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_{\rm LH}} - \langle m(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}_{\rm LH}}, \end{aligned} \tag{5.11}$$

where $\boldsymbol{\theta}_{\rm LH}$ denotes $\boldsymbol{\theta} \in \Theta_{\rm LH}$ and $\Theta_{\rm LH}$ is a set of 1000 parameters in the LH set. Here, we write λ as λ_θ because in the LH set, λ are already determined depending on $\boldsymbol{\theta}$. Assuming that the average of simulations over the

LH set can average not only the simulations from different parameters but also marginalize their simulation uncertainties such that it can be approximately equal to the average of the uncertainty-marginalized ideal simulations over the LH set—i.e., $\langle g(\boldsymbol{\theta}, \delta(\lambda_{\boldsymbol{\theta}}, \boldsymbol{\theta}))\rangle_{\boldsymbol{\theta}_{\rm LH}} \sim \langle\langle g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta}))\rangle_{\lambda}\rangle_{\boldsymbol{\theta}_{\rm LH}} \equiv \langle \bar{g}(\boldsymbol{\theta})\rangle_{\boldsymbol{\theta}_{\rm LH}}$, we obtain

$$b_{\rm LH} \simeq -\langle \epsilon(\boldsymbol{\theta})\rangle_{\boldsymbol{\theta}_{\rm LH}} - \langle m(\boldsymbol{\theta})\rangle_{\boldsymbol{\theta}_{\rm LH}} \,. \tag{5.12}$$

The empirical biases $b_{\rm LH}$ for the SFRD and SMFs are 0.0026 dex and -0.0014 dex, respectively. Compared to the standard deviations of the simulation uncertainty for the SFRD and SMFs (~ 0.06 dex and ~ 0.1 dex), the emulators have relatively small bias, which indicates that the mean of emulators and the mean of simulations are in a good agreement.

Secondly, we estimate the variance of emulator prediction with respect to the uncertainty-marginalized ideal simulation in terms of the LH set as follows:

$$\begin{aligned}\hat{\sigma}^2_{\rm LH} &= \left\langle \left(\bar{g}(\boldsymbol{\theta}) - f(\boldsymbol{\theta})\right)^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} \\ &= \left\langle \left(\bar{g}(\boldsymbol{\theta}) - g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta})) + g(\boldsymbol{\theta}, \delta(\lambda, \boldsymbol{\theta})) - f(\boldsymbol{\theta})\right)^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} .\end{aligned} \tag{5.13}$$

With the definitions $A \equiv \bar{g}(\boldsymbol{\theta}) - g(\boldsymbol{\theta}, \delta(\lambda_{\boldsymbol{\theta}}, \boldsymbol{\theta}))$, $B \equiv g(\boldsymbol{\theta}, \delta(\lambda_{\boldsymbol{\theta}}, \boldsymbol{\theta})) - f(\boldsymbol{\theta})$ and $C \equiv \bar{g}(\boldsymbol{\theta}) - f(\boldsymbol{\theta})$, we obtain

$$\begin{aligned}\tilde{\sigma}^2_{\rm LH} = \left\langle C^2 \right\rangle_{\boldsymbol{\theta}\rm LH} &= \left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + \left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + 2\left\langle AB \right\rangle_{\boldsymbol{\theta}_{\rm LH}} \\ &= \left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + \left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + 2\left\langle A(C-A) \right\rangle_{\boldsymbol{\theta}_{\rm LH}} \\ &= -\left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + \left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + 2\left\langle AC \right\rangle_{\boldsymbol{\theta}_{\rm LH}} \\ &\simeq -\left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} + \left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}} .\end{aligned} \tag{5.14}$$

In the third line, $\langle AC\rangle_{\boldsymbol{\theta}_{\rm LH}}$ is the covariance of A and C where A and C are the simulation uncertainty and the deviations of the emulator from the

ideally-marginalized prediction, respectively. Since the error of emulator and simulation uncertainty are, in principle, fully independent, we assume that $\langle AC \rangle_{\boldsymbol{\theta}_{\rm LH}}$ approximately vanishes. $\left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}}$ is the variance of simulation uncertainty and $\left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}}$ is the variance of emulator with respect to simulation in the LH set, both of which are measurable quantities. $\left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}}$ and $\left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}}$ for the SFRD are 0.036 dex^2 and 0.004 dex^2, and $\left\langle A^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}}$ and $\left\langle B^2 \right\rangle_{\boldsymbol{\theta}_{\rm LH}}$ for the SMFs are 0.369 dex^2 and 0.035 dex^2. Thus, $\hat{\sigma}^2_{\rm LH}$ for the SFRD and SMFs are 0.032 dex^2 and 0.334 dex^2, respectively. As a result, the emulators have negligibly small biases but are deviated from the mean of the simulation uncertainty at a similar level to the variance of the simulation uncertainty. Therefore, we can conclude that the emulators cannot properly marginalize the simulation uncertainty.

Chapter 6

Conclusion

In the era of technology, the development of computers and telescopes has led to remarkable successes in both numerical simulations and observations, and greatly improved our understanding of astrophysical and cosmological phenomena, from formation of the first star to the large-scale structure. However, simulations and observations have been sundered and studied almost independently due to several factors, such as various uncertainties in both observation and simulation, limitations of physical models for both observation and simulation, and the absence of delicate and sophisticated comparisons. In this dissertation, we harness the power of machine learning to overcome the computational limitations of numerical simulations and extend our understanding of formation and evolution of galaxy in a cosmological context.

In the first part of the dissertation (Chapter 4), we build a machine-learning model that estimates the baryonic properties of halos, taking as input dark-matter halo properties without running actual hydrodynamic simulations. In doing so, we improve our model by adopting several methods: inclusion of historical and environmental properties of halos, application of logarithmic scaling, and use of two-stage learning. Machine accuracy has appreciably been improved as follows. The application of logarithmic scaling alleviates the problem of disproportionality in output features such as stellar mass so that we can have consistent accuracy across mass scale. The inclusion of historical and environmental properties leads to higher correlation between input and output features by providing more physical information to the machine. The two-stage learning helps reduce the complexity of the machine (i.e., structure of decision trees), leading to increase in both

training efficiency and accuracy (refer to Sections 4.2.3 and 4.3.1). By applying our improved model to larger-scale dark matter-only simulations, we can address topics like baryonic acoustic oscillations, with galaxies having basic properties, which is usually unavailable with smaller-size cosmological hydrodynamic simulations that the model is trained on.

One of the main benefits of our model is that we can transplant the baryon physics of galaxy-scale hydrodynamic simulations onto a much larger volume of dark matter-only simulations, only within a fraction of time needed for a hydrodynamic simulation. Therefore, we can obtain a large-scale galactic catalog which is not feasible to obtain using hydrodynamic simulations. In addition, the ERT algorithm that we adopt assesses the relative importances of input features in estimating each baryonic properties (Section 4.4). This has a full of potential to study the complex relations between dark matter and baryon properties.

In Chapter 5, we have studied how we can bridge the gap between observations and simulations by addressing practical computational-efficiency problems using machine learning. For computational efficiency, we trained emulators on cosmological simulation to predict observables, taking as input the simulation parameters—cosmological and astrophysical parameters. Using the emulators, we have performed simulation-based inference on the cosmological and astrophysical parameters ($\Omega_{\rm m}$, σ_8, stellar wind feedback, and kinetic black hole feedback) from the cosmic star formation rate density (SFRD) and stellar mass functions at different redshifts (SMFs), and retrieved 6-dimensional posterior distributions of the parameters. In a performance test, the posterior distributions successfully recover both SFRD and SMFs with relative errors of less than 1%. We fine that the SFRD requires an order of magnitude more training data to converge than the SMFs do due to the weak correlation between the SFRD and parameters. In the inference from the emulated SFRD, there exists a bimodal posterior distribution that has two *degenerate* peaks. However, the bimodality is broken

with the SMFs, which indicates that the SMFs provide stronger constraints for the parameters.

In the inference from observations, we successfully inferred a set of parameters from the target observed SFRD with a relatively high accuracy, whereas the inferred SMFs show significant discrepancies with the target observed SMFs. These discrepancies can stem from the intrinsic physical limitations of cosmological simulations, such as subgrid models, resolutions and dimensionality of parameters. We might resolve discrepancies by extending the parameter space, updating the subgrid models, and/or resolution of cosmological simulations. In addition, physical interpretation of the inferred parameters is highly sensitive to emulator accuracy, resolution effects, simulation uncertainty, and inaccuracy from limited parameter space. Thus, clearly there is a room for improvements in the future. Nevertheless, as a cornerstone, this shows the full potential of calibrating cosmological simulations and improving our understanding of galaxy formation and evolution in a cosmological context.

Bibliography

[1] S. Agarwal, R. Davé, and B. A. Bassett. Painting galaxies into dark matter haloes using machine learning. , 478:3410–3422, Aug 2018. doi: 10.1093/mnras/sty1169.

[2] H. Aihara, N. Arimoto, R. Armstrong, S. Arnouts, N. A. Bahcall, S. Bickerton, J. Bosch, K. Bundy, P. L. Capak, J. H. H. Chan, M. Chiba, J. Coupon, E. Egami, M. Enoki, F. Finet, H. Fujimori, S. Fujimoto, H. Furusawa, J. Furusawa, T. Goto, A. Goulding, J. P. Greco, J. E. Greene, J. E. Gunn, T. Hamana, Y. Harikane, Y. Hashimoto, T. Hattori, M. Hayashi, Y. Hayashi, K. G. Hełminiak, R. Higuchi, C. Hikage, P. T. P. Ho, B.-C. Hsieh, K. Huang, S. Huang, H. Ikeda, M. Imanishi, A. K. Inoue, K. Iwasawa, I. Iwata, A. T. Jaelani, H.-Y. Jian, Y. Kamata, H. Karoji, N. Kashikawa, N. Katayama, S. Kawanomoto, I. Kayo, J. Koda, M. Koike, T. Kojima, Y. Komiyama, A. Konno, S. Koshida, Y. Koyama, H. Kusakabe, A. Leauthaud, C.-H. Lee, L. Lin, Y.-T. Lin, R. H. Lupton, R. Mandelbaum, Y. Matsuoka, E. Medezinski, S. Mineo, S. Miyama, H. Miyatake, S. Miyazaki, R. Momose, A. More, S. More, Y. Moritani, T. J. Moriya, T. Morokuma, S. Mukae, R. Murata, H. Murayama, T. Nagao, F. Nakata, M. Niida, H. Niikura, A. J. Nishizawa, Y. Obuchi, M. Oguri, Y. Oishi, N. Okabe, S. Okamoto, Y. Okura, Y. Ono, M. Onodera, M. Onoue, K. Osato, M. Ouchi, P. A. Price, T.-S. Pyo, M. Sako, M. Sawicki, T. Shibuya, K. Shimasaku, A. Shimono, M. Shirasaki, J. D. Silverman, M. Simet, J. Speagle, D. N. Spergel, M. A. Strauss, Y. Sugahara, N. Sugiyama, Y. Suto, S. H. Suyu, N. Suzuki, P. J. Tait, M. Takada, T. Takata, N. Tamura, M. M. Tanaka, M. Tanaka, M. Tanaka, Y. Tanaka, T. Terai, Y. Terashima, Y. Toba, N. Tominaga, J. Toshikawa, E. L. Turner, T. Uchida, H. Uchiyama, K. Umetsu, F. Uraguchi, Y. Urata, T. Usuda, Y. Utsumi, S.-Y. Wang, W.-H. Wang, K. C. Wong, K. Yabe, Y. Yamada, H. Yamanoi, N. Yasuda,

S. Yeh, A. Yonehara, and S. Yuma. The Hyper Suprime-Cam SSP Survey: Overview and survey design. , 70:S4, Jan. 2018. doi: 10.1093/pasj/psx066.

[3] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv e-prints*, art. arXiv:1907.10902, July 2019.

[4] J. Alsing, B. D. Wandelt, and S. M. Feeney. Optimal proposals for Approximate Bayesian Computation. *arXiv e-prints*, art. arXiv:1808.06040, Aug. 2018.

[5] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. , 488(3):4440–4458, Sept. 2019. doi: 10.1093/mnras/stz1960.

[6] D. Anglés-Alcázar, R. Davé, F. Özel, and B. D. Oppenheimer. Cosmological Zoom Simulations of z = 2 Galaxies: The Impact of Galactic Outflows. , 782(2):84, Feb. 2014. doi: 10.1088/0004-637X/782/2/84.

[7] R. E. Angulo, V. Springel, S. D. M. White, A. Jenkins, C. M. Baugh, and C. S. Frenk. Scaling relations for galaxy clusters in the Millennium-XXL simulation. , 426(3):2046–2062, Nov 2012. doi: 10.1111/j.1365-2966.2012.21830.x.

[8] I. K. Baldry, S. P. Driver, J. Loveday, E. N. Taylor, L. S. Kelvin, J. Liske, P. Norberg, A. S. G. Robotham, S. Brough, A. M. Hopkins, S. P. Bamford, J. A. Peacock, J. Bland-Hawthorn, C. J. Conselice, S. M. Croom, D. H. Jones, H. R. Parkinson, C. C. Popescu, M. Prescott, R. G. Sharp, and R. J. Tuffs. Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at z ¡ 0.06. , 421(1): 621–634, Mar. 2012. doi: 10.1111/j.1365-2966.2012.20340.x.

[9] C. M. Baugh. A primer on hierarchical galaxy formation: the semi-analytical approach. *Reports on Progress in Physics*, 69:3101–3156, Dec 2006. doi: 10.1088/0034-4885/69/12/R02.

[10] P. Behroozi, R. Wechsler, A. Hearin, and C. Conroy. UniverseMachine: The Correlation between Galaxy Growth and Dark Matter Halo Assembly from z=0-10. *arXiv e-prints*, art. arXiv:1806.07893, Jun 2018.

[11] P. S. Behroozi, R. H. Wechsler, and C. Conroy. The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8. , 770:57, Jun 2013. doi: 10.1088/0004-637X/770/1/57.

[12] A. J. Benson. Galaxy formation theory. , 495:33–86, Oct 2010. doi: 10.1016/j.physrep.2010.06.001.

[13] A. J. Benson. Building a predictive model of galaxy formation - I. Phenomenological model constrained to the z = 0 stellar mass function. , 444(3):2599–2636, Nov. 2014. doi: 10.1093/mnras/stu1630.

[14] P. Bett, V. Eke, C. S. Frenk, A. Jenkins, J. Helly, and J. Navarro. The spin and shape of dark matter haloes in the Millennium simulation of a Λ cold dark matter universe. , 376:215–232, Mar 2007. doi: 10.1111/j.1365-2966.2007.11432.x.

[15] H. Bondi. On spherically symmetrical accretion. , 112:195, Jan. 1952. doi: 10.1093/mnras/112.2.195.

[16] H. Bondi and F. Hoyle. On the mechanism of accretion by stars. , 104:273, Jan. 1944. doi: 10.1093/mnras/104.5.273.

[17] R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. , 370(2):645–655, Aug. 2006. doi: 10.1111/j.1365-2966.2006.10519.x.

[18] M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. , 398(3):1150–1164, Sep 2009. doi: 10.1111/j.1365-2966.2009.15191.x.

[19] M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. , 398(3):1150–1164, Sept. 2009. doi: 10.1111/j.1365-2966.2009.15191.x.

[20] H. Brunner, T. Liu, G. Lamer, A. Georgakakis, A. Merloni, M. Brusa, E. Bulbul, K. Dennerl, S. Friedrich, A. Liu, C. Maitra, K. Nandra, M. E. Ramos-Ceja, J. S. Sanders, I. M. Stewart, T. Boller, J. Buchner, N. Clerc, J. Comparat, T. Dwelly, D. Eckert, A. Finoguenov, M. Freyberg, V. Ghirardini, A. Gueguen, F. Haberl, I. Kreykenbohm, M. Krumpe, S. Osterhage, F. Pacaud, P. Predehl, T. H. Reiprich, J. Robrade, M. Salvato, A. Santangelo, T. Schrabback, A. Schwope, and J. Wilms. The eROSITA Final Equatorial Depth Survey (eFEDS): The X-ray catalog. *arXiv e-prints*, art. arXiv:2106.14517, June 2021.

[21] G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. , 344:1000–1028, Oct 2003. doi: 10.1046/j.1365-8711.2003.06897.x.

[22] J. S. Bullock, A. Dekel, T. S. Kolatt, A. V. Kravtsov, A. A. Klypin, C. Porciani, and J. R. Primack. A Universal Angular Momentum Profile for Galactic Halos. , 555:240–257, Jul 2001. doi: 10.1086/321477.

[23] J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel. Profiles of dark haloes: evolution, scatter and environment. , 321:559–575, Mar 2001. doi: 10.1046/j.1365-8711.2001.04068.x.

[24] D. Ceverino and A. Klypin. The role of stellar feedback in the formation of galactic disks and bulges in a ΛCDM Universe. In M. Bureau, E. Athanassoula, and B. Barbuy, editors, *Formation and Evolution of Galaxy Bulges*, volume 245, pages 33–34, July 2008. doi: 10.1017/S1743921308017213.

[25] G. Chabrier. Galactic Stellar and Substellar Initial Mass Function. , 115(809):763–795, July 2003. doi: 10.1086/376392.

[26] K. T. E. Chua, A. Pillepich, M. Vogelsberger, and L. Hernquist. Shape of dark matter haloes in the Illustris simulation: effects of baryons. , 484(1):476–493, Mar 2019. doi: 10.1093/mnras/sty3531.

[27] S. Cole and C. Lacey. The structure of dark matter haloes in hierarchical clustering models. , 281:716, Jul 1996. doi: 10.1093/mnras/281.2.716.

[28] S. A. Cora, C. A. Vega-Martínez, T. Hough, A. N. Ruiz, Á. A. Orsi, A. M. Muñoz Arancibia, I. D. Gargiulo, F. Collacchioni, N. D. Padilla, S. Gottlöber, and G. Yepes. Semi-analytic galaxies - I. Synthesis of environmental and star-forming regulation mechanisms. , 479:2–24, Sep 2018. doi: 10.1093/mnras/sty1131.

[29] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020. ISSN 0027-8424. doi: 10.1073/pnas.1912789117. URL https://www.pnas.org/content/117/48/30055.

[30] D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. , 365(1):11–28, Jan. 2006. doi: 10.1111/j.1365-2966.2005.09675.x.

[31] D. J. Croton, A. R. H. Stevens, C. Tonini, T. Garel, M. Bernyk, A. Bibiano, L. Hodkinson, S. J. Mutch, G. B. Poole, and G. M. Shattow. Semi-Analytic Galaxy Evolution (SAGE): Model Calibration and Basic Results. , 222:22, Feb 2016. doi: 10.3847/0067-0049/222/2/22.

[32] W. Cui and Y. Zhang. *The Impact of Baryons on the Large-Scale Structure of the Universe*, page 7. 2017. doi: 10.5772/68116.

[33] W. Cui, S. Borgani, K. Dolag, G. Murante, and L. Tornatore. The effects of baryons on the halo mass function. , 423(3):2279–2287, Jul 2012. doi: 10.1111/j.1365-2966.2012.21037.x.

[34] N. Dalmasso, D. Zhao, R. Izbicki, and A. B. Lee. Likelihood-Free Frequentist Inference: Bridging Classical Statistics and Machine Learning in Simulation and Uncertainty Quantification. *arXiv e-prints*, art. arXiv:2107.03920, July 2021.

[35] R. Davé, R. Thompson, and P. F. Hopkins. MUFASA: galaxy formation simulations with meshless hydrodynamics. , 462(3):3265–3284, Nov. 2016. doi: 10.1093/mnras/stw1862.

[36] R. Davé, R. Thompson, and P. F. Hopkins. MUFASA: galaxy formation simulations with meshless hydrodynamics. , 462:3265–3284, Nov 2016. doi: 10.1093/mnras/stw1862.

[37] R. Davé, D. Anglés-Alcázar, D. Narayanan, Q. Li, M. H. Rafieferantsoa, and S. Appleby. Simba: Cosmological Simulations with Black Hole Growth and Feedback. *arXiv e-prints*, art. arXiv:1901.10203, Jan 2019.

[38] R. Davé, D. Anglés-Alcázar, D. Narayanan, Q. Li, M. H. Rafieferantsoa, and S. Appleby. SIMBA: Cosmological simulations with black hole growth and feedback. , 486(2):2827–2849, June 2019. doi: 10.1093/mnras/stz937.

[39] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. , 292:371–394, May 1985. doi: 10.1086/163168.

[40] A. Dey, D. J. Schlegel, D. Lang, R. Blum, K. Burleigh, X. Fan, J. R. Findlay, D. Finkbeiner, D. Herrera, S. Juneau, M. Landriau, M. Levi, I. McGreer, A. Meisner, A. D. Myers, J. Moustakas, P. Nugent, A. Patej, E. F. Schlafly, A. R. Walker, F. Valdes, B. A. Weaver, C. Yèche, H. Zou, X. Zhou, B. Abareshi, T. M. C. Abbott,

B. Abolfathi, C. Aguilera, S. Alam, L. Allen, A. Alvarez, J. Annis, B. Ansarinejad, M. Aubert, J. Beechert, E. F. Bell, S. Y. BenZvi, F. Beutler, R. M. Bielby, A. S. Bolton, C. Briceño, E. J. Buckley-Geer, K. Butler, A. Calamida, R. G. Carlberg, P. Carter, R. Casas, F. J. Castander, Y. Choi, J. Comparat, E. Cukanovaite, T. Delubac, K. DeVries, S. Dey, G. Dhungana, M. Dickinson, Z. Ding, J. B. Donaldson, Y. Duan, C. J. Duckworth, S. Eftekharzadeh, D. J. Eisenstein, T. Etourneau, P. A. Fagrelius, J. Farihi, M. Fitzpatrick, A. Font-Ribera, L. Fulmer, B. T. Gänsicke, E. Gaztanaga, K. George, D. W. Gerdes, S. G. A. Gontcho, C. Gorgoni, G. Green, J. Guy, D. Harmer, M. Hernandez, K. Honscheid, L. W. Huang, D. J. James, B. T. Jannuzi, L. Jiang, R. Joyce, A. Karcher, S. Karkar, R. Kehoe, J.-P. Kneib, A. Kueter-Young, T.-W. Lan, T. R. Lauer, L. Le Guillou, A. Le Van Suu, J. H. Lee, M. Lesser, L. Perreault Levasseur, T. S. Li, J. L. Mann, R. Marshall, C. E. Martínez-Vázquez, P. Martini, H. du Mas des Bourboux, S. McManus, T. G. Meier, B. Ménard, N. Metcalfe, A. Muñoz-Gutiérrez, J. Najita, K. Napier, G. Narayan, J. A. Newman, J. Nie, B. Nord, D. J. Norman, K. A. G. Olsen, A. Paat, N. Palanque-Delabrouille, X. Peng, C. L. Poppett, M. R. Poremba, A. Prakash, D. Rabinowitz, A. Raichoor, M. Rezaie, A. N. Robertson, N. A. Roe, A. J. Ross, N. P. Ross, G. Rudnick, S. Safonova, A. Saha, F. J. Sánchez, E. Savary, H. Schweiker, A. Scott, H.-J. Seo, H. Shan, D. R. Silva, Z. Slepian, C. Soto, D. Sprayberry, R. Staten, C. M. Stillman, R. J. Stupak, D. L. Summers, S. Sien Tie, H. Tirado, M. Vargas-Magaña, A. K. Vivas, R. H. Wechsler, D. Williams, J. Yang, Q. Yang, T. Yapici, D. Zaritsky, A. Zenteno, K. Zhang, T. Zhang, R. Zhou, and Z. Zhou. Overview of the DESI Legacy Imaging Surveys. , 157(5): 168, May 2019. doi: 10.3847/1538-3881/ab089d.

[41] M. Donnari, A. Pillepich, D. Nelson, M. Vogelsberger, S. Genel, R. Weinberger, F. Marinacci, V. Springel, and L. Hernquist. The star formation activity of IllustrisTNG galaxies: main sequence, UVJ diagram, quenched fractions, and systematics. , 485(4):4817–4840, June

2019. doi: 10.1093/mnras/stz712.

[42] M. Donnari, A. Pillepich, G. D. Joshi, D. Nelson, S. Genel, F. Marinacci, V. Rodriguez-Gomez, R. Pakmor, P. Torrey, M. Vogelsberger, and L. Hernquist. Quenched fractions in the IllustrisTNG simulations: the roles of AGN feedback, environment, and pre-processing. , 500(3):4004–4024, Jan. 2021. doi: 10.1093/mnras/staa3006.

[43] Y. Dubois, R. Gavazzi, S. Peirani, and J. Silk. AGN-driven quenching of star formation: morphological and dynamical implications for early-type galaxies. , 433(4):3297–3313, Aug. 2013. doi: 10.1093/mnras/stt997.

[44] Y. Dubois, C. Pichon, C. Welker, D. Le Borgne, J. Devriendt, C. Laigle, S. Codis, D. Pogosyan, S. Arnouts, K. Benabed, E. Bertin, J. Blaizot, F. Bouchet, J. F. Cardoso, S. Colombi, V. de Lapparent, V. Desjacques, R. Gavazzi, S. Kassin, T. Kimm, H. McCracken, B. Milliard, S. Peirani, S. Prunet, S. Rouberol, J. Silk, A. Slyz, T. Sousbie, R. Teyssier, L. Tresse, M. Treyer, D. Vibert, and M. Volonteri. Dancing in the dark: galactic properties trace spin swings along the cosmic web. , 444(2):1453–1468, Oct. 2014. doi: 10.1093/mnras/stu1227.

[45] Y. Dubois, C. Pichon, C. Welker, D. Le Borgne, J. Devriendt, C. Laigle, S. Codis, D. Pogosyan, S. Arnouts, K. Benabed, E. Bertin, J. Blaizot, F. Bouchet, J. F. Cardoso, S. Colombi, V. de Lapparent, V. Desjacques, R. Gavazzi, S. Kassin, T. Kimm, H. McCracken, B. Milliard, S. Peirani, S. Prunet, S. Rouberol, J. Silk, A. Slyz, T. Sousbie, R. Teyssier, L. Tresse, M. Treyer, D. Vibert, and M. Volonteri. Dancing in the dark: galactic properties trace spin swings along the cosmic web. , 444:1453–1468, Oct 2014. doi: 10.1093/mnras/stu1227.

[46] A. R. Duffy, J. Schaye, S. T. Kay, C. Dalla Vecchia, R. A. Battye, and C. M. Booth. Impact of baryon physics on dark matter structures: a

detailed simulation study of halo density profiles. , 405(4):2161–2178, Jul 2010. doi: 10.1111/j.1365-2966.2010.16613.x.

[47] C. Durkan, I. Murray, and G. Papamakarios. On Contrastive Learning for Likelihood-free Inference. *arXiv e-prints*, art. arXiv:2002.03712, Feb. 2020.

[48] D. Elbaz, M. Dickinson, H. S. Hwang, T. Díaz-Santos, G. Magdis, B. Magnelli, D. Le Borgne, F. Galliano, M. Pannella, P. Chanial, L. Armus, V. Charmandaris, E. Daddi, H. Aussel, P. Popesso, J. Kartaltepe, B. Altieri, I. Valtchanov, D. Coia, H. Dannerbauer, K. Dasyra, R. Leiton, J. Mazzarella, D. M. Alexander, V. Buat, D. Burgarella, R. R. Chary, R. Gilli, R. J. Ivison, S. Juneau, E. Le Floc'h, D. Lutz, G. E. Morrison, J. R. Mullaney, E. Murphy, A. Pope, D. Scott, M. Brodwin, D. Calzetti, C. Cesarsky, S. Charlot, H. Dole, P. Eisenhardt, H. C. Ferguson, N. Förster Schreiber, D. Frayer, M. Giavalisco, M. Huynh, A. M. Koekemoer, C. Papovich, N. Reddy, C. Surace, H. Teplitz, M. S. Yun, and G. Wilson. GOODS-Herschel: an infrared main sequence for star-forming galaxies. , 533:A119, Sep 2011. doi: 10.1051/0004-6361/201117239.

[49] E. J. Elliott, C. M. Baugh, and C. G. Lacey. Efficient exploration and calibration of a semi-analytical model of galaxy formation with deep learning. , 506(3):4011–4030, Sept. 2021. doi: 10.1093/mnras/stab1837.

[50] A. C. Fabian. Observational Evidence of Active Galactic Nuclei Feedback. , 50:455–489, Sept. 2012. doi: 10.1146/annurev-astro-081811-125521.

[51] Y. Feng, T. Di-Matteo, R. A. Croft, S. Bird, N. Battaglia, and S. Wilkins. The BlueTides simulation: first galaxies and reionization. , 455(3):2778–2791, Jan. 2016. doi: 10.1093/mnras/stv2484.

[52] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods

for astronomical data analyses. , 384(2):449–463, Feb. 2008. doi: 10.1111/j.1365-2966.2007.12353.x.

[53] F. Fontanot, A. Calabrò, M. Talia, F. Mannucci, M. Castellano, G. Cresci, G. De Lucia, A. Gallazzi, M. Hirschmann, L. Pentericci, L. Xie, R. Amorin, M. Bolzonella, A. Bongiorno, O. Cucciati, F. Cullen, J. P. U. Fynbo, N. Hathi, P. Hibon, R. J. McLure, and L. Pozzetti. The evolution of the mass-metallicity relations from the VANDELS survey and the GAEA semi-analytic model. , 504(3): 4481–4492, July 2021. doi: 10.1093/mnras/stab1213.

[54] M. Gabrié, A. Manoel, C. Luneau, J. Barbier, N. Macris, F. Krzakala, and L. Zdeborová. Entropy and mutual information in models of deep neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 12(12):124014, Dec. 2019. doi: 10.1088/1742-5468/ab3430.

[55] A. Gallazzi, S. Charlot, J. Brinchmann, S. D. M. White, and C. A. Tremonti. The ages and metallicities of galaxies in the local universe. , 362(1):41–58, Sept. 2005. doi: 10.1111/j.1365-2966.2005.09321.x.

[56] S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, D. Nelson, G. Snyder, V. Rodriguez-Gomez, P. Torrey, and L. Hernquist. Introducing the Illustris project: the evolution of galaxy populations across cosmic time. , 445:175–200, Nov 2014. doi: 10.1093/mnras/stu1654.

[57] S. Genel, M. Vogelsberger, V. Springel, D. Sijacki, D. Nelson, G. Snyder, V. Rodriguez-Gomez, P. Torrey, and L. Hernquist. Introducing the Illustris project: the evolution of galaxy populations across cosmic time. , 445(1):175–200, Nov. 2014. doi: 10.1093/mnras/stu1654.

[58] S. Genel, G. L. Bryan, V. Springel, L. Hernquist, D. Nelson, A. Pillepich, R. Weinberger, R. Pakmor, F. Marinacci, and M. Vogelsberger. A Quantification of the Butterfly Effect in Cosmological

Simulations and Implications for Galaxy Scaling Relations. , 871(1): 21, Jan. 2019. doi: 10.3847/1538-4357/aaf4bb.

[59] F. Gerardi, S. M. Feeney, and J. Alsing. Unbiased likelihood-free inference of the Hubble constant from light standard sirens. , 104(8): 083531, Oct. 2021. doi: 10.1103/PhysRevD.104.083531.

[60] Z. Goldfeld, E. van den Berg, K. Greenewald, I. Melnyk, N. Nguyen, B. Kingsbury, and Y. Polyanskiy. Estimating Information Flow in Deep Neural Networks. *arXiv e-prints*, art. arXiv:1810.05728, Oct. 2018.

[61] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. , 23(2):347–356, Jan. 1981. doi: 10.1103/PhysRevD.23.347.

[62] K. Heitmann, N. Frontiere, C. Sewell, S. Habib, A. Pope, H. Finkel, S. Rizzi, J. Insley, and S. Bhattacharya. The Q Continuum Simulation: Harnessing the Power of GPU Accelerated Supercomputers. , 219(2):34, Aug 2015. doi: 10.1088/0067-0049/219/2/34.

[63] M. A. Henson, D. J. Barnes, S. T. Kay, I. G. McCarthy, and J. Schaye. The impact of baryons on massive galaxy clusters: halo structure and cluster mass estimates. , 465(3):3361–3378, Mar 2017. doi: 10.1093/mnras/stw2899.

[64] J. Hermans, V. Begy, and G. Louppe. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv e-prints*, art. arXiv:1903.04057, Mar. 2019.

[65] M. Hirschmann, K. Dolag, A. Saro, L. Bachmann, S. Borgani, and A. Burkert. Cosmological simulations of black hole growth: AGN luminosities and downsizing. , 442(3):2304–2324, Aug. 2014. doi: 10.1093/mnras/stu1023.

[66] M.-F. Ho, S. Bird, and C. R. Shelton. Multifidelity emulation for

the matter power spectrum using Gaussian processes. , 509(2):2551–2565, Jan. 2022. doi: 10.1093/mnras/stab3114.

[67] E. Holmberg. On the Clustering Tendencies among the Nebulae. II. a Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure. , 94:385, Nov. 1941. doi: 10.1086/144344.

[68] F. Hoyle and R. A. Lyttleton. The effect of interstellar matter on climatic variation. *Proceedings of the Cambridge Philosophical Society*, 35(3):405, Jan. 1939. doi: 10.1017/S0305004100021150.

[69] D. A. Hubber, S. A. E. G. Falle, and S. P. Goodwin. Convergence of AMR and SPH simulations - I. Hydrodynamical resolution and convergence tests. , 432(1):711–727, June 2013. doi: 10.1093/mnras/stt509.

[70] D. Huppenkothen and M. Bachetti. Accurate X-ray Timing in the Presence of Systematic Biases With Simulation-Based Inference. , Dec. 2021. doi: 10.1093/mnras/stab3437.

[71] T. Ishiyama, M. Enoki, M. A. R. Kobayashi, R. Makiya, M. Nagashima, and T. Oogi. The ν^2GC simulations: Quantifying the dark side of the universe in the Planck cosmology. , 67(4):61, Aug 2015. doi: 10.1093/pasj/psv021.

[72] A. Jenkins, C. S. Frenk, F. R. Pearce, P. A. Thomas, J. M. Colberg, S. D. M. White, H. M. P. Couchman, J. A. Peacock, G. Efstathiou, and A. H. Nelson. Evolution of Structure in Cold Dark Matter Universes. , 499:20–40, May 1998. doi: 10.1086/305615.

[73] H. M. Kamdar, M. J. Turk, and R. J. Brunner. Machine learning and cosmological simulations - I. Semi-analytical models. , 455:642–658, Jan 2016. doi: 10.1093/mnras/stv2310.

[74] H. M. Kamdar, M. J. Turk, and R. J. Brunner. Machine learning and

cosmological simulations - II. Hydrodynamical simulations. , 457: 1162–1179, Apr 2016. doi: 10.1093/mnras/stv2981.

[75] N. Katz, D. H. Weinberg, and L. Hernquist. Cosmological Simulations with TreeSPH. , 105:19, Jul 1996. doi: 10.1086/192305.

[76] B. W. Keller, J. W. Wadsley, L. Wang, and J. M. D. Kruijssen. Chaos and variance in galaxy formation. , 482(2):2244–2261, Jan. 2019. doi: 10.1093/mnras/sty2859.

[77] N. Khandai, T. Di Matteo, R. Croft, S. Wilkins, Y. Feng, E. Tucker, C. DeGraf, and M.-S. Liu. The MassiveBlack-II simulation: the evolution of haloes and galaxies to z $\sim$ 0. , 450(2):1349–1374, June 2015. doi: 10.1093/mnras/stv627.

[78] J. B. Kinney and G. S. Atwal. Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Science*, 111(9):3354–3359, Mar. 2014. doi: 10.1073/pnas.1309933111.

[79] A. Klypin, S. Gottlöber, A. V. Kravtsov, and A. M. Khokhlov. Galaxies in N-Body Simulations: Overcoming the Overmerging Problem. , 516:530–551, May 1999. doi: 10.1086/307122.

[80] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. , 457(4):4340–4359, Apr. 2016. doi: 10.1093/mnras/stw248.

[81] A. Klypin, G. Yepes, S. Gottlöber, F. Prada, and S. Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. , 457:4340–4359, Apr 2016. doi: 10.1093/mnras/stw248.

[82] A. A. Klypin, S. Trujillo-Gomez, and J. Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. , 740(2):102, Oct. 2011. doi: 10.1088/0004-637X/740/2/102.

[83] A. Knebe, D. Stoppacher, F. Prada, C. Behrens, A. Benson, S. A. Cora, D. J. Croton, N. D. Padilla, A. N. Ruiz, M. Sinha, A. R. H. Stevens, C. A. Vega-Martínez, P. Behroozi, V. Gonzalez-Perez, S. Gottlöber, A. A. Klypin, G. Yepes, H. Enke, N. I. Libeskind, K. Riebe, and M. Steinmetz. MULTIDARK-GALAXIES: data release and first results. , 474:5206–5231, Mar 2018. doi: 10.1093/mnras/stx2662.

[84] D. Kodi Ramanah, R. Wojtak, and N. Arendse. Simulation-based inference of dynamical galaxy cluster masses with 3D convolutional neural networks. , 501(3):4080–4091, Mar. 2021. doi: 10.1093/mnras/staa3922.

[85] J. Kormendy and K. Gebhardt. Supermassive black holes in galactic nuclei. In J. C. Wheeler and H. Martel, editors, *20th Texas Symposium on relativistic astrophysics*, volume 586 of *American Institute of Physics Conference Series*, pages 363–381, Oct. 2001. doi: 10.1063/1.1419581.

[86] J. Kormendy and L. C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. , 51(1):511–653, Aug. 2013. doi: 10.1146/annurev-astro-082708-101811.

[87] J. Kormendy and R. D. McClure. The Nucleus of M33. , 105:1793, May 1993. doi: 10.1086/116555.

[88] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. , 69(6):066138, June 2004. doi: 10.1103/PhysRevE.69.066138.

[89] K. Kuijken, C. Heymans, A. Dvornik, H. Hildebrandt, J. T. A. de Jong, A. H. Wright, T. Erben, M. Bilicki, B. Giblin, H. Y. Shan, F. Getman, A. Grado, H. Hoekstra, L. Miller, N. Napolitano, M. Paolilo, M. Radovich, P. Schneider, W. Sutherland, M. Tewes, C. Tortora, E. A. Valentijn, and G. A. Verdoes Kleijn. The fourth data release of the Kilo-Degree Survey: ugri imaging and nine-band optical-IR

photometry over 1000 square degrees. , 625:A2, May 2019. doi: 10.1051/0004-6361/201834918.

[90] A. M. C. Le Brun, I. G. McCarthy, J. Schaye, and T. J. Ponman. Towards a realistic population of simulated galaxy groups and clusters. , 441(2):1270–1290, June 2014. doi: 10.1093/mnras/stu608.

[91] J. Lee, J. Shin, O. N. Snaith, Y. Kim, C. G. Few, J. Devriendt, Y. Dubois, L. M. Cox, S. E. Hong, O.-K. Kwon, C. Park, C. Pichon, J. Kim, B. K. Gibson, and C. Park. The Horizon Run 5 Cosmological Hydrodynamical Simulation: Probing Galaxy Formation from Kilo- to Gigaparsec Scales. , 908(1):11, Feb. 2021. doi: 10.3847/1538-4357/abd08b.

[92] R. Legin, Y. Hezaveh, L. Perreault Levasseur, and B. Wandelt. Simulation-Based Inference of Strong Gravitational Lensing Parameters. *arXiv e-prints*, art. arXiv:2112.05278, Dec. 2021.

[93] J. Leja, B. D. Johnson, C. Conroy, P. van Dokkum, J. S. Speagle, G. Brammer, I. Momcheva, R. Skelton, K. E. Whitaker, M. Franx, and E. J. Nelson. An Older, More Quiescent Universe from Panchromatic SED Fitting of the 3D-HST Survey. , 877(2):140, June 2019. doi: 10.3847/1538-4357/ab1d5a.

[94] J. Leja, J. S. Speagle, B. D. Johnson, C. Conroy, P. van Dokkum, and M. Franx. A New Census of the 0.2 ¡ z ¡ 3.0 Universe. I. The Stellar Mass Function. , 893(2):111, Apr. 2020. doi: 10.3847/1538-4357/ab7e27.

[95] J. Leja, J. S. Speagle, Y.-S. Ting, B. D. Johnson, C. Conroy, K. E. Whitaker, E. J. Nelson, P. van Dokkum, and M. Franx. A New Census of the 0.2¡ z ¡3.0 Universe, Part II: The Star-Forming Sequence. *arXiv e-prints*, art. arXiv:2110.04314, Oct. 2021.

[96] P. Lemos, N. Jeffrey, L. Whiteway, O. Lahav, N. Libeskind, and Y. Hoffman. Sum of the masses of the Milky Way and M31: A

likelihood-free inference approach. , 103(2):023009, Jan. 2021. doi: 10.1103/PhysRevD.103.023009.

[97] G. Lemson and G. Kauffmann. Environmental influences on dark matter haloes and consequences for the galaxies within them. , 302: 111–117, Jan 1999. doi: 10.1046/j.1365-8711.1999.02090.x.

[98] J. Lequeux, M. Peimbert, J. F. Rayo, A. Serrano, and S. Torres-Peimbert. Reprint of 1979A&A....80..155L. Chemical composition and evolution of irregular and blue compact galaxies. , 500:145–156, Dec. 1979.

[99] C. Lia, G. Carraro, and P. Salucci. The dark and luminous matter coupling in the formation of spheroids: a SPH investigation. , 360: 76–84, Aug. 2000.

[100] C. Lia, L. Portinari, and G. Carraro. Star formation and chemical evolution in smoothed particle hydrodynamics simulations: a statistical approach. , 330(4):821–836, Mar. 2002. doi: 10.1046/j.1365-8711.2002.05118.x.

[101] Y. Lu, H. J. Mo, N. Katz, and M. D. Weinberg. Bayesian inference of galaxy formation from the K-band luminosity function of galaxies: tensions between theory and observation. , 421(2):1779–1796, Apr. 2012. doi: 10.1111/j.1365-2966.2012.20435.x.

[102] Y. Lu, H. J. Mo, Z. Lu, N. Katz, and M. D. Weinberg. Bayesian inferences of galaxy formation from the K-band luminosity and H I mass functions of galaxies: constraining star formation and feedback. , 443(2):1252–1266, Sept. 2014. doi: 10.1093/mnras/stu1200.

[103] J.-M. Lueckmann, G. Bassetto, T. Karaletsos, and J. H. Macke. Likelihood-free inference with emulator networks. *arXiv e-prints*, art. arXiv:1805.09294, May 2018.

[104] P. Madau and M. Dickinson. Cosmic Star-Formation History. , 52: 415–486, Aug. 2014. doi: 10.1146/annurev-astro-081811-125615.

[105] P. Madau, J. Diemand, and M. Kuhlen. Dark Matter Subhalos and the Dwarf Satellites of the Milky Way. , 679:1260–1271, Jun 2008. doi: 10.1086/587545.

[106] F. Marinacci, M. Vogelsberger, R. Pakmor, P. Torrey, V. Springel, L. Hernquist, D. Nelson, R. Weinberger, A. Pillepich, J. Naiman, and S. Genel. First results from the IllustrisTNG simulations: radio haloes and magnetic fields. , 480(4):5113–5139, Nov. 2018. doi: 10.1093/mnras/sty2206.

[107] D. Martizzi, R. Teyssier, B. Moore, and T. Wentz. The effects of baryon physics, black holes and active galactic nucleus feedback on the mass distribution in clusters of galaxies. , 422(4):3081–3091, Jun 2012. doi: 10.1111/j.1365-2966.2012.20879.x.

[108] I. G. McCarthy, J. Schaye, S. Bird, and A. M. C. Le Brun. The BAHAMAS project: calibrated hydrodynamical simulations for large-scale structure cosmology. , 465(3):2936–2965, Mar. 2017. doi: 10.1093/mnras/stw2792.

[109] S. S. McGaugh, J. M. Schombert, G. D. Bothun, and W. J. G. de Blok. The Baryonic Tully-Fisher Relation. , 533(2):L99–L102, Apr. 2000. doi: 10.1086/312628.

[110] S. S. McGaugh, F. Lelli, J. M. Schombert, P. Li, T. Visgaitis, K. S. Parker, and M. S. Pawlowski. The Baryonic Tully-Fisher Relation in the Local Group and the Equivalent Circular Velocity of Pressure-supported Dwarfs. , 162(5):202, Nov. 2021. doi: 10.3847/1538-3881/ac2502.

[111] D. J. McLeod, R. J. McLure, J. S. Dunlop, F. Cullen, A. C. Carnall, and K. Duncan. The evolution of the galaxy stellar-mass function over the last 12 billion years from a combination of ground-based and HST surveys. , 503(3):4413–4435, May 2021. doi: 10.1093/mnras/stab731.

[112] D. Merritt and L. Ferrarese. The M-σ Relation for Supermassive Black Holes. , 547(1):140–145, Jan. 2001. doi: 10.1086/318372.

[113] S. Mohamed and B. Lakshminarayanan. Learning in Implicit Generative Models. *arXiv e-prints*, art. arXiv:1610.03483, Oct. 2016.

[114] B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi. Dark Matter Substructure within Galactic Halos. , 524: L19–L22, Oct 1999. doi: 10.1086/312287.

[115] B. P. Moster, T. Naab, and S. D. M. White. EMERGE - an empirical model for the formation of galaxies since z 10. , 477:1822–1852, Jun 2018. doi: 10.1093/mnras/sty655.

[116] L. A. Mowla, P. van Dokkum, G. B. Brammer, I. Momcheva, A. van der Wel, K. Whitaker, E. Nelson, R. Bezanson, A. Muzzin, M. Franx, J. MacKenty, J. Leja, M. Kriek, and D. Marchesini. COSMOS-DASH: The Evolution of the Galaxy Size-Mass Relation since z $\sim$ 3 from New Wide-field WFC3 Imaging Combined with CANDELS/3D-HST. , 880(1):57, July 2019. doi: 10.3847/1538-4357/ab290a.

[117] K. Nagamine. Star Formation History of Dwarf Galaxies in Cosmological Hydrodynamic Simulations. *Advances in Astronomy*, 2010: 651621, Jan. 2010. doi: 10.1155/2010/651621.

[118] J. P. Naiman, A. Pillepich, V. Springel, E. Ramirez-Ruiz, P. Torrey, M. Vogelsberger, R. Pakmor, D. Nelson, F. Marinacci, L. Hernquist, R. Weinberger, and S. Genel. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium. , 477(1):1206–1224, June 2018. doi: 10.1093/mnras/sty618.

[119] J. P. Naiman, A. Pillepich, V. Springel, E. Ramirez-Ruiz, P. Torrey, M. Vogelsberger, R. Pakmor, D. Nelson, F. Marinacci, L. Hernquist, R. Weinberger, and S. Genel. First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution

of magnesium and europium. , 477:1206–1224, Jun 2018. doi: 10.1093/mnras/sty618.

[120] J. F. Navarro, C. S. Frenk, and S. D. M. White. A Universal Density Profile from Hierarchical Clustering. , 490:493–508, Dec 1997. doi: 10.1086/304888.

[121] D. Nelson, A. Pillepich, S. Genel, M. Vogelsberger, V. Springel, P. Torrey, V. Rodriguez-Gomez, D. Sijacki, G. F. Snyder, B. Griffen, F. Marinacci, L. Blecha, L. Sales, D. Xu, and L. Hernquist. The illustris simulation: Public data release. *Astronomy and Computing*, 13:12–37, Nov. 2015. doi: 10.1016/j.ascom.2015.09.003.

[122] D. Nelson, A. Pillepich, V. Springel, R. Weinberger, L. Hernquist, R. Pakmor, S. Genel, P. Torrey, M. Vogelsberger, G. Kauffmann, F. Marinacci, and J. Naiman. First results from the IllustrisTNG simulations: the galaxy colour bimodality. , 475:624–647, Mar 2018. doi: 10.1093/mnras/stx3040.

[123] D. Nelson, A. Pillepich, V. Springel, R. Weinberger, L. Hernquist, R. Pakmor, S. Genel, P. Torrey, M. Vogelsberger, G. Kauffmann, F. Marinacci, and J. Naiman. First results from the IllustrisTNG simulations: the galaxy colour bimodality. , 475(1):624–647, Mar. 2018. doi: 10.1093/mnras/stx3040.

[124] D. Nelson, V. Springel, A. Pillepich, V. Rodriguez-Gomez, P. Torrey, S. Genel, M. Vogelsberger, R. Pakmor, F. Marinacci, R. Weinberger, L. Kelley, M. Lovell, B. Diemer, and L. Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.

[125] K. G. Noeske, B. J. Weiner, S. M. Faber, C. Papovich, D. C. Koo, R. S. Somerville, K. Bundy, C. J. Conselice, J. A. Newman, D. Schiminovich, E. Le Floc'h, A. L. Coil, G. H. Rieke, J. M. Lotz, J. R. Primack, P. Barmby, M. C. Cooper, M. Davis, R. S. Ellis, G. G. Fazio, P. Guhathakurta, J. Huang, S. A. Kassin, D. C. Martin, A. C.

Phillips, R. M. Rich, T. A. Small, C. N. A. Willmer, and G. Wilson. Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies. , 660(1):L43–L46, May 2007. doi: 10.1086/517926.

[126] B. K. Oh, B. D. Smith, J. A. Peacock, and S. Khochfar. Calibration of a star formation and feedback model for cosmological simulations with ENZO. , 497(4):5203–5219, Oct. 2020. doi: 10.1093/mnras/staa2318.

[127] M. J. Page, M. Symeonidis, J. D. Vieira, B. Altieri, A. Amblard, V. Arumugam, H. Aussel, T. Babbedge, A. Blain, J. Bock, A. Boselli, V. Buat, N. Castro-Rodríguez, A. Cava, P. Chanial, D. L. Clements, A. Conley, L. Conversi, A. Cooray, C. D. Dowell, E. N. Dubois, J. S. Dunlop, E. Dwek, S. Dye, S. Eales, D. Elbaz, D. Farrah, M. Fox, A. Franceschini, W. Gear, J. Glenn, M. Griffin, M. Halpern, E. Hatziminaoglou, E. Ibar, K. Isaak, R. J. Ivison, G. Lagache, L. Levenson, N. Lu, S. Madden, B. Maffei, G. Mainetti, L. Marchetti, H. T. Nguyen, B. O'Halloran, S. J. Oliver, A. Omont, P. Panuzzo, A. Papageorgiou, C. P. Pearson, I. Pérez-Fournon, M. Pohlen, J. I. Rawlings, D. Rigopoulou, L. Riguccini, D. Rizzo, G. Rodighiero, I. G. Roseboom, M. Rowan-Robinson, M. S. Portal, B. Schulz, D. Scott, N. Seymour, D. L. Shupe, A. J. Smith, J. A. Stevens, M. Trichas, K. E. Tugwell, M. Vaccari, I. Valtchanov, M. Viero, L. Vigroux, L. Wang, R. Ward, G. Wright, C. K. Xu, and M. Zemcov. The suppression of star formation by powerful active galactic nuclei. , 485(7397):213–216, May 2012. doi: 10.1038/nature11096.

[128] R. Pakmor and V. Springel. Simulations of magnetic fields in isolated disc galaxies. , 432(1):176–193, June 2013. doi: 10.1093/mnras/stt428.

[129] R. Pakmor and V. Springel. Simulations of magnetic fields in isolated disc galaxies. , 432:176–193, Jun 2013. doi: 10.1093/mnras/stt428.

[130] R. Pakmor, A. Bauer, and V. Springel. Magnetohydrodynamics on an unstructured moving grid. , 418:1392–1401, Dec 2011. doi: 10.1111/j.1365-2966.2011.19591.x.

[131] L. Paninski. Estimation of Entropy and Mutual Information. *Neural Computation*, 15(6):1191–1253, 06 2003. ISSN 0899-7667. doi: 10.1162/089976603321780272. URL https://doi.org/10.1162/089976603321780272.

[132] G. Papamakarios. Neural Density Estimation and Likelihood-free Inference. *arXiv e-prints*, art. arXiv:1910.13233, Oct. 2019.

[133] G. Papamakarios and I. Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, art. arXiv:1605.06376, May 2016.

[134] G. Papamakarios, D. C. Sterratt, and I. Murray. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *arXiv e-prints*, art. arXiv:1805.07226, May 2018.

[135] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[136] A. Pillepich, D. Nelson, L. Hernquist, V. Springel, R. Pakmor, P. Torrey, R. Weinberger, S. Genel, J. P. Naiman, F. Marinacci, and M. Vogelsberger. First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies. , 475(1):648–675, Mar. 2018. doi: 10.1093/mnras/stx3112.

[137] A. Pillepich, V. Springel, D. Nelson, S. Genel, J. Naiman, R. Pakmor, L. Hernquist, P. Torrey, M. Vogelsberger, R. Weinberger, and F. Marinacci. Simulating galaxy formation with the IllustrisTNG model. , 473:4077–4106, Jan 2018. doi: 10.1093/mnras/stx2656.

[138] A. Pillepich, V. Springel, D. Nelson, S. Genel, J. Naiman, R. Pakmor, L. Hernquist, P. Torrey, M. Vogelsberger, R. Weinberger, and F. Marinacci. Simulating galaxy formation with the IllustrisTNG model. , 473(3):4077–4106, Jan. 2018. doi: 10.1093/mnras/stx2656.

[139] J. M. Piotrowska, A. F. L. Bluck, R. Maiolino, and Y. Peng. On the quenching of star formation in observed and simulated central galaxies: evidence for the role of integrated AGN feedback. , 512(1): 1052–1090, May 2022. doi: 10.1093/mnras/stab3673.

[140] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barand ela, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. Bobin, J. J. Bock, A. Bonaldi, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bridges, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, B. Cappellini, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, X. Chen, H. C. Chiang, L. Y. Chiang, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Delouis, F. X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, T. C. Gaier, S. Galeotta, S. Galli, K. Ganga, M. Giard, G. Giardino, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Haissinski, J. Hamann, F. K. Hansen, D. Hanson, D. Harrison, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrand t, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, Z. Hou, W. Hovest, K. M. Huffenberger, A. H. Jaffe, T. R. Jaffe, J. Jewell, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner,

R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, R. J. Laureijs, C. R. Lawrence, S. Leach, J. P. Leahy, R. Leonardi, J. León-Tavares, J. Lesgourgues, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, B. Maffei, D. Maino, N. Mand olesi, M. Maris, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, F. Matthai, P. Mazzotta, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, E. Menegoni, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, S. Osborne, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, D. Pearson, T. J. Pearson, H. V. Peiris, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, P. Platania, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, C. Rosset, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, E. P. S. Shellard, L. D. Spencer, J. L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2013 results. XVI. Cosmological parameters. , 571:A16, Nov 2014. doi: 10.1051/0004-6361/201321591.

[141] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, J. Chluba, P. R. Christensen, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, M. Farhang, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, T. Ghosh, M. Giard, Y. Giraud-Héraud, E. Giusarma, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. Hanson, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrand t, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, Z. Huang, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marchini, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Masi, S. Matarrese, P. McGehee, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A.

Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, R. Paladini, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, L. Popa, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, B. Rouillé d'Orfeuil, M. Rowan-Robinson, J. A. Rubiño-Martín, B. Rusholme, N. Said, V. Salvatelli, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, M. D. Seiffert, P. Serra, E. P. S. Shellard, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Wilkinson, D. Yvon, A. Zacchei, and A. Zonca. Planck 2015 results. XIII. Cosmological parameters. , 594:A13, Sep 2016. doi: 10.1051/0004-6361/201525830.

[142] F. Prada, A. A. Klypin, E. Simonneau, J. Betancort-Rijo, S. Patiri, S. Gottlöber, and M. A. Sanchez-Conde. How Far Do They Go? The Outer Structure of Galactic Dark Matter Halos. , 645:1001–1011, Jul 2006. doi: 10.1086/504456.

[143] F. Prada, A. A. Klypin, A. J. Cuesta, J. E. Betancort-Rijo, and J. Primack. Halo concentrations in the standard Λ cold dark matter cosmology. , 423(4):3018–3030, Jul 2012. doi: 10.1111/j.1365-2966.2012.21007.x.

[144] A. Ragagnin, K. Dolag, V. Biffi, M. Cadolle Bel, N. J. Hammer,

A. Krukau, M. Petkova, and D. Steinborn. A web portal for hydrodynamical, cosmological simulations. *Astronomy and Computing*, 20: 52–67, July 2017. doi: 10.1016/j.ascom.2017.05.001.

[145] J. A. Regan, P. H. Johansson, and M. G. Haehnelt. Numerical resolution effects on simulations of massive black hole seeds. , 439(1): 1160–1175, Mar. 2014. doi: 10.1093/mnras/stu068.

[146] K. Riebe, A. M. Partl, H. Enke, J. Forero-Romero, S. Gottlöber, A. Klypin, G. Lemson, F. Prada, J. R. Primack, M. Steinmetz, and V. Turchaninov. The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. *Astronomische Nachrichten*, 334(7):691–708, Aug 2013. doi: 10.1002/asna.201211900.

[147] K. Riebe, A. M. Partl, H. Enke, J. Forero-Romero, S. Gottlöber, A. Klypin, G. Lemson, F. Prada, J. R. Primack, M. Steinmetz, and V. Turchaninov. The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. *Astronomische Nachrichten*, 334:691–708, Aug 2013. doi: 10.1002/asna.201211900.

[148] V. Rodriguez-Gomez, S. Genel, M. Vogelsberger, D. Sijacki, A. Pillepich, L. V. Sales, P. Torrey, G. Snyder, D. Nelson, V. Springel, C.-P. Ma, and L. Hernquist. The merger rate of galaxies in the Illustris simulation: a comparison with observations and semi-empirical models. , 449:49–64, May 2015. doi: 10.1093/mnras/stv264.

[149] A. Rodríguez-Puebla, P. Behroozi, J. Primack, A. Klypin, C. Lee, and D. Hellinger. Halo and subhalo demographics with Planck cosmological parameters: Bolshoi-Planck and MultiDark-Planck simulations. , 462(1):893–916, Oct 2016. doi: 10.1093/mnras/stw1705.

[150] A. Rodríguez-Puebla, J. R. Primack, V. Avila-Reese, and S. M. Faber. Constraining the galaxy-halo connection over the last 13.3 Gyr: star formation histories, galaxy mergers and structural properties. , 470: 651–687, Sep 2017. doi: 10.1093/mnras/stx1172.

[151] Y. M. Rosas-Guevara, R. G. Bower, J. Schaye, M. Furlong, C. S. Frenk, C. M. Booth, R. A. Crain, C. Dalla Vecchia, M. Schaller, and T. Theuns. The impact of angular momentum on black hole accretion rates in simulations of galaxy formation. , 454(1):1038–1057, Nov. 2015. doi: 10.1093/mnras/stv2056.

[152] B. C. Ross. Mutual Information between Discrete and Continuous Data Sets. *PLoS ONE*, 9(2):e87357, Feb. 2014. doi: 10.1371/journal.pone.0087357.

[153] S. Salim, J. C. Lee, S. Janowiecki, E. da Cunha, M. Dickinson, M. Boquien, D. Burgarella, J. J. Salzer, and S. Charlot. GALEX-SDSS-WISE Legacy Catalog (GSWLC): Star Formation Rates, Stellar Masses, and Dust Attenuations of 700,000 Low-redshift Galaxies. , 227(1):2, Nov 2016. doi: 10.3847/0067-0049/227/1/2.

[154] T. Sawala, C. S. Frenk, R. A. Crain, A. Jenkins, J. Schaye, T. Theuns, and J. Zavala. The abundance of (not just) dark matter haloes. , 431 (2):1366–1382, May 2013. doi: 10.1093/mnras/stt259.

[155] J. Schaye and C. Dalla Vecchia. On the relation between the Schmidt and Kennicutt-Schmidt star formation laws and its implications for numerical simulations. , 383:1210–1222, Jan 2008. doi: 10.1111/j.1365-2966.2007.12639.x.

[156] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M. Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A. Thomas, and J. Trayford. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. , 446:521–554, Jan 2015. doi: 10.1093/mnras/stu2058.

[157] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M.

Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A. Thomas, and J. Trayford. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. , 446(1): 521–554, Jan. 2015. doi: 10.1093/mnras/stu2058.

[158] S. Shen, H. J. Mo, S. D. M. White, M. R. Blanton, G. Kauffmann, W. Voges, J. Brinkmann, and I. Csabai. The size distribution of galaxies in the Sloan Digital Sky Survey. , 343(3):978–994, Aug. 2003. doi: 10.1046/j.1365-8711.2003.06740.x.

[159] J. Silk and G. A. Mamon. The current status of galaxy formation. *Research in Astronomy and Astrophysics*, 12:917–946, Aug 2012. doi: 10.1088/1674-4527/12/8/004.

[160] S. W. Skillman, M. S. Warren, M. J. Turk, R. H. Wechsler, D. E. Holz, and P. M. Sutter. Dark Sky Simulations: Early Data Release. *arXiv e-prints*, art. arXiv:1407.2600, Jul 2014.

[161] O. N. Snaith, C. Park, J. Kim, and J. Rosdahl. Resolution convergence in cosmological hydrodynamical simulations using adaptive mesh refinement. , 477(1):983–1003, June 2018. doi: 10.1093/mnras/sty673.

[162] J. S. Speagle, C. L. Steinhardt, P. L. Capak, and J. D. Silverman. A Highly Consistent Framework for the Evolution of the Star-Forming "Main Sequence" from z ~0-6. , 214(2):15, Oct. 2014. doi: 10.1088/0067-0049/214/2/15.

[163] V. Springel. The cosmological simulation code GADGET-2. , 364: 1105–1134, Dec 2005. doi: 10.1111/j.1365-2966.2005.09655.x.

[164] V. Springel. E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. , 401(2):791–851, Jan. 2010. doi: 10.1111/j.1365-2966.2009.15715.x.

[165] V. Springel and L. Hernquist. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. , 339:289–311, Feb 2003. doi: 10.1046/j.1365-8711.2003.06206.x.

[166] V. Springel and L. Hernquist. The history of star formation in a Λ cold dark matter universe. , 339:312–334, Feb 2003. doi: 10.1046/j.1365-8711.2003.06207.x.

[167] V. Springel, S. D. M. White, G. Tormen, and G. Kauffmann. Populating a cluster of galaxies - I. Results at [formmu2]z=0. , 328:726–750, Dec 2001. doi: 10.1046/j.1365-8711.2001.04912.x.

[168] V. Springel, T. Di Matteo, and L. Hernquist. Modelling feedback from stars and black holes in galaxy mergers. , 361:776–794, Aug 2005. doi: 10.1111/j.1365-2966.2005.09238.x.

[169] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. , 435:629–636, Jun 2005. doi: 10.1038/nature03597.

[170] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. , 435(7042):629–636, June 2005. doi: 10.1038/nature03597.

[171] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White. The Aquarius Project: the subhaloes of galactic haloes. , 391:1685–1711, Dec 2008. doi: 10.1111/j.1365-2966.2008.14066.x.

[172] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, and J. Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. , 475(1):676–698, Mar. 2018. doi: 10.1093/mnras/stx3304.

[173] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, and J. Naiman. First results from the IllustrisTNG simulations: matter and galaxy clustering. , 475(1):676–698, Mar. 2018. doi: 10.1093/mnras/stx3304.

[174] K.-Y. Su, P. F. Hopkins, G. L. Bryan, R. S. Somerville, C. C. Hayward, D. Anglés-Alcázar, C.-A. Faucher-Giguère, S. Wellons, J. Stern, B. A. Terrazas, T. K. Chan, M. E. Orr, C. Hummels, R. Feldmann, and D. Kereš. Which AGN jets quench star formation in massive galaxies? , 507(1):175–204, Oct. 2021. doi: 10.1093/mnras/stab2021.

[175] M. Sugiyama, T. Suzuki, and T. Kanamori. *Density Ratio Estimation in Machine Learning*. Cambridge books online. Cambridge University Press, 2012. ISBN 9780521190176. URL `https://books.google.co.kr/books?id=NOQHkhcFJ0oC`.

[176] S.-I. Tam, K. Umetsu, and A. Amara. Likelihood-free Forward Modeling for Cluster Weak Lensing and Cosmology. *arXiv e-prints*, art. arXiv:2109.09741, Sept. 2021.

[177] A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. Gonçalves, D. Greenberg, and J. Macke. sbi: A toolkit for simulation-based inference. *The Journal of Open Source Software*, 5(52):2505, Aug. 2020. doi: 10.21105/joss.02505.

[178] B. A. Terrazas, E. F. Bell, A. Pillepich, D. Nelson, R. S. Somerville, S. Genel, R. Weinberger, M. Habouzit, Y. Li, L. Hernquist, and M. Vogelsberger. The relationship between black hole mass and galaxy properties: examining the black hole feedback model in IllustrisTNG. , 493(2):1888–1906, Apr. 2020. doi: 10.1093/mnras/staa374.

[179] M. Tremmel, M. Karcher, F. Governato, M. Volonteri, T. R. Quinn, A. Pontzen, L. Anderson, and J. Bellovary. The Romulus cosmolog-

ical simulations: a physical approach to the formation, dynamics and accretion models of SMBHs. , 470(1):1121–1139, Sept. 2017. doi: 10.1093/mnras/stx1160.

[180] M. Tremmel, M. Karcher, F. Governato, M. Volonteri, T. R. Quinn, A. Pontzen, L. Anderson, and J. Bellovary. The Romulus cosmological simulations: a physical approach to the formation, dynamics and accretion models of SMBHs. , 470:1121–1139, Sep 2017. doi: 10.1093/mnras/stx1160.

[181] R. Trotta, G. Jóhannesson, I. V. Moskalenko, T. A. Porter, R. Ruiz de Austri, and A. W. Strong. Constraints on Cosmic-ray Propagation Models from A Global Bayesian Analysis. , 729(2):106, Mar. 2011. doi: 10.1088/0004-637X/729/2/106.

[182] R. B. Tully and J. R. Fisher. Reprint of 1977A&A....54..661T. A new method of determining distance to galaxies. , 500:105–117, Feb. 1977.

[183] A. van der Wel, M. Franx, P. G. van Dokkum, R. E. Skelton, I. G. Momcheva, K. E. Whitaker, G. B. Brammer, E. F. Bell, H. W. Rix, S. Wuyts, H. C. Ferguson, B. P. Holden, G. Barro, A. M. Koekemoer, Y.-Y. Chang, E. J. McGrath, B. Häussler, A. Dekel, P. Behroozi, M. Fumagalli, J. Leja, B. F. Lundgren, M. V. Maseda, E. J. Nelson, D. A. Wake, S. G. Patel, I. Labbé, S. M. Faber, N. A. Grogin, and D. D. Kocevski. 3D-HST+CANDELS: The Evolution of the Galaxy Size-Mass Distribution since z = 3. , 788(1):28, June 2014. doi: 10.1088/0004-637X/788/1/28.

[184] A. Venhola, R. Peletier, E. Laurikainen, H. Salo, E. Iodice, S. Mieske, M. Hilker, C. Wittmann, T. Lisker, M. Paolillo, M. Cantiello, J. Janz, M. Spavone, R. D'Abrusco, G. Ven, N. Napolitano, G. Kleijn, N. Maddox, M. Capaccioli, A. Grado, E. Valentijn, J. Falcón-Barroso, and L. Limatola. The Fornax Deep Survey with the VST. IV. A size and magnitude limited catalog of dwarf galaxies in the area of the

Fornax cluster. , 620:A165, Dec. 2018. doi: 10.1051/0004-6361/201833933.

[185] F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, D. N. Spergel, R. S. Somerville, R. Dave, A. Pillepich, L. Hernquist, D. Nelson, P. Torrey, D. Narayanan, Y. Li, O. Philcox, V. La Torre, A. Maria Delgado, S. Ho, S. Hassan, B. Burkhart, D. Wadekar, N. Battaglia, G. Contardo, and G. L. Bryan. The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations. , 915(1):71, July 2021. doi: 10.3847/1538-4357/abf7ba.

[186] P. Villanueva-Domingo, F. Villaescusa-Navarro, S. Genel, D. Anglés-Alcázar, L. Hernquist, F. Marinacci, D. N. Spergel, M. Vogelsberger, and D. Narayanan. Weighing the Milky Way and Andromeda with Artificial Intelligence. *arXiv e-prints*, art. arXiv:2111.14874, Nov. 2021.

[187] M. Vogelsberger, S. Genel, D. Sijacki, P. Torrey, V. Springel, and L. Hernquist. A model for cosmological simulations of galaxy formation physics. , 436(4):3031–3067, Dec. 2013. doi: 10.1093/mnras/stt1789.

[188] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson, and L. Hernquist. Properties of galaxies reproduced by a hydrodynamic simulation. , 509(7499):177–182, May 2014. doi: 10.1038/nature13316.

[189] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson, and L. Hernquist. Properties of galaxies reproduced by a hydrodynamic simulation. , 509:177–182, May 2014. doi: 10.1038/nature13316.

[190] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in

the Universe. , 444(2):1518–1547, Oct. 2014. doi: 10.1093/mnras/stu1536.

[191] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. , 444:1518–1547, Oct 2014. doi: 10.1093/mnras/stu1536.

[192] Z. Wang, H. Xu, X. Yang, Y. Jing, K. Wang, H. Guo, F. Dong, and M. He. The clustering of galaxies in the DESI imaging legacy surveys DR8: I. The luminosity and color dependent intrinsic clustering. *Science China Physics, Mechanics, and Astronomy*, 64(8):289811, Aug. 2021. doi: 10.1007/s11433-021-1707-6.

[193] W. A. Watson, I. T. Iliev, A. D'Aloisio, A. Knebe, P. R. Shapiro, and G. Yepes. The halo mass function through the cosmic ages. , 433(2): 1230–1245, Aug 2013. doi: 10.1093/mnras/stt791.

[194] R. Weinberger, V. Springel, L. Hernquist, A. Pillepich, F. Marinacci, R. Pakmor, D. Nelson, S. Genel, M. Vogelsberger, J. Naiman, and P. Torrey. Simulating galaxy formation with black hole driven thermal and kinetic feedback. , 465(3):3291–3308, Mar. 2017. doi: 10.1093/mnras/stw2944.

[195] R. Weinberger, V. Springel, L. Hernquist, A. Pillepich, F. Marinacci, R. Pakmor, D. Nelson, S. Genel, M. Vogelsberger, J. Naiman, and P. Torrey. Simulating galaxy formation with black hole driven thermal and kinetic feedback. , 465:3291–3308, Mar 2017. doi: 10.1093/mnras/stw2944.

[196] R. Weinberger, V. Springel, R. Pakmor, D. Nelson, S. Genel, A. Pillepich, M. Vogelsberger, F. Marinacci, J. Naiman, P. Torrey, and L. Hernquist. Supermassive black holes and their feedback effects in the IllustrisTNG simulation. , 479:4056–4072, Sep 2018. doi: 10.1093/mnras/sty1733.

[197] R. Weinberger, V. Springel, and R. Pakmor. The AREPO Public Code Release. , 248(2):32, June 2020. doi: 10.3847/1538-4365/ab908c.

[198] S. D. M. White, M. Davis, G. Efstathiou, and C. S. Frenk. Galaxy distribution in a cold dark matter universe. , 330:451–453, Dec 1987. doi: 10.1038/330451a0.

[199] S. D. M. White, C. S. Frenk, M. Davis, and G. Efstathiou. Clusters, Filaments, and Voids in a Universe Dominated by Cold Dark Matter. , 313:505, Feb 1987. doi: 10.1086/164990.

[200] R. P. C. Wiersma, J. Schaye, and B. D. Smith. The effect of photoionization on the cooling rates of enriched, astrophysical plasmas. , 393: 99–107, Feb 2009. doi: 10.1111/j.1365-2966.2008.14191.x.

[201] R. P. C. Wiersma, J. Schaye, T. Theuns, C. Dalla Vecchia, and L. Tornatore. Chemical enrichment in cosmological, smoothed particle hydrodynamics simulations. , 399:574–600, Oct 2009. doi: 10.1111/j.1365-2966.2009.15331.x.

[202] Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. , 5:84–89, Mar. 1970.

[203] X. Zhao, Y. Mao, C. Cheng, and B. D. Wandelt. Simulation-based Inference of Reionization Parameters from 3D Tomographic 21 cm Light-cone Images. , 926(2):151, Feb. 2022. doi: 10.3847/1538-4357/ac457d.

[204] X. Zhao, Y. Mao, and B. D. Wandelt. Implicit Likelihood Inference of Reionization Parameters from the 21 cm Power Spectrum. *arXiv e-prints*, art. arXiv:2203.15734, Mar. 2022.

[205] E. Zinger, A. Pillepich, D. Nelson, R. Weinberger, R. Pakmor, V. Springel, L. Hernquist, F. Marinacci, and M. Vogelsberger. Ejective and preventative: the IllustrisTNG black hole feedback and its effects on the thermodynamics of the gas within and around galaxies. , 499(1):768–792, Nov. 2020. doi: 10.1093/mnras/staa2607.

국문초록

수십 년 동안의 우주론적 모의 실험과 관측의 눈부신 발전은 우주 거대 구조의 형성과 진화와 같은 다양한 천체 물리학 및 천문 현상에 대한 이해를 눈부시게 확장시켰고, 우주론적 모의 실험은 천체 물리학을 연구하는 데에 없어서는 안될 필수불가결한 요소가 되었습니다. 그러나 관측과 모의 실험의 다양한 불확실성, 관측과 모의 실험의 물리적 모델의 한계, 모의 실험의 계산 비용, 수학적으로 정교한 비교의 방법론적 부재와 같은 여러 요인으로 인해 모의 실험과 관측은 융합되어 연구되지 못하였습니다. 한편, 컴퓨터의 발전과 함께 부상한 기계 학습은 위에서 제기된 대부분의 문제를 해결할 수 있을 만큼의 잠재력을 보여주었습니다. 이 논문에서는 기계 학습을 활용하여 크게 두 가지 문제를 해결하는 것을 목표로 합니다. (1) 계산적으로 매우 효율적인 방식으로 우주론적 유체역학 모의 실험을 지원할 수 있는 모델을 수립하고, (2) 수학적으로 보다 정교한 방법으로 관측에 대해 우주론적 모의 실험을 보정하는 골조를 구축하였습니다.

첫 번째 부분에서는 기계 학습을 사용하여 고해상도 유체 역학 모의 실험으로 암흑 물질으로만 이루어진 암흑 물질 전용 모의 실험에서 암흑 물질 헤일로의 중입자적 특성(즉, 항성이나 기체 등의 특성)을 추정하는 모델을 제시합니다. 모델을 훈련 및 개선시키기 위해 무작위 숲 알고리즘과 함께 두단계 학습, 개선된 오차 함수, 헤일로의 주변 환경과 성정 과정 등을 추가적인 입력값으로 사용하였습니다. 이러한 개선 사항들을 통해 저희 모델의 중입자적 특성을 예측하는 정확도가 크게 향상되었음을 보여줍니다. 이런 모델을 아주 큰 암흑 물질 전용 모의 실험에 적용하여 암흑 물질의 정보로만 이루어진 암흑 물질 헤일로 목록에서 중입자적 특성까지 포함할 수 있는 은하 목록을 기존 모의 실험에 비해 비약적으로 짧은 시간 안에 생성할 수 있는 골조을 구축하였습니다. 저희 모델은 은하 규모의 유체역학 모의 실험의 중입자 물리학을 큰 규모의 암흑 물질 전용 모의 실험에 효율적으로 이식할 수 있는 획기적인 방법이 될 수 있습니다. 또

한, 기계 학습의 장점과 이러한 접근 방식을 이용하여 우주론적 현상들에 대한 물리학적 이해를 증대하기 위한 방법들에 대해 논의하였습니다.

두 번째 부분에서는 우도함수를 사용하지 않는 추론이라고도 하는 모의 실험 기반 추론을 사용하여, 기존에는 모의 실험의 높은 계산 비용으로 실현할 수 없었던, 관측에 대한 우주론적 모의 실험의 보정을 수행하였습니다. 계산 효율성을 위해 우리는 약 천여개의 우주론적 모의 실험을 사용하여 에뮬레이터를 훈련하여 우주론 및 천체 물리학 매개변수를 입력으로 받아 모의 실험의 관측량을 추정하고, 이러한 에뮬레이터를 우주론 모의 실험의 대체제로 사용합니다. 우주 항성 형성 속도 밀도 및 항성 질량 함수를 사용하여 선택된 우주 및 천체 물리학 매개변수($\Omega_{\rm m}$, σ_8, 초신성 폭발 피드백 및 블랙홀 피드백)에 대해 모의 실험 기반 추론을 수행하여 6차원 사후 확률 분포를 얻습니다. 저희는 우주 항성 형성 속도 밀도에서 추론된 매개변수 사이에 축퇴가 존재한다는 것을 발견했으며, 이는 새로운 우주론적 모의 실험으로도 확인되었습니다. 또한, 항성 질량 함수가 우주 항상 형성 속도 밀도의 축퇴를 깨뜨릴 수 있음을 발견했습니다. 이는 항성 질량 함수가 매개변수에 대해 더 강력한 제약을 제공할 수 있음을 나타냅니다. 저희는 관측에서 추론된 항성 형성 속도 밀도에서 추론된 매개변수 집합이 관측된 항성 형성 속도 밀도를 매우 잘 재현하는 반면, 항성 질량 함수의 경우, 추론 및 관찰된 항성 질량 함수의 상당한 불일치로부터 은하 형성 모델링의 한계가 있다는 것을 발견했습니다.

주요어 : 우주론 모의 실험, 모의 실험 기반 추론, 은하 형성과 진화, 우주의 거대 구조 형성
학번 : 2014-22379