

Creating halos with autoregressive multistage networks

Shivam Pandey¹, Chirag Modi^{2,3}, Benjamin D. Wandelt^{3,4}, Deaglan J. Bartlett⁴, Adrian E. Bayer^{5,3},
Greg L. Bryan¹, Matthew Ho^{1,4}, Guilhem Lavaux⁴, T. Lucas Makinen⁶, and Francisco Villaescusa-Navarro^{3,5}

¹*Columbia Astrophysics Laboratory, Columbia University,
550 West 120th Street, New York, New York 10027, USA*

²*Center for Computational Mathematics, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA*

³*Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA*

⁴*CNRS and Sorbonne Université, Institut d'Astrophysique de Paris (IAP),
UMR 7095, 98 bis boulevard Arago, F-75014 Paris, France*

⁵*Department of Astrophysical Sciences, Princeton University,
4 Ivy Lane, Princeton, New Jersey 08544, USA*

⁶*Imperial Centre for Inference and Cosmology (ICIC), Imperial College London,
Prince Consort Road, London SW7 2AZ, United Kingdom*



(Received 19 December 2024; accepted 16 October 2025; published 5 November 2025)

To maximize the amount of information extracted from cosmological datasets, simulations that accurately represent these observations are necessary. However, traditional simulations that evolve particles under gravity by estimating particle-particle interactions (N -body simulations) are computationally expensive and prohibitive to scale to the large volumes and resolutions necessary for the upcoming datasets. Moreover, modeling the distribution of galaxies typically involves identifying virialized dark matter halos, which is also a time- and memory-consuming process for large N -body simulations, further exacerbating the computational cost. In this study, we introduce CHARM, a novel method for creating mock halo catalogs by matching the spatial, mass, and velocity statistics of halos directly from the large-scale distribution of the dark matter density field. We develop multistage neural spline flow-based networks to learn this mapping at redshift $z = 0.5$ directly with computationally cheaper low-resolution particle mesh simulations instead of relying on the high-resolution N -body simulations. We show that the mock halo catalogs and painted galaxy catalogs have the same statistical properties as obtained from N -body simulations in both real space and redshift space. Finally, we use these mock catalogs for cosmological inference using redshift-space galaxy power spectrum, bispectrum, and wavelet-based statistics using simulation-based inference, performing the first inference with accelerated forward model simulations and finding unbiased cosmological constraints with well-calibrated posteriors.

DOI: [10.1103/vlm2-tm6k](https://doi.org/10.1103/vlm2-tm6k)

I. INTRODUCTION

The standard model of cosmology describes the evolution of the Universe using a set of free cosmological parameters. Constraining these parameters with observations of the Universe is one of the primary goals of cosmological studies [1–4]. Over approximately 13.7 billion years of evolution, the hierarchical structure formation process transforms an initially Gaussian distribution of matter into a highly non-Gaussian field comprising halos, voids, and filaments. The observed galaxies occupy the collapsed and bound structures of dark matter called halos. The cosmological parameters can be constrained by analyzing the statistical distribution of the observed galaxies and comparing them to predictions from theoretical models or simulations. Traditional techniques limit this comparison to simple two-point summary statistics, such as the power spectrum at large scales, as theoretical models break down

for higher-order statistics and nonlinear small scales [5–8]. Since the evolved matter distribution is non-Gaussian, higher-order statistics, as well as small-scale two-point correlations, carry a significant amount of complementary information about the cosmological parameters [9–16].

To extract this information, we need to rely on accurate N -body simulations as forward models and employ simulation-based inference (SBI) techniques [12,17–20]. SBI involves using computational forward models to simulate the data for a set of cosmological parameters, measuring the statistics of interest, and comparing them with the observed data using machine-learning techniques to constrain the parameters (see [21] for a review). For galaxy clustering surveys, these forward models involve evolving the dark matter particles under gravity, identifying the dark matter halos, and then populating these halos with galaxies. However, simulating the high-mass halos (which are very rare) in various environmental conditions requires a large

simulation box, while simulating the lower halo masses demands high resolution. Particularly, to reliably analyze the current generation of galaxy surveys, the number of particles and the volume of the simulation required are so large that the computational cost of running these simulations at a grid of cosmological parameters is prohibitive. To put things in context, analyzing the last generation of cosmological surveys, which ended a decade ago, with this approach would require running at least 2000 simulations with 2.7×10^{10} particles, taking more than 270 million CPU hours for running the simulations alone [22]. Furthermore, finding the halos accurately, which requires processing phase space information of particles [23] in these N -body simulations, also adds to the computational cost.

However, physically, we expect the number, mass, and velocity distributions of halos to depend on the large-scale matter distribution (see [24] for a review). For example, the overdense regions of the Universe will have more matter to collapse and will be able to form more numerous and massive halos. Therefore, accurately learning this relationship and generating fast approximations of the dark matter distribution on large scales can accelerate mock halo catalog generation and ultimately generate observed data with end-to-end simulations. This motivates us to use deep-learning techniques to learn these highly nonlinear and nonlocal [25] relationships between the dark matter and halo distributions.

To accelerate the simulations on large scales, particle mesh (PM) approximations can be used (e.g., [26,27]). These approximations estimate the gravitational forces by interpolating cold dark matter particles on a uniform grid, enabling the use of techniques such as fast Fourier transforms to solve the equations of motion. Due to this grid interpolation, they lose information on scales smaller than the grid resolution, resulting in poor halo catalogs [28,29], particularly for low-mass halos. However, on scales larger than the grid size, they accurately capture the matter distribution. Since these PM simulations are orders of magnitude faster than N -body simulations, our goal is to learn the relationship between N -body halos and matter density fields obtained from paired (ran with same initial conditions) low-resolution PM simulations.

In this work, we introduce CHARM: a generative model for creating halo catalogs using multistage neural spline flows to transform the low-resolution PM simulations to discrete mock catalogs expected from a high-resolution N -body simulation. It consists of four stages after extracting features from the surrounding dark matter density in PM simulations at any location: (i) learn the number of halos expected, (ii) learn the mass of the heaviest halo, (iii) autoregressively learn lower halo masses, and (iv) autoregressively learn the 3D halo velocities. Developing a methodology like this which is accurate, is fast enough to be used for cosmological inference, and is scalable to larger volumes is crucial

for using simulation-based inference techniques to analyze large-volume data and maximize the information gained from current and future galaxy surveys, a key goal of, for example, the Learning the Universe Collaboration [30]. We will test the performance of our model across all of these three axes—accuracy, inference, and scalability for the analysis of galaxy survey data.

In recent years, there have been other studies with related goals, but they provide different solutions compared to what is desired here. In [31,32], a similar mapping is learned using physically motivated networks with a significantly reduced number of free parameters, but they assume an explicit form of the likelihood for halo occupation, which breaks down for high-mass halos and small scales, which are of interest in this study. Moreover, these models have been trained on a fixed cosmology simulation in real space. In contrast, we aim to obtain a forward model that generalizes to different cosmologies and to redshift space. There have also been attempts, as in [33,34], that do not impose a likelihood form. However, the methodology of [33] only works for continuous biased scalar fields like total halo mass, whereas here we aim to obtain discrete halo catalogs. The methodology of [34] is designed to work only with dark matter density obtained from high-resolution N -body simulations, thus requiring large computational resources.

In [35,36], displacement corrections to PM simulations were provided to make them resemble their N -body counterparts. However, these corrections, applied at the particle level, only improve the statistics of relatively high mass halos ($M > 10^{14} M_{\odot}/h$), are calibrated on relatively low-resolution simulations, and are expensive to train. Note that the corrections can be improved if additional force evaluations are included [37]. In [29], the authors present methods to correct PM simulations, leading to a faithful reproduction of halos even on small scales. However, they require high-resolution PM runs, whereas here we focus on obtaining N -body-like halo catalogs while using low-resolution PM simulations that can be scaled to large volumes. Nevertheless, these corrections could be used to augment the PM simulations used here and improve the accuracy of the model in the future.

The paper is organized as follows: in Sec. II we describe the simulation dataset used, in Sec. III we describe the architecture of our model, in Sec. IV we describe the setup for evaluating the performance of the model, in Sec. V we show the results, and we finally conclude in Sec. VI. A table of the run-time of the simulation products used in this study as well as for the acronyms used in this study is given in Appendix. B.

II. DATASET

A. Simulations

We use the public simulation suite from the N -body Quijote project [22], which simulates a volume of

$(1000 \text{ Mpc}/h)^3$. These simulations cover a wide range of cosmologies and have enough volume and resolution to provide reliable dark matter halo catalogs with masses above $\sim 5 \times 10^{12} M_\odot/h$. Therefore, these simulations provide a good suite to build a reliable model of how the distribution of halos relates to the surrounding dark matter field, which can then be applied to significantly larger volume simulations. In this work, we process the high-resolution Latin hypercube (LH) and fiducial cosmology suite of simulations, varying five cosmological parameters: matter density Ω_m , baryon density Ω_b , matter clustering strength σ_8 , primordial power spectrum tilt n_s , and the expansion rate of the Universe h . These five cosmological parameters are varied with a wide uniform prior of $\Omega_m \in \mathcal{U}[0.1, 0.5]$, $\sigma_8 \in \mathcal{U}[0.6, 1.0]$, $\Omega_b \in \mathcal{U}[0.03, 0.07]$, $n_s \in \mathcal{U}[0.8, 1.2]$, and $h \in \mathcal{U}[0.5, 0.9]$. In this study, we use only the density fields and halo catalogs at $z = 0.5$. We use subvolumes from 1800 LH simulations and one full fiducial cosmology simulation for training while reserving the remaining 200 LH simulations for testing. Each N -body simulation evolves an initial Gaussian distribution of 1024^3 cold dark matter particles to the present time and takes approximately 5000 CPU hours to complete.

B. Input dark matter fields

For the approximate PM simulations, which form our input, we run paired simulations (i.e., matching Gaussian initial conditions and cosmology to the Quijote suite) using the FastPM algorithm [26]. Here, we evolve only 384^3 particles over the same volume, and each simulation takes only 5 CPU hours.¹ We compute the matter density fields (ρ_m) and 3D matter velocity fields ($[v_m^x, v_m^y, v_m^z]$) from the PM simulations on a regular grid with 128^3 voxels using cloud-in-cell interpolation. Note that each voxel has a physical size of approximately $7.8 \text{ Mpc}/h$. It is also important to note that, since the size of a typical halo is less than $1 \text{ Mpc}/h$, each voxel can host multiple halos. We concatenate the density and velocity fields, which form the input to our model.

C. Target halo catalog

From the N -body simulations, we use the halo catalog obtained by applying the Rockstar halo finder to the particle distribution. The algorithm uses the 6D phase space information of the particles to identify particles in collapsed structures. We use halos with masses $M_{200c} \geq 5 \times 10^{12} M_\odot/h$ in the catalog, which is the resolution limit of the Quijote simulations. Note that the spherical overdensity mass M_{200c} is the mass within r_{200c} of the halo which is defined such that the average enclosed density

within a sphere of radius r_{200c} is equal to 200 times the critical density of the Universe, $\rho_c(z)$. The low-resolution PM simulations used here do not account for small-scale particle-particle interactions, resulting in biased small-scale matter velocity and density fields. As Rockstar relies on an accurate small-scale velocity distribution of matter particles, it heavily underestimates the distribution of low-mass halos [28,29]. However, the spatial distribution and velocity of these halos are correlated with the large-scale fields (which are correctly captured by the PM simulations; e.g., see [40]), and we aim to learn this mapping using CHARM.

We voxelize our target halo distribution on the same 128^3 grid using the nearest-grid point (NGP) mass assignment scheme. Within this scheme, for each voxel i , we count the number of halos inside it (N_{tot}^i). If it is nonzero, we also store the halo masses in decreasing order ($[M_1^i, M_2^i, \dots, M_{N_{\text{tot}}^i}^i]$, where $M_1^i > M_2^i > \dots > M_{N_{\text{tot}}^i}^i$). In our training set, we have at most approximately 12 halos in any voxel, so we fix $N_{\text{max}} = 12$, and $N_{\text{tot}}^i \leq N_{\text{max}}$ for all i .

While the PM simulations underestimate the small-scale velocity distribution, they capture the large-scale coherent velocity fields correctly. Therefore, we estimate the 3D PM velocity for each halo by linearly interpolating the 3D matter velocity field from PM simulations ($\vec{v}_{\text{FastPM}}^i$) and aim to estimate the difference, $\Delta \vec{v} = \vec{v}_{\text{Quijote}} - \vec{v}_{\text{FastPM}}^i$, for each halo in the voxel. We order the halos in the same order as for masses, going from heaviest to least heavy.

D. Batching

We divide the 3D simulation boxes, each of size 128^3 , into subboxes of size 16^3 , resulting in 512 subboxes from each simulation. Each subbox has a physical size of $125 \text{ Mpc}/h$. We select the corresponding subboxes from the PM set to create the input dark matter density field, paired with the N -body halos for training. To facilitate feature extraction, we pad the input density field from the PM subboxes so that the output after convolutions preserves the size of the subbox (16^3). Therefore, we add a padding of four voxels on each side from the original periodic simulations.

For training, we use one full simulation (512 subboxes) at the fiducial cosmology while using only 16 randomly selected subboxes for 2000 simulations with varying cosmologies in the LH grid (using only about $\sim 3\%$ of the available volume). This is done primarily due to the memory limitations of GPUs and to accelerate the training process. With one full simulation at a fiducial cosmology, the network learns the dependence of halo properties on the environment at a fixed cosmology, while with the limited data from the LH set, the network learns the variation of properties with cosmological parameters.

¹GPU implementations of these algorithms can further increase computational efficiency [38,39].

III. METHODOLOGY

The task of this paper is to obtain a discrete mock halo catalog when provided with approximate dark matter overdensity and velocity fields from PM simulations. Halo formation is a complex nonlinear process that depends on the 3D matter distribution on large scales. For example, there is a higher probability of forming a heavy halo at the intersection of large dark matter filaments compared to in a void region. Therefore, to extract the features of dark matter density and velocity fields that correlate with the halo distribution, we stack two 3D residual network (ResNet) layers [41]. We input the dark matter density and 3D velocity fields as described in Sec. II as four channels and extract 20 features from ResNet layers. These features, extracted from a physical region of approximately 70 Mpc/ h , are used as conditioning for a multilevel generative model for the halo distribution, as described below. Note that we also append the values of the five cosmological parameters for each simulation subbox to these 20 features.

To create a mock halo catalog and its masses, we need to estimate a discrete distribution of halos conditioned on a feature vector for each voxel. To achieve this, we split the problem into three parts. For each voxel i , we first predict the total number of halos (N_{tot}^i), which provides a mask as well as an occupation number to train the mass distribution. Then, for the voxels that have a nonzero number of halos, we predict the mass of the heaviest halo (M_1^i). Afterward, for voxels that have more than one halo, we train the prediction for the masses of lighter halos ($M_2^i, \dots, M_{N_{\text{tot}}^i}^i$) in an autoregressive fashion. This means that, as dictated by the physics of structure formation, we always condition the probability of lighter halo masses on the masses of all the heavier halos in the same voxel. Our final loss function is a sum of the losses from these three steps and the details of each step are as follows (also see Fig. 1 for a summary of this inference pipeline).

- (i) We model the probability distribution of the total number of halos as a mixture of N_{max} Gaussians. We take the discrete distribution of the number of halos in each voxel of the training simulations and make it continuous by adding a small Gaussian noise with a known variance ($\sigma^2 = 0.0025$). With input from the learned features of the ResNet, we predict the probability of each Gaussian using a fully connected neural network (FCNN).² In this case, the loss is modeled as the forward Kullback-Leibler (KL) divergence [42] between the modeled Gaussian

²The FCNN layer is a stack of three linear layers interspersed with two nonlinear layers that use the hyperbolic tangent as the activation function. The first linear layer takes the input dimension of the features and maps it to the hidden dimension (here equal to 8) and the third linear layer maps the hidden dimension to the output dimension.

mixture with predicted probabilities (with fixed mean and variance) and the true distribution. Note that the KL divergence is a measure of how one probability distribution $q(x)$ differs from a reference distribution $p(x)$, quantifying the information lost when $q(x)$ is used to approximate $p(x)$. It is defined as $D_{\text{KL}}(p||q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx$ for continuous distributions.

- (ii) To model the mass of the heaviest halo, we first transform a base distribution using a stack of five spline transformations [43] with eight knots. The base distribution here is not the standard Gaussian, as is often used with normalizing flows, but the probability distribution function estimated from the unconditional halo mass function, as described in [44] (see the upper-central black-dashed histogram in Fig. 1). We found that using this physically motivated base distribution was crucial for obtaining accurate predictions for the heaviest halo. The parameters of the transformations are learned using a separate FCNN. The loss is then calculated as the KL divergence between a known base distribution and this transformed distribution.
- (iii) We learn the distribution of lower halo masses using an autoregressive neural spline flow [43]. We condition the transformation on the masses of all the heavier halos. Additionally, to ensure a decreasing order of halo masses, for the j th halo in the voxel, we learn the mass difference $M_{j-1} - M_j$ and ensure that this difference is positive. Here, we stack two such spline flows, and their parameters are once again learned using an FCNN. As having more halos in the same voxel becomes a rarer phenomenon, we model the base distribution as proportional to the Gumbel distribution [45], which provides a good initial estimate of extreme value statistics (see the bottom-central histogram in Fig. 1).

Finally, we train an independent autoregressive network to obtain the velocities of the halos conditioned on the features extracted from the PM simulation. As described in Sec. II, we use the difference between the true halo velocity and its interpolated value estimated from the PM simulation. The PM simulation velocity fields smoothed at ~ 8 Mpc/ h resolution already capture the large-scale coherent velocity component of the halos. Therefore, with this network, we learn the remaining small-scale component of the velocities (related to the fingers-of-god effect [46]). We start with a Gaussian distribution as our base distribution and use a stack of two neural spline flows to learn the 3D velocity difference of halos at each voxel.

Once trained, this model can then be used to obtain a catalog of mock halo positions in real space (\vec{x}_h), their masses, and their velocities (\vec{v}_h) on test simulations by inputting the corresponding PM density and velocity fields. To bring the model closer to the observations, we need to

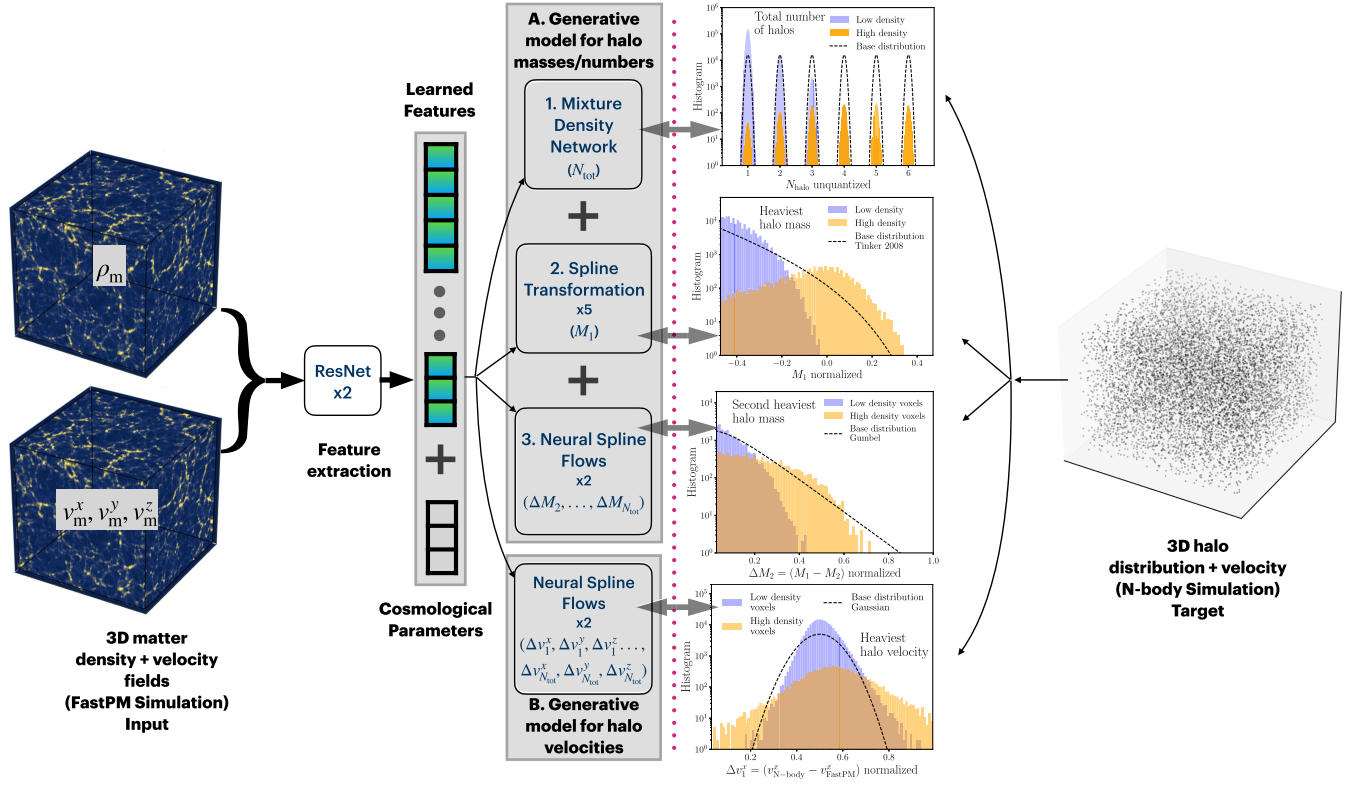


FIG. 1. Visualization of the data products and network architecture used in this study. On the left side of the dotted line, we show the matter density and velocity fields from the PM simulation that serve as the input to the ResNet layers to extract features. These features are then used to predict the halo distribution and velocities in four parts: the total number of halos is modeled using a mixture density network, the heaviest halo mass is modeled using a stack of spline transformations, and lower halo masses as well as halo velocities are modeled using a stack of autoregressive neural spline flows. On the right side, we display the target distribution of the halos from paired true N -body simulation. Additionally, we present histograms of the four quantities in both low- and high-density voxels, illustrating their dependence on the dark matter distribution.

place the objects in redshift space. With the knowledge of \vec{x}_h and \vec{v}_h , we can calculate the position of halos in redshift space (\vec{s}_h) with $\vec{s}_h = \vec{x}_h + (\eta_{\text{los}}) \cdot (\vec{v}_h \times (1+z)/H(z))$, where η_{los} is the line of sight direction (which we assume to be parallel to the x axis of the box in this study) and $H(z)$ is the Hubble constant at redshift z .

IV. DIAGNOSTICS AND INFERENCE SETUP

A. Statistical summaries

Once the model is trained, we obtain mock halo and galaxy catalogs on test LH simulations and compare their performance by calculating various statistics of interest for cosmological studies. In addition to comparing the environment-dependent histograms of mock and true halo catalogs, we also compare the higher-order statistical properties. We calculate three different statistics that are sensitive to two-point, three-point and higher-order correlations between the galaxies.

1. Power spectrum

We calculate the real-space 3D power spectrum on 16 scales between $k_{\text{min}} = 0.01$ and $k_{\text{max}} = 0.32$ and use the

nearest-grid point estimator to calculate the density fields. We use the public PYLIANS code to estimate the power spectra [47]. In redshift space, we calculate and compare the power spectra multipoles $P_\ell(k)$ for $\ell \in [0, 2, 4]$.

2. Bispectrum

We calculate the bispectrum for isosceles triangles for three different scales, $k \in [0.08, 0.16, 0.32]h/\text{Mpc}$ and eight different values of angle between the equal edges, ranging between 0.1 and 3.04 rad. These are also calculated using the PYLIANS code. In redshift space, we only calculate and compare the monopole of bispectrum $B_{\ell=0}(k)$.

3. Wavelets

We follow [48] for calculating the wavelet scattering statistics using their public code [49]. Given the 3D density field (ρ) constructed from the mock galaxy or halo catalogs, we calculate three different wavelet coefficients by convolving the halo field with various wavelets and evaluating the moments:

$$S_0 = \langle |\rho|^p \rangle, \quad (1)$$

$$S_1 = \langle |\psi_{\lambda_0} * \rho|^p \rangle, \quad (2)$$

$$S_2 = \langle \langle |\psi_{\lambda_1} * \rho| * \psi_{\lambda_2}|^p \rangle \rangle, \quad (3)$$

where ψ_{λ_j} is the wavelet, λ_j is its index, and p is the exponent. The wavelets are constructed from a directional mother wavelet which is defined in Fourier space. Its form is localized in real space and has a cutoff scale (k_c) in Fourier space. We set $k_c = 4\pi k_{\max}/3$, where we again use $k_{\max} = 0.32 h/\text{Mpc}$. We set the number of octaves (doublings of scales) to $J = 3$ and set $Q = 4$ which controls the number of scales per octave. We evaluate the wavelet coefficients for three different values of exponent $p \in [1/2, 1, 2]$.

B. Galaxies

Instead of dark matter halos, we typically observe galaxies and so we need to predict the statistics of galaxies from a given halo catalog. There are several ways to establish the halo-galaxy connection ranging from simple empirical fitting functions [50–55] to more physical models (see [56] for a review) that describe the galaxy formation through cosmic times using hydrodynamical simulations (e.g., [57,58]) or semianalytic models (e.g., [59]). These methods have a wide range of applicability and complexity and in this study where we aim to target spectroscopic red galaxy catalogs, such as the CMASS galaxy sample of the Sloan Digital Sky Survey (SDSS-CMASS), we choose the empirical halo occupation distribution (HOD) approach. This approach has been successfully applied in several analyses for both standard two-point power spectra (e.g., [60]) and higher-order correlations (e.g., [48,61]). In this first study to use an accelerated forward model for halos, we use a simple five-parameter HOD model [55] which assumes that the occupation probability of galaxies in any halo depends only on its mass. We defer the study with a more realistic halo-galaxy connection to a future investigation.

The [55] HOD model parametrizes the number of central (N_c) and satellite (N_s) galaxies as a function of halo mass with five free parameters: $\log(M_{\min})$, $\sigma_{\log M}$, $\log(M_0)$, $\log(M_1)$, and α_{sat} . We sample ten random values for the HOD parameters for each LH simulation, resulting in 20 000 mock galaxy catalogs. The parameters $\sigma_{\log M}$ and α_{sat} are sampled with wide uniform priors: $\sigma_{\log M} \in \mathcal{U}[0.3, 0.5]$ and $\alpha_{\text{sat}} \in \mathcal{U}[0.3, 0.5]$. The other three HOD parameters are constrained to vary with the cosmology ($\vec{\theta}_{\text{cosmo}}$) of each LH simulation: $\log(M_{\min}) \in \mathcal{U}[\log(M_{\min}^{\vec{\theta}_{\text{cosmo}}}) \pm 0.15]$, $\log(M_0) \in \mathcal{U}[\log(M_0^{\vec{\theta}_{\text{cosmo}}}) \pm 0.2]$, and $\log(M_1) \in \mathcal{U}[\log(M_1^{\vec{\theta}_{\text{cosmo}}}) \pm 0.3]$. The cosmology-dependent central values of the priors for these parameters [$\log(M_{\min}^{\vec{\theta}_{\text{cosmo}}})$, $\log(M_0^{\vec{\theta}_{\text{cosmo}}})$, and $\log(M_1^{\vec{\theta}_{\text{cosmo}}})$] are fixed such that the number density of the mock galaxy catalog

approximately matches the CMASS sample comoving galaxy density, $\bar{n}_g \sim 3 \times 10^{-4} (\text{Mpc}/h)^{-3}$, and the satellite fraction $f_{\text{sat}} \sim 0.2$ [60].

Assuming \bar{n}_g and f_{sat} , we estimate the expected number of central (\bar{N}_c) and satellite (\bar{N}_s) galaxies. Then, we calculate the minimum halo mass ($\log M_h^{\vec{\theta}_{\text{cosmo}}}$) that results in \bar{N}_c halos, each hosting a central galaxy, and set $\log(M_{\min}^{\vec{\theta}_{\text{cosmo}}}) = \log(M_0^{\vec{\theta}_{\text{cosmo}}}) = \log M_h^{\vec{\theta}_{\text{cosmo}}}$. Finally, assuming a fiducial value of $\alpha_{\text{sat}} = 0.7$, we determine $\log(M_1^{\vec{\theta}_{\text{cosmo}}})$ that results in \bar{N}_s satellite galaxies.

We assume that the central galaxy, which is placed at the center, has the same velocity as the parent halo, whereas the satellite galaxies are distributed around the halo following the Navarro-Frenk-White profile [62] and receive an additional velocity contribution due to their virial motion. We place the galaxies at the real-space positions of the halos, assign their velocities, and then move them to redshift space. We then measure the statistical summaries from this distribution as described below.

However, note that CHARM does not resolve the positions of the halos within the 8 Mpc/h voxel, so it places them at the voxel centers. To account for this lack of knowledge of galaxy positions on smaller scales, we apply smoothing to the positions of the galaxies in redshift space by adding a random Gaussian scatter to their 3D positions with a standard deviation of $\sigma_G = 8 \text{ Mpc}/h$. We apply this same smoothing to the true galaxy catalogs in redshift space generated from the test Quijote simulations, against which we compare the performance of the CHARM model.

C. Inference setup

We use neural posterior estimation (NPE) to obtain the posteriors on (cosmological and HOD) parameters of our forward model of the statistics of galaxies in redshift space. The NPE algorithm trains a conditional normalizing flow [$q_\phi(\vec{\theta}|\vec{x}_{\text{data}})$], where $\vec{\theta}$ are the model parameters, \vec{x}_{data} is the data vector of summary statistics compressed using an embedding network and ϕ are the hyperparameters of the normalizing flow model and embedding networks. We use the standard masked autoregressive architecture [63] for normalizing flow which approximates the true posterior distribution [$p(\vec{\theta}, \vec{x})$] by approximating it with a sequence of autoregressive affine transformations of a base distribution, with known probability density (here a standard Gaussian). The flow is trained by minimizing the KL divergence:

$$\mathcal{L}(\phi) = D_{\text{KL}}[p(\vec{\theta}, \vec{x}) || q_\phi(\vec{\theta}|\vec{x}_{\text{data}})p(\vec{x}_{\text{data}})], \quad (4)$$

which can be shown to be equivalent to maximizing the training score:

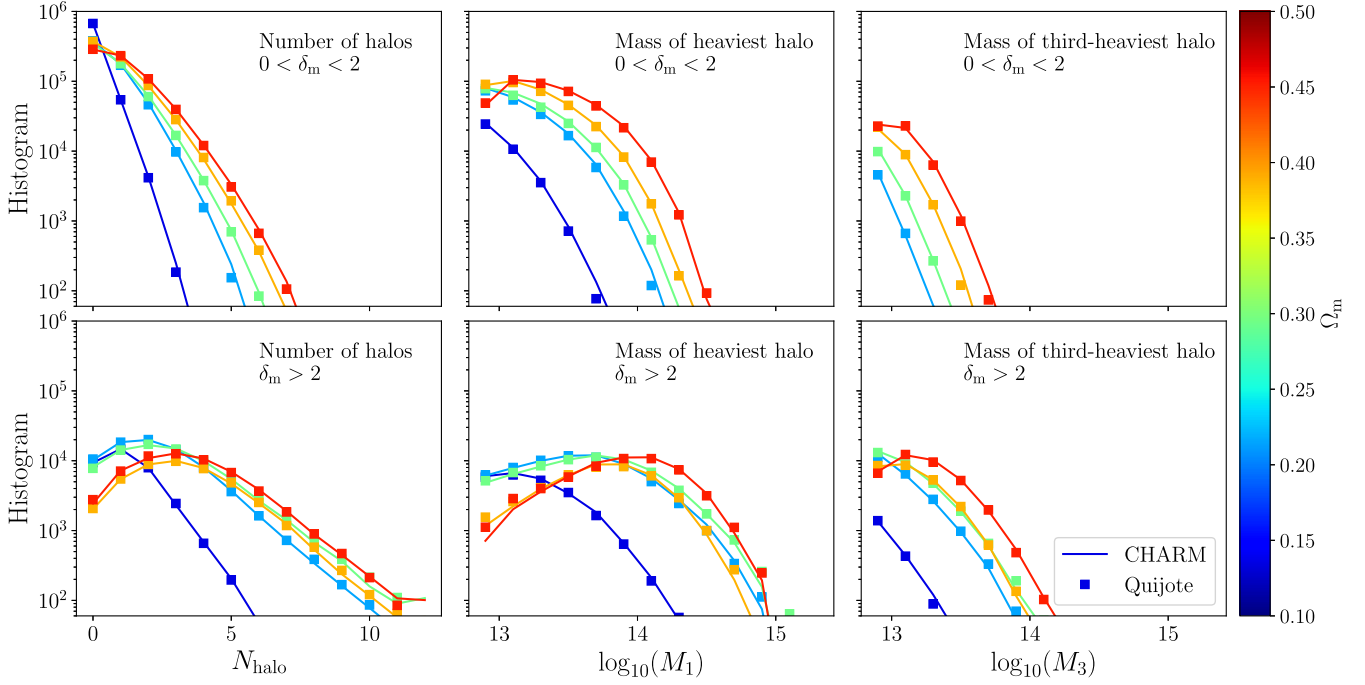


FIG. 2. Local performance of the model with halos in real space: comparison of the one-point statistics between the true (square markers) and mock (solid lines) halo catalogs in test simulations with different cosmologies. The left, middle, and right columns compare the number of halos, the mass of the heaviest halo, and the mass of the third heaviest halo, respectively, in 128^3 voxels for each cosmology. In the top row, we show the histogram for subselections of voxels with dark matter density in the range $0 < \delta_m < 2$, while in the bottom row, we use high-density voxels with $\delta_m > 2$. This test individually compares the performance of each of the three stages of the halo mass network (as described in Sec. III), demonstrating the high fidelity of the mock catalogs obtained using CHARM.

$$\mathcal{S}(\phi) = \sum_{j \in \text{Train}} \log q_{\phi}(\vec{\theta}_j | \vec{x}_{\text{data},j}), \quad (5)$$

where the summation runs over all the training examples. We use the public SBI package to perform this inference [64,65].

To maximize the training score, we first train 256 networks with random values of hyperparameters ϕ of the network. The hyperparameters include the number of fully connected layers, hidden features and batch size to obtain parameters of affine transformations, as well as parameters of the embedding network compressing the input data vector. We then select the five configurations with the best validation loss and generate an ensemble flow with these five trained networks by linearly combining them with uniform weights. We defer further optimization of this procedure, such as using a more optimal weighting scheme to combine the flows in an ensemble and using more expressive affine transformations such as neural spline flows, to a future study.

V. RESULTS

The model is now trained using 1800 LH simulations and one fiducial-cosmology simulation from the Quijote suite (see Sec. II). We train the generative model for halo masses and numbers on four Nvidia-H100 GPUs in

12 hours while training the model for halo velocities uses four Nvidia-H100 GPUs for 8 hours. However, once the network is trained, obtaining the samples of halos and their velocities on test simulation takes less than a minute on a single Nvidia-H100 GPU. We reserve the remaining 200 LH simulations as a test suite to compare the performance of the trained model.

We test the performance of the model across three axes—accuracy, inference, and scaling. We use a set of different statistical summaries described in Sec. IVA to test the accuracy of the model, then use these summaries for galaxies in redshift space to perform inference on cosmological parameters and, finally, we test how well the model scales to larger volume simulations.

A. Statistical summaries comparison

Firstly, to test the performance of our trained network, we calculate the one-point statistics (histogram), two-point statistics (power spectrum), three-point statistics (bispectrum) as well as higher-order statistics (wavelet scattering) from CHARM and true N -body halo catalogs in test simulations with varying cosmologies.

In Fig. 2, we compare the one-point statistics of the total number of halos (N_{halo}), the mass of the heaviest halo (M_1), and the mass of the third-heaviest halo (M_3). We compare these histograms in both low-density ($0 < \delta_m < 2$) and

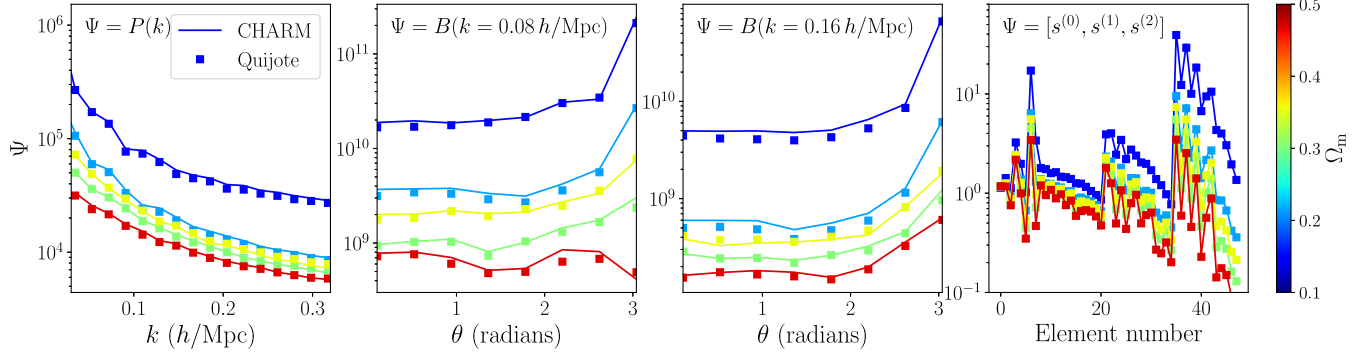


FIG. 3. Global performance of the model with halos in real space: comparison of the mass-weighted power spectrum (left), bispectrum on large scales (middle left), bispectrum on small scales (middle right), and wavelet scattering transform (right) statistics between the true (square markers) and mock (solid lines) halo catalogs in real comoving space across five different test simulations, colored by the value of the Ω_m cosmological parameter. When calculating the density field to compute these statistics, each halo is assigned a mass-dependent weight of $w = (M/M')^\alpha$, where we fix $M' = 10^{14}M_\odot/h$ and $\alpha = 0.7$ to mimic the effect of galaxies.

high-density ($\delta_m > 2$) environments in five test simulations with different cosmologies to highlight the large differences in these histograms and their dependence on the underlying dark matter density field. We observe a good match for all three histograms in both environmental conditions and all cosmologies. Note that this one-point comparison individually tests the performance of each of the three parts of the pipeline to infer the halo numbers and masses as mentioned in Sec. III. We see that, for all the cases where the number of halos is significant, our network is accurate at the percent level.

To calculate beyond one-point statistics, we take the mock and truth halo catalogs in the 1000 Mpc/ h box and calculate the 3D weighted-density fields by applying a halo mass-dependent weight $w = (M/M')^\alpha$, where $M' = 10^{14}M_\odot/h$ and $\alpha = 0.7$. This power-law scaling mimics the weighting applied by the halo occupation distribution of SDSS-CMASS galaxies at $z = 0.5$ [19,60] and makes the statistics sensitive to environment-dependent halo count and halo masses. In Fig. 3, we compare the mock and true mean power spectrum, bispectrum, and wavelet coefficients from the test simulations in the real comoving space. We show the comparison for five of the 200 LH test simulations, which highlights the large variation in the statistics due to changing cosmologies. We find good agreement in all cases, where the performance of CHARM is either better than 5% or within the cosmic variance uncertainty due to varying initial conditions.

We now test the performance of the model when including the inference of the halo velocities as well. For this, we compare the power spectra multipoles, bispectra monopoles, and wavelet scattering monopoles in redshift space. We show the comparison of truth and mock catalog statistics using markers and solid lines, respectively, in Fig. 4 for five cosmologies in the test simulation suite. We again find similar performance as for comparison in the real space. Note that, as the weighted

power spectra, bispectra, and wavelet scattering transform probe the correlations between voxels—as well as between halo counts, masses, and velocities—it is a stringent test of the joint performance of all four stages of the network. In Sec. VI, we show the performance of the model using residuals of the summary statistics between the mock and true halo catalogs for all the 200 test simulations.

Now we paste in galaxies on the halo catalogs from both the mock CHARM catalogs and true N -body simulations using a HOD formulation and measure the summary statistics in redshift space. Note that we additionally apply a Gaussian smoothing to the galaxy positions from both the mock and true catalogs (with a standard deviation of $\sigma_G \sim 8$ Mpc/ h) to account for the inability of our model to resolve halo positions below 8 Mpc/ h (see Sec. IV B for details). We compare different summary statistics for five test simulations (with varying cosmologies and random HOD parameters) in Fig. 5. We see that similar to the results for halo catalogs, we find that the model can infer the statistics of galaxies correctly up to the scales of $k < 0.32$ h/Mpc.

B. Cosmological inference

We now use the summary statistics of 18000 mock galaxy catalogs obtained from CHARM (1800 LH cosmologies and ten realizations with different values of the HOD parameters for each LH cosmology) which forms our training suite. Note that these are the same simulations that were used to train the CHARM network and we concatenate all the summary statistics described in Sec. III ($P_{\ell=0}, P_{\ell=2}, P_{\ell=4}, B_{\ell=0}$ for $k \in [0.08, 0.16, 0.32]$, and $s_{\ell=0}^{(0)}, s_{\ell=0}^{(1)}, s_{\ell=0}^{(2)}$) to form the input to the inference network. Moreover, as described in Sec. IV C, we train a set of 256 NPE and embedding networks with random hyperparameters to predict the five cosmological and five HOD parameters given the summary statistics and choose the five best networks with the lowest validation error. We form an

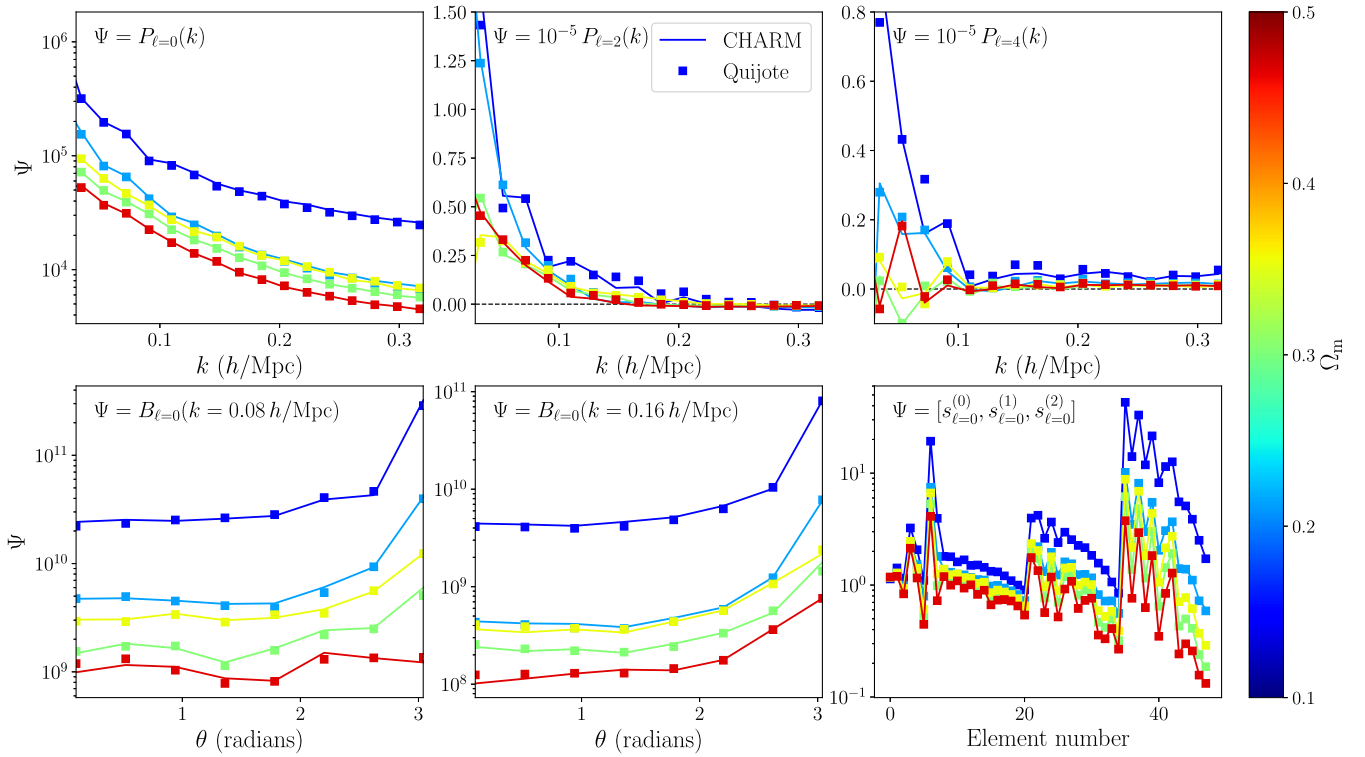


FIG. 4. Global performance of the model with halos in redshift space: comparison of the power spectrum multipoles (top row), bispectrum monopole on large scales (bottom row, left column), bispectrum monopole on small scales (bottom row, center column), and wavelet scattering transform (bottom row, right column) statistics between the true (square markers) and mock (solid lines) halo catalogs in the test LH simulations, colored by the value of the Ω_m cosmological parameter.

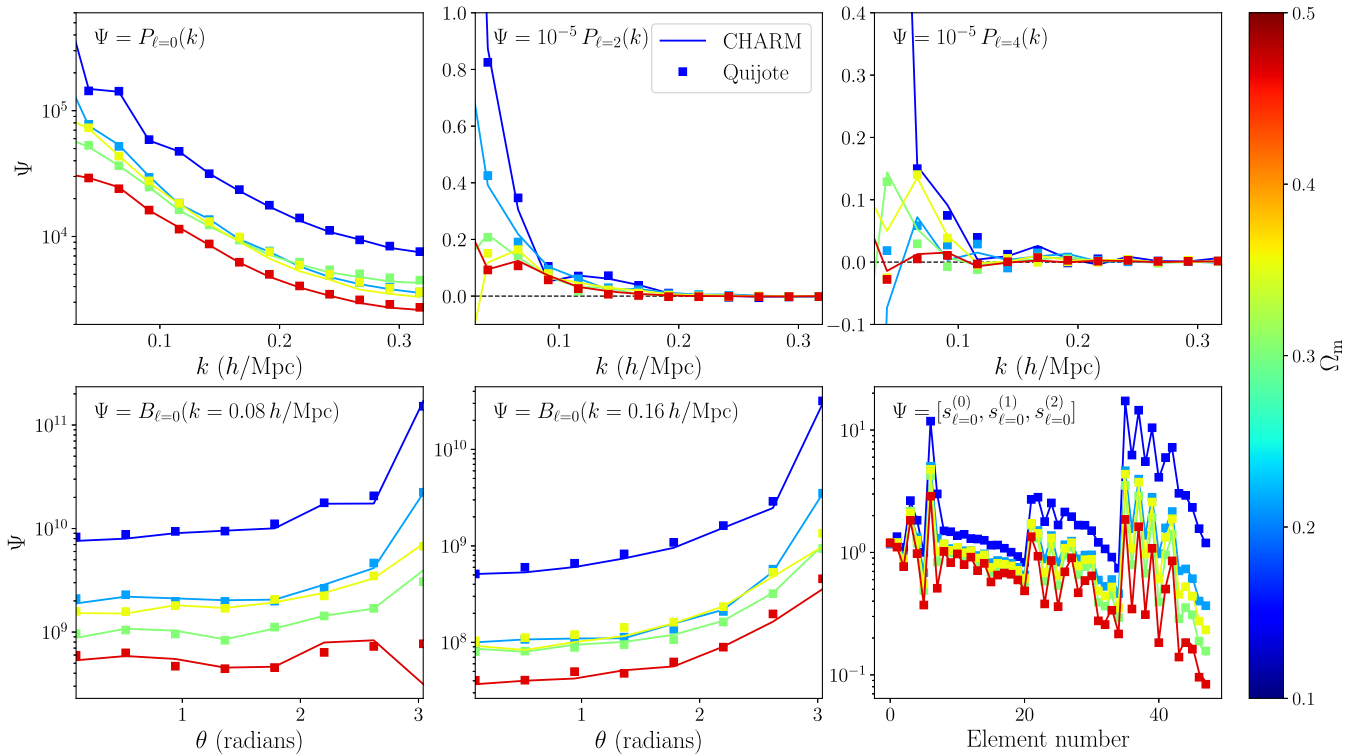


FIG. 5. Global performance of the model with galaxies in redshift space: the same as Fig. 4 but for galaxies in redshift space and each curve for different cosmology uses a random set of HOD parameters.

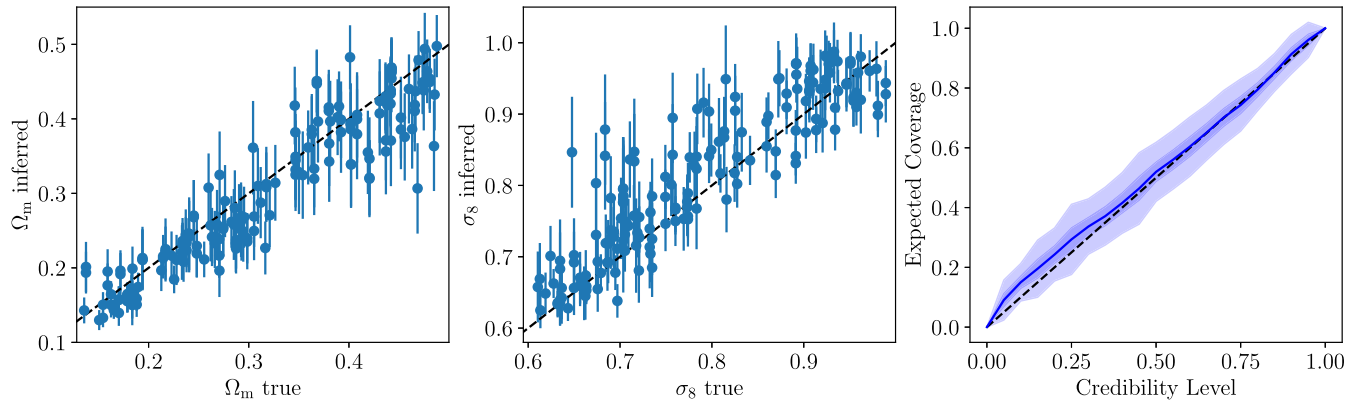


FIG. 6. Parameter inference with galaxies in redshift space: comparison of the true values of cosmological parameters in test simulations with the predicted values and their 1σ uncertainties using simulation-based inference with the redshift-space power spectra of galaxies ($P_{\ell=0}, P_{\ell=2}, P_{\ell=4}$) for $k < 0.32 h/\text{Mpc}$, monopole equilateral bispectra (with $\theta_k \in [0.1, 3.04]$ rad) at $k \in \{0.06, 0.12, 0.32\} h/\text{Mpc}$, and monopole first-, second-, and third-order wavelets: $s_{\ell=0}^{(0)}, s_{\ell=0}^{(1)}, s_{\ell=0}^{(2)}$, as described in Sec. IV A. In the right panel, we also show the coverage of the posterior and its 1σ and 3σ uncertainty bands, using the methodology presented in [66], with the diagonal line representing a well-calibrated posterior.

ensemble of these five networks to jointly infer the parameter constraints based on the summary statistics measured directly from 200 test Quijote simulations (corresponding to the remaining 200 LH simulations and using a random HOD for each simulation).

In Fig. 6, we show the scatter between the true and predicted values of two of the cosmological parameters we expect to constrain with galaxy clustering statistics: Ω_m and σ_8 . We show the mean and 1σ uncertainty on the parameters using the samples obtained from the learned posterior. We find that the network can accurately learn and constrain Ω_m and σ_8 from the mock galaxy catalogs. Note that this inference marginalizes over other cosmological and HOD parameters. We also quantify the calibration of the uncertainty of our model in the parameter space by calculating the credibility level of the inferred parameters and comparing it against expected coverage. We use the method described in [66], which uses random points in the parameter space to estimate the expected coverage probabilities and also outputs the error in the estimated values using the bootstrap method. We show the coverage plot in the right panel of Fig. 6, where we find that our inferred constraints agree with the diagonal line which corresponds to correctly calibrated posteriors.

C. Applications to larger volume simulations

The main advantage of a forward model like CHARM is that, once trained, it can be applied to significantly larger volume simulations without the need for retraining since halo formation is only affected by the dark matter environment at scales less than $100 \text{ Mpc}/h$ scales which we capture accurately here. To test this, we run large-volume N -body simulations with a box with $3 \text{ Gpc}/h$ side length. Due to computational constraints, we ran it at a

lower resolution (with 1536^3 particles) than the runs used to train CHARM. The N -body simulation takes approximately 8000 CPU hours and can only resolve halos with masses above $\log_{10}(M/[M_{\odot}/h]) > 13.7$. We run a paired FastPM simulation with the same initial condition in approximately 150 CPU hours. Afterward, we infer the halo catalogs from the trained CHARM network as described above which takes less than 2 GPU-minutes on one Nvidia-H100 GPU. We compare the statistics of the inferred and true halo catalogs in Fig. 7, showing the histogram of halo masses on the left, real-space power spectra in the middle, and wavelet scattering transform on the right. The power spectra and wavelet scattering transform are calculated using halos resolved by the full N -body simulation, $13.7 < \log(M/[M_{\odot}/h]) < 14.9$. We see that the mock catalogs from CHARM match the statistics of halos in these larger volume simulations to within 5% without any extra training. Moreover, the mock catalogs extend down to lower halo masses than the ones resolved in N -body simulation which we validate by comparing with the halo mass fitting function described in [44] which uses several high-resolution simulations that can resolve low mass halos. Note that running a $3 \text{ Gpc}/h$ box-size N -body simulation that can resolve halos with masses of $\log_{10}(M/[M_{\odot}/h]) > 12.7$ would have taken approximately 80000 CPU hours.

VI. DISCUSSION

In this work, we describe how accelerating cosmological simulations and identifying collapsed dark matter structures (halos) are crucial for maximizing the information gained from current and upcoming galaxy surveys. We develop a multistage generative model that learns the relationship

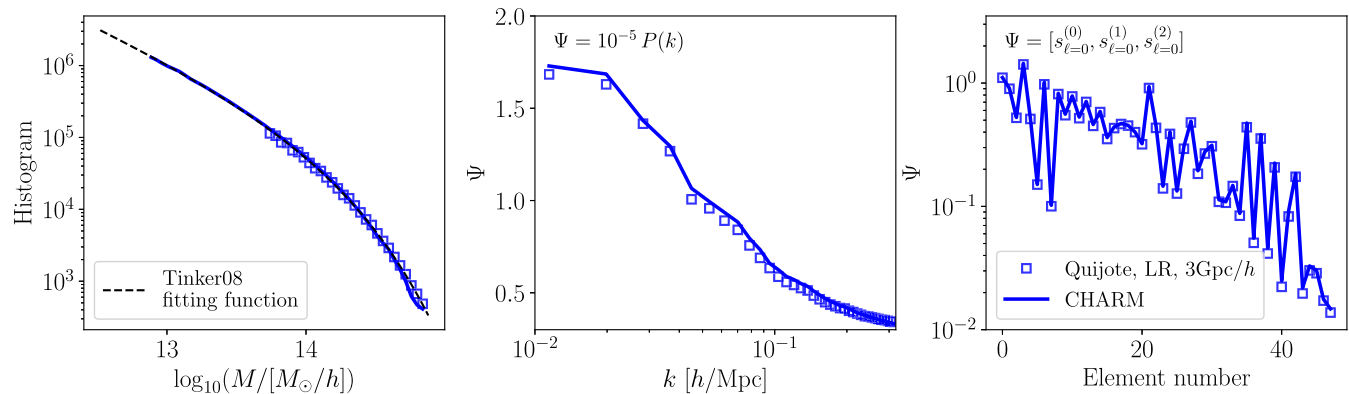


FIG. 7. Comparison of statistics of mock catalogs from CHARM with a low-resolution 3 Gpc/h box-size N -body simulation. In the left panel we compare the histogram of halo masses from the N -body simulation (unfilled square markers) which has a halo mass resolution limit of $\log(M/[M_{\odot}/h]) > 13.7$ and the one predicted from trained CHARM network (solid line) which has a mass limit of $\log(M/[M_{\odot}/h]) > 12.7$. We also compare them to the fitting function presented in [44] finding that CHARM can extrapolate to lower halo masses well. In the middle and right panels, we compare the power spectra and wavelet scattering transforms of halos in real space in the mass range resolved by the N -body simulation ($13.7 < \log(M/[M_{\odot}/h]) < 14.9$).

between the halo distribution and the surrounding dark matter distribution on large scales. Our approach uses a faster alternative to N -body simulations, which accurately captures dark matter densities on these scales. We validate our mock catalogs, obtained using CHARM, by comparing various statistics under different environmental conditions and cosmologies, demonstrating the method’s accuracy.

We then paint galaxies onto the mock halo catalogs generated by the trained CHARM network and use the velocity information to calculate observable galaxy clustering statistics in redshift space. Comparisons with results from full N -body simulations show that, for scales of interest ($k < 0.32 h/\text{Mpc}$), CHARM produces accurate galaxy catalogs. Subsequently, we train a neural density network to infer cosmological and HOD parameters from observed galaxy clustering statistics and find that the calibrated posterior constrains cosmology accurately. Finally, we illustrate how CHARM, trained on high-resolution, small-volume N -body simulations, can generate highly accurate halo catalogs in large-volume simulations at a fraction of the computational cost of running the full N -body simulation.

In this study, we trained the network with a physical resolution of 8 Mpc/h and used a simple model of the galaxy-halo connection. Future work will focus on training the network on higher-resolution simulations to utilize even more information from galaxy catalogs. Additionally, incorporating more realistic halo-galaxy connection models will require information about secondary halo properties beyond mass, such as concentration and merger history [67]. We also trained the network at a fixed redshift of $z = 0.5$, relevant for SDSS-CMASS-like galaxies. As we probe fainter galaxies with upcoming surveys, we will need to develop models that generalize to different redshifts. Moreover, in this study we have trained the model to recover

the Rockstar halo catalogs with spherical overdensity mass definition, but the simulation-based inference of cosmological parameters can be sensitive to this choice [68]. However, a similar network can be trained to generate catalogs corresponding to other halo finders, such as friends of friends. Finally, as the trained generative model is differentiable, it can be optimized to find the maximum *a posteriori* estimate of the mock halo catalog conditioned on the surrounding dark matter properties, which will have a high cross-correlation coefficient with the true halo catalogs from the full N -body simulation. We plan to explore these developments in future studies.

ACKNOWLEDGMENTS

We thank Francois Lanusse for valuable discussions. S. P. thanks the CCA and CCM at the Flatiron Institute for hospitality while (a portion of) this research was carried out. The computations reported in this paper were performed using resources made available by the Flatiron Institute. The Flatiron Institute is supported by the Simons Foundation. This work is supported by the Simons Collaboration on “Learning the Universe.”

DATA AVAILABILITY

The data that support the findings of this article are openly available [22].

APPENDIX A: HALO STATISTICS RESIDUAL PERFORMANCE

In Fig. 8, we compare the performance of the model by plotting the residual of the predicted and true summary statistics of halos in redshift space. We show the residuals for all 200 test LH Quijote simulations. We color each

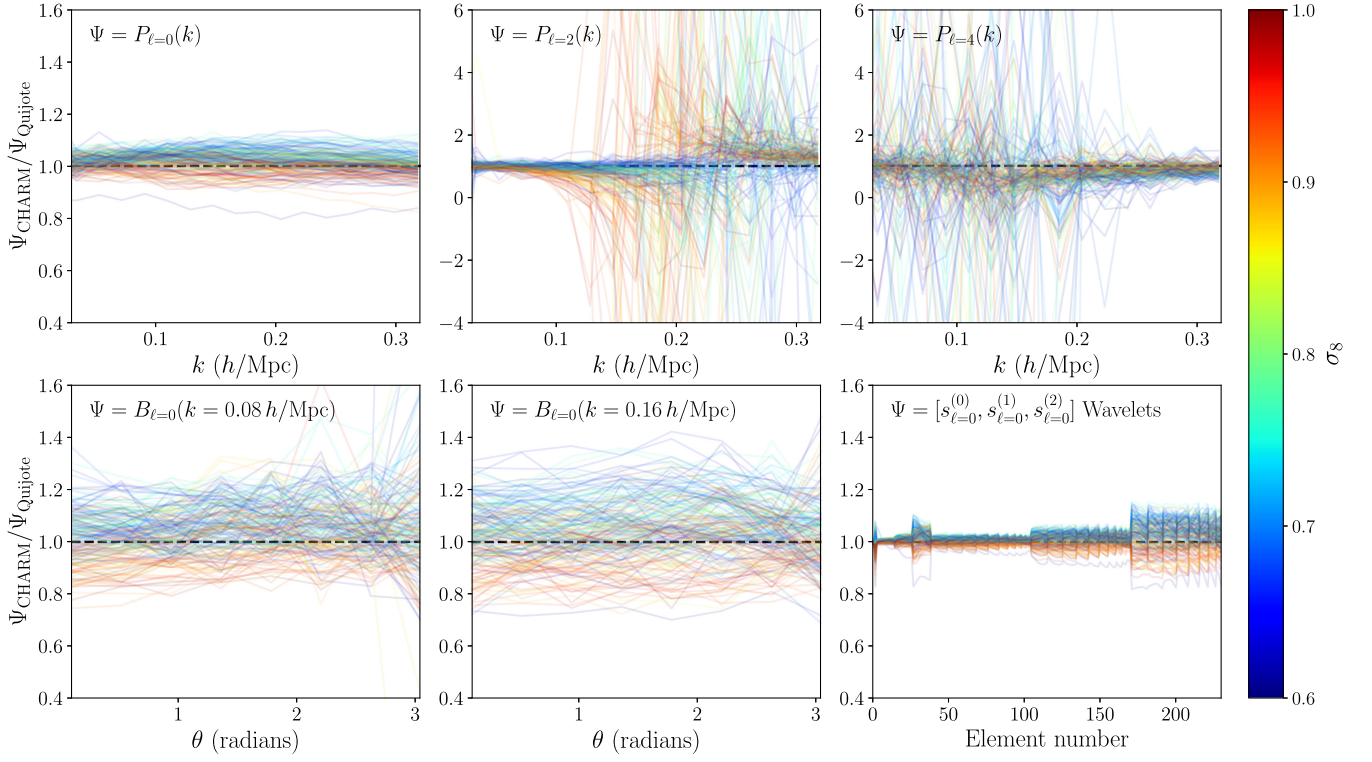


FIG. 8. Residuals of the global performance of the model with halos in redshift space: the same as Fig. 4 but showing the residuals of each of the summary statistics. Here we show the residuals for all the 200 test simulations and color the lines with the cosmological parameter σ_8 .

residual by the value of σ_8 parameter in that LH simulation. We find that the trained model works well either at approximately 5% level or within the expected cosmic variance for a majority of the test simulations. We find a slight degradation of the performance for simulations having extreme values of the cosmological parameters but this does not bias our cosmological inference (see Sec. V). It is important to note that the quadrupole [$P_{\ell=2}(k)$] power spectra have a zero crossing at $k \sim 0.2 h/\text{Mpc}$ and

hexadecapole [$P_{\ell=4}(k)$] also crosses zero due to large noise, which makes the residual statistics to have large values.

APPENDIX B: SIMULATION RUN TIMES AND ACRONYMS

In Table I we compare the run-time of various simulations presented in the main text. A description of the acronyms used in the main text are detailed in Table II.

TABLE I. Simulation run-time comparison.

Simulation name	Run-time
Quijote-hres (1 Gpc/h box, 1024^3 particles)	5000 CPU hours
PM-lres (1 Gpc/h box, 384^3 particles)	5 CPU hours
Quijote-lres (1 Gpc/h box, 1536^3 particles)	80 000 CPU hours
PM-lres (1 Gpc/h box, 1152^3 particles)	150 CPU hours

TABLE II. Acronym definitions.

Acronym	Full form
CHARM	Creating halos with autoregressive multistage networks
SBI	Simulation-based inference
PM	Particle mesh simulations
LH	Latin hypercube set of simulations
NGP	Nearest grid point
ResNet	Residual networks
FCNN	Fully connected neural network
KL	Kullback-Leibler
SDSS	Sloan Digital Sky Survey
CMASS	Constant stellar mass at constant redshift
HOD	Halo occupation distribution
NPE	Neural posterior estimation

- [1] S. Alam *et al.*, *Mon. Not. R. Astron. Soc.* **470**, 2617 (2017).
- [2] A. Albrecht *et al.*, [arXiv:astro-ph/0609591](https://arxiv.org/abs/astro-ph/0609591).
- [3] F. Bernardeau, S. Colombi, E. Gaztanaga, and R. Scoccimarro, *Phys. Rep.* **367**, 1 (2002).
- [4] Planck Collaboration *et al.*, *Astron. Astrophys.* **641**, A6 (2020).
- [5] S.-F. Chen, Z. Vlah, and M. White, *J. Cosmol. Astropart. Phys.* **03** (2021) 100.
- [6] G. D'Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, *J. Cosmol. Astropart. Phys.* **05** (2024) 059.
- [7] T. Nishimichi, G. D'Amico, M. M. Ivanov, L. Senatore, M. Simonović, M. Takada, M. Zaldarriaga, and P. Zhang, *Phys. Rev. D* **102**, 123541 (2020).
- [8] O. H. E. Philcox and M. M. Ivanov, *Phys. Rev. D* **105**, 043517 (2022).
- [9] A. Banerjee and T. Abel, **500**, 5479 (2021).
- [10] A. E. Bayer *et al.*, *Astrophys. J.* **919**, 24 (2021).
- [11] M. Eickenberg *et al.*, *Wavelet Moments for Cosmological Parameter Estimation* (2022).
- [12] C. Hahn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218810120 (2023).
- [13] T. L. Makinen, T. Charnock, P. Lemos, N. Porqueres, A. F. Heavens, and B. D. Wandelt, *J. Cosmol. Astropart. Phys.* **11** (2021) 049.
- [14] E. Massara, F. Villaescusa-Navarro, S. Ho, N. Dalal, and D. N. Spergel, *Phys. Rev. Lett.* **126**, 011301 (2021).
- [15] K. Naidoo, E. Massara, and O. Lahav, *Mon. Not. R. Astron. Soc.* **513**, 3596 (2022).
- [16] G. Valogiannis and C. Dvorkin, *Phys. Rev. D* **105**, 103534 (2022).
- [17] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, *Mon. Not. R. Astron. Soc.* **488**, 4440 (2019).
- [18] Beyond-2pt Collaboration *et al.*, *Astrophys. J.* **990**, 99 (2025).
- [19] C. Hahn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218810120 (2023).
- [20] M. Ho *et al.*, *Open J. Astrophys.* **7**, 54 (2024).
- [21] K. Cranmer, J. Brehmer, and G. Louppe, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055 (2020).
- [22] F. Villaescusa-Navarro *et al.*, *Astrophys. J. Suppl. Ser.* **250**, 2 (2020).
- [23] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, *Astrophys. J.* **762**, 109 (2013).
- [24] V. Desjacques, D. Jeong, and F. Schmidt, *Phys. Rep.* **733**, 1 (2018).
- [25] D. J. Bartlett, M. Ho, and B. D. Wandelt, *Astrophys. J. Lett.* **977**, L44 (2024).
- [26] Y. Feng, M.-Y. Chu, U. Seljak, and P. McDonald, *Mon. Not. R. Astron. Soc.* **463**, 2273 (2016).
- [27] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein, *J. Cosmol. Astropart. Phys.* **06** (2013) 036.
- [28] L. Doerer, D. Jamieson, S. Stopyra, G. Lavaux, F. Leclercq, and J. Jasche, *Mon. Not. R. Astron. Soc.* **535**, 1258 (2024).
- [29] Y. Wu, H. Guo, and V. Springel, **531**, 4944 (2024).
- [30] <https://learning-the-universe.org>.
- [31] T. Charnock, G. Lavaux, B. D. Wandelt, S. Sarma Boruah, J. Jasche, and M. J. Hudson, *Mon. Not. R. Astron. Soc.* **494**, 50 (2020).
- [32] S. Ding, G. Lavaux, and J. Jasche, *Astron. Astrophys.* **690**, A236 (2024).
- [33] C. Modi, Y. Feng, and U. Seljak, *J. Cosmol. Astropart. Phys.* **10** (2018) 028.
- [34] X. Zhang, Y. Wang, W. Zhang, Y. Sun, S. He, G. Contardo, F. Villaescusa-Navarro, and S. Ho, [arXiv:1902.05965](https://arxiv.org/abs/1902.05965).
- [35] D. Jamieson, Y. Li, R. Alves de Oliveira, F. Villaescusa-Navarro, S. Ho, and D. N. Spergel, *Field Level Neural Network Emulator for Cosmological N-body Simulations* (2022).
- [36] D. Jamieson, Y. Li, F. Villaescusa-Navarro, S. Ho, and D. N. Spergel, *J. Cosmol. Astropart. Phys.* **03** (2025) 072.
- [37] D. J. Bartlett, M. Chiarenza, L. Doerer, and F. Leclercq, *Astron. Astrophys.* **694**, A287 (2025).
- [38] Y. Li, C. Modi, D. Jamieson, Y. Zhang, L. Lu, Y. Feng, F. Lanusse, and L. Greengard, *Astrophys. J. Suppl. Ser.* **270**, 36 (2024).
- [39] C. Modi, F. Lanusse, and U. Seljak, *Astron. Comput.* **37**, 100505 (2021).
- [40] A. E. Bayer, C. Modi, and S. Ferraro, *J. Cosmol. Astropart. Phys.* **06** (2023) 046.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, [arXiv:1512.03385](https://arxiv.org/abs/1512.03385).
- [42] S. Kullback and R. A. Leibler, *Ann. Math. Stat.* **22**, 79 (1951).
- [43] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, [arXiv:1906.04032](https://arxiv.org/abs/1906.04032).
- [44] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz, *Astrophys. J.* **688**, 709 (2008).
- [45] E. Gumbel, *Ann. Inst. Henri Poincaré* **5**, 115 (1935).
- [46] J. C. Jackson, *Mon. Not. R. Astron. Soc.* **156**, 1P (1972).
- [47] <https://pylians3.readthedocs.io/en/master/>.
- [48] B. Régalo-Saint Blancard *et al.*, *Phys. Rev. D* **109**, 083535 (2024).
- [49] https://github.com/bregaldo/galactic_wavelets/tree/main.
- [50] P. S. Behroozi, C. Conroy, and R. H. Wechsler, *Astrophys. J.* **717**, 379 (2010).
- [51] C. Conroy, R. H. Wechsler, and A. V. Kravtsov, *Astrophys. J.* **647**, 201 (2006).
- [52] B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser, *Astrophys. J.* **710**, 903 (2010).
- [53] B. P. Moster, T. Naab, and S. D. M. White, **477**, 1822 (2018).
- [54] X. Yang, H. J. Mo, and F. C. van den Bosch, *Astrophys. J.* **695**, 900 (2009).
- [55] Z. Zheng, A. L. Coil, and I. Zehavi, *Astrophys. J.* **667**, 760 (2007).
- [56] R. S. Somerville and R. Davé, **53**, 51 (2015).
- [57] F. Villaescusa-Navarro *et al.*, *Astrophys. J.* **929**, 132 (2022).
- [58] M. Vogelsberger *et al.*, *Mon. Not. R. Astron. Soc.* **444**, 1518 (2014).
- [59] R. S. Somerville and J. R. Primack, **310**, 1087 (1999).
- [60] B. A. Reid, H.-J. Seo, A. Leauthaud, J. L. Tinker, and M. White, *Mon. Not. R. Astron. Soc.* **444**, 476 (2014).
- [61] C. Hahn *et al.*, *Phys. Rev. D* **109**, 083534 (2024).

- [62] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **490**, 493 (1997).
- [63] G. Papamakarios, T. Pavlakou, and I. Murray, [arXiv:1705.07057](https://arxiv.org/abs/1705.07057).
- [64] A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke, *J. Open Source Software* **5**, 2505 (2020).
- [65] <https://github.com/sbi-dev/sbi>.
- [66] P. Lemos, A. Coogan, Y. Hezaveh, and L. Perreault-Levasseur, *Proceedings of the 40th International Conference on Machine Learning* (2023), Vol. 202, p. 19256.
- [67] C. K. Jespersen, M. Cranmer, P. Melchior, S. Ho, R. S. Somerville, and A. Gabrielpillai, *Astrophys. J.* **941**, 7 (2022).
- [68] C. Modi, S. Pandey, M. Ho, C. Hahn, B. R'egaldo-Saint Blancard, and B. Wandelt, *Mon. Not. R. Astron. Soc.* **536**, 254 (2024).