

Galaxy clustering analysis with SimBIG and the wavelet scattering transform

Bruno Régaldo-Saint Blancard^{1,*} ChangHoon Hahn,² Shirley Ho,³ Jiamin Hou,^{4,5} Pablo Lemos,^{6,7} Elena Massara,^{8,9}
Chirag Modi,^{1,3} Azadeh Moradinezhad Dizgah,¹⁰ Liam Parker,¹¹ Yuling Yao,¹ and Michael Eickenberg¹

(SimBIG Collaboration)

¹*Center for Computational Mathematics, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA*

²*Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA*

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

⁴*Department of Astronomy, University of Florida, 211 Bryant Space Science Center,
Gainesville, Florida 32611, USA*

⁵*Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstrasse 1,
85748 Garching bei München, Germany*

⁶*Department of Physics, Université de Montréal,
Montréal, 1375 Avenue Thérèse-Lavoie-Roux, Québec H2V 0B3, Canada*

⁷*Mila—Quebec Artificial Intelligence Institute, Montréal,
6666 Rue Saint-Urbain, Québec H2S 3H1, Canada*

⁸*Waterloo Centre for Astrophysics, University of Waterloo,
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada*

⁹*Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West,
Waterloo, Ontario N2L 3G1, Canada*

¹⁰*Département de Physique Théorique, Université de Genève,
24 quai Ernest Ansermet, 1211 Genève 4, Switzerland*

¹¹*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*



(Received 12 July 2023; accepted 19 February 2024; published 30 April 2024)

The non-Gaussian spatial distribution of galaxies traces the large-scale structure of the Universe and therefore constitutes a prime observable to constrain cosmological parameters. We conduct Bayesian inference of the Λ CDM parameters Ω_m , Ω_b , h , n_s , and σ_8 from the Baryon Oscillation Spectroscopic Survey CMASS galaxy sample by combining the wavelet scattering transform (WST) with a simulation-based inference approach enabled by the SimBIG forward model. We design a set of reduced WST statistics that leverage symmetries of redshift-space data. Posterior distributions are estimated with a conditional normalizing flow trained on 20,000 simulated SimBIG galaxy catalogs with survey realism. We assess the accuracy of the posterior estimates using simulation-based calibration and quantify generalization and robustness to the change of forward model using a suite of 2000 test simulations. When probing scales down to $k_{\max} = 0.5 h/\text{Mpc}$, we are able to derive accurate posterior estimates that are robust to the change of forward model for all parameters, except σ_8 . We mitigate the robustness issues with σ_8 by removing the WST coefficients that probe scales smaller than $k \sim 0.3 h/\text{Mpc}$. Applied to the Baryon Oscillation Spectroscopic Survey CMASS sample, our WST analysis yields seemingly improved constraints obtained from a standard perturbation-theory-based power spectrum analysis with $k_{\max} = 0.25 h/\text{Mpc}$ for all parameters except h . However, we still raise concerns on these results. The observational predictions significantly vary across different normalizing flow architectures, which we interpret as a form of model misspecification. This highlights a key challenge for forward modeling approaches when using summary statistics that are sensitive to detailed model-specific or observational imprints on galaxy clustering.

DOI: [10.1103/PhysRevD.109.083535](https://doi.org/10.1103/PhysRevD.109.083535)

*bregaldo@flatironinstitute.org

I. INTRODUCTION

The evolution of the Universe is remarkably well understood under the lens of the standard model of cosmology, or Λ CDM model. The spatial distribution of galaxies traces the large-scale structure of the Universe (LSS) and constitutes a prime observable to constrain the parameters of this model. In the past decades, spectroscopic galaxy surveys such as the Sloan Digital Sky Survey III: Baryon Oscillation Spectroscopic Survey (BOSS) [1,2] provided us with redshift measurements for over a million galaxies spanning an effective volume of several Gpc^3 . Upcoming instruments, including the Dark Energy Spectroscopy Instrument [3], the Subaru Prime Focus Spectrograph [4], the *Euclid* satellite [5], and the Nancy Grace Roman Space Telescope [6] will surpass these numbers by more than an order of magnitude. This tremendous amount of data will provide a unique opportunity to precisely measure the expansion and growth histories of the Universe.

Cosmological inference from the observed galaxy distribution is commonly formulated within a Bayesian framework. Given a prior distribution on the target parameters, a set of *summary statistics* describing the galaxy clustering, and a model for the corresponding likelihood function, Bayesian inference yields a posterior distribution on the parameters. Typically, the likelihood is assumed to follow a Gaussian functional form, although this assumption is hardly justifiable [7,8].¹ In terms of summary statistics, galaxy clustering analyses thus far have focused primarily on the galaxy power spectrum, $P_\ell(k)$ [e.g. [9–16]], using perturbation theory (PT) [17,18] to model its dependency on the cosmological parameters. These standard analyses are limited to (large) scales in the linear and weakly nonlinear regime, where PT is valid. As a result, they do not exploit cosmological information available in the strongly nonlinear regime or in the non-Gaussian properties of the galaxy distribution.

The literature has investigated several summary statistics beyond P_ℓ , including the bispectrum [19–21], marked power spectrum [22], and skew spectra [23]. These works firmly demonstrate that non-Gaussian information significantly improves the precision on cosmological constraints. Unfortunately, traditional PT-based analyses hardly extend to these statistics, which makes analytic connections to the cosmological parameters unclear if not intractable, especially in the nonlinear regime. On the other hand, simulation-based inference² (SBI), which has grown in popularity in astronomy and cosmology, offers an alternative approach to model the likelihood function by leveraging high-fidelity numerical simulations, and thus avoiding strong assumptions on the likelihood. Typically, neural density estimators are trained on these simulations to enable efficient posterior sampling from the target observation [24,25].

¹The functional form of the likelihood *a priori* depends directly on the summary statistics.

²SBI is also often called *likelihood-free inference* or *implicit likelihood inference* in the literature.

In this context, [26,27] introduced the SimBIG forward modeling framework, which enables SBI for arbitrary summary statistics. SimBIG accurately models galaxy clustering down to nonlinear scales and includes systematic effects in the observations. To demonstrate the framework, the authors adapted SimBIG to the BOSS CMASS [2] galaxy sample, and conducted SBI of Ω_m , Ω_b , h , n_s , and σ_8 from the galaxy power spectrum. Following in the footsteps of this study, we lead a similar analysis using the wavelet scattering transform (WST).

The WST, first introduced by [28,29] from a data science perspective, defines a set of descriptive statistics well suited to the characterization of non-Gaussian fields resulting from nonlinear multiscale physics. In the past years, it has notably found interest in astrophysics for the modeling of the interstellar medium [30–32], and in cosmology for the analysis of weak lensing data [33,34], line-intensity mapping [35], or the study of the epoch of reionization [36,37]. A 3D version of the WST first introduced in [38] for quantum chemistry applications enabled the extension of these statistics to the characterization of LSS [39–41]. Fisher forecasts have shown that these statistics are much more constraining than the power spectrum on simulated matter density fields [39,40,42]. Recently, [41] showed that this is still the case on BOSS data in a Bayesian inference setting with strong assumptions on the likelihood function (including a Gaussian assumption). We now go further and conduct an SBI analysis of BOSS using the WST. We introduce a new variant of the WST statistics designed to account for the specificity of galaxy surveys, as well as further methodological developments to the SimBIG inference pipeline for posterior estimation and validation.

The rest of this paper is organized as follows. In Sec. II, we present our methodology. We specify our target observational data, and then introduce our inference pipeline and the WST statistics used as summary statistic. In Sec. III, we present our results, which comprise a validation of our pipeline and the posterior distributions of the Λ CDM parameters. Finally, in Sec. IV, we draw our conclusions and discuss perspectives. This paper also includes three Appendices.

II. METHODOLOGY

Our goal is to infer the Λ CDM cosmological parameters Ω_m , Ω_b , h , n_s , and σ_8 from the BOSS galaxy survey using SBI. In Sec. II A, we introduce the target galaxy sample. We then give an overview of our forward model in Sec. II B. In Sec. II C, we present the WST summary statistics that we have designed for this work. Finally, in Sec. II D, we describe the SBI methodology.

A. Observations: BOSS CMASS galaxies

In this paper, we analyze the CMASS galaxy sample of the BOSS Data Release 12. We limit this analysis to the

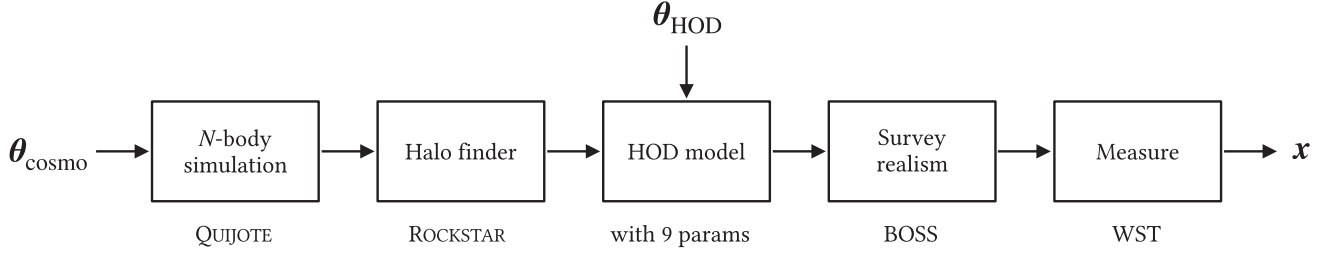


FIG. 1. Diagram representing the different components of the SimBIG forward model.

southern galactic cap (SGC), which is further restricted in angular coordinates to $\text{Dec} > -6$ deg and $-25 < \text{RA} < 28$ deg and in redshift to $0.45 < z < 0.6$. In total, this catalog comprises 109,636 galaxies. We refer to [26,27] for visuals of the sample.

B. SimBIG forward modeling framework

We frame our inference problem in a Bayesian setting. Given a vector of measurements, \mathbf{x} , obtained from our galaxy sample, we aim to infer the posterior distribution of the vector of model parameters, $\boldsymbol{\theta}$. Bayes' theorem decomposes this posterior distribution as $p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(\mathbf{x}|\boldsymbol{\theta})$ is the likelihood function and $p(\boldsymbol{\theta})$ is the prior distribution.

With SBI (for a review, see [24]), the likelihood function is learned from simulated observations. This prevents us from using strong assumptions on its functional form. SBI only requires the ability to simulate a \mathbf{x} for any given $\boldsymbol{\theta}$ through our forward model. For this analysis, we employ the SimBIG forward model. We give below a general overview of this model and its parameters, and refer to [26,27] for more details. Figure 1 gives a schematic representation of the components of the SimBIG forward model. These include the following steps:

- (1) An N -body simulation that solves the dynamical evolution of a dark matter fluid in a Λ CDM cosmology parametrized by $\boldsymbol{\theta}_{\text{cosmo}} = (\Omega_m, \Omega_b, h, n_s, \sigma_8)$. We use the existing high-resolution simulations of the Quijote suite [43].
- (2) A dark matter halo finder, which identifies halos from the dark matter particles. Here, we use Rockstar [44].
- (3) A halo occupation distribution (HOD) framework designed to populate halos with galaxies in a phenomenological way. The HOD introduces another set of nine HOD parameters $\boldsymbol{\theta}_{\text{HOD}}$.
- (4) The application of survey realism to mimic the BOSS CMASS SGC sample. This includes a reproduction of the survey geometry as well as observational systematics, including fiber collisions.
- (5) The measurements of summary statistics from the final mock galaxy survey. This paper focuses on WST statistical measurements, described in detail in Sec. II C.

The target parameters $\boldsymbol{\theta}$ of the inference are $\boldsymbol{\theta}_{\text{cosmo}}$ and $\boldsymbol{\theta}_{\text{HOD}}$. We use the same prior distribution $p(\boldsymbol{\theta})$ as in [27]. We recall in Table I the priors on the cosmological parameters $\boldsymbol{\theta}_{\text{cosmo}}$. These are conservative priors set by the range of parameters used to run the Quijote simulations. We refer to Ref. [27] (Table 1) for a specification of the priors on $\boldsymbol{\theta}_{\text{HOD}}$.

C. Summary statistics

In this work, we focus on WST statistics to capture the non-Gaussian information stemming from the nonlinear dynamical evolution of the LSS. We handcraft a new variant of WST statistics that leverage the symmetries of galaxy surveys. These include the power spectrum information as well as a quantification of interactions between scales.

1. Wavelets

Given an initial anisotropic 3D wavelet ψ called the *mother wavelet*, we build a family of wavelets $\{\psi_\lambda\}$ by dilation and rotation of ψ . With $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$, the standard basis of \mathbb{R}^3 , we introduce spherical coordinates (r, ϑ, ϕ) . r is the radius, ϑ is the polar angle defined with respect to \mathbf{e}_z , and ϕ is the azimuthal angle. We also introduce the unit radial vector $\mathbf{e}_{\vartheta,\phi} = \sin \vartheta \cos \phi \mathbf{e}_x + \sin \vartheta \sin \phi \mathbf{e}_y + \cos \vartheta \mathbf{e}_z$. Calling j the index of dilation, the wavelet ψ_λ with $\lambda = (j, \vartheta, \phi)$ is a dilated (by a factor 2^j) and rotated version of ψ , formally defined as follows:

$$\psi_{j,\vartheta,\phi}(\mathbf{r}) = 2^{-3j} \psi(T_{\vartheta,\phi} 2^{-j} \mathbf{r}), \quad (1)$$

 TABLE I. Prior distribution $p(\boldsymbol{\theta})$ and fiducial cosmologies $\boldsymbol{\theta}_{\text{fid}}$ used for the Quijote and Abacussummit simulations.

| Cosmological parameter | Prior | Quijote fiducial | Abacussummit fiducial |
|------------------------|---------------------------|------------------|-----------------------|
| Ω_m | $\mathcal{U}(0.1, 0.5)$ | 0.3175 | 0.3152 |
| Ω_b | $\mathcal{U}(0.03, 0.07)$ | 0.049 | 0.0493 |
| h | $\mathcal{U}(0.5, 0.9)$ | 0.6711 | 0.6736 |
| n_s | $\mathcal{U}(0.8, 1.2)$ | 0.9624 | 0.9649 |
| σ_8 | $\mathcal{U}(0.6, 1.0)$ | 0.834 | 0.8080 |

where $T_{\vartheta,\varphi}$ is a rotation matrix of SO(3) such that $T_{\vartheta,\varphi}\mathbf{e}_z = \mathbf{e}_{\vartheta,\varphi}$.³ Note that this only constrains $T_{\vartheta,\varphi}$ up to a rotation around $\mathbf{e}_{\vartheta,\varphi}$. However, since we later choose the mother wavelet ψ to be axisymmetric, this definition will not depend on the particular choice of $T_{\vartheta,\varphi}$.

Inspired by [39], we define the mother wavelet ψ in Fourier space as the product of a radial component R and an angular component α , that is

$$\hat{\psi}(\mathbf{k}) = R(k)\alpha(\vartheta(\mathbf{k}), \varphi(\mathbf{k})). \quad (2)$$

The radial component R is chosen to be the Fourier transform of the spline (isotropic) wavelet introduced in [45]. This wavelet benefits from a good localization in physical space while being reasonably regular. Its definition is given in Appendix A. Note that it depends on a cutoff frequency k_c that delimits its support in Fourier space, with $R(k) = 0$ for $k > 2k_c$. For the angular component α , we choose a Gaussian function that reads:

$$\alpha(\vartheta) = \exp\left(-\frac{\vartheta^2}{2\sigma_\vartheta^2}\right), \quad (3)$$

with σ_ϑ the angular width that parametrizes the localization of the function. Since α only depends on ϑ , and not on φ , the resulting mother wavelet ψ is axisymmetric. This feature will be helpful in the following to exploit the statistical invariance properties of redshift-space data.

Similarly to [39], we employ dilation indices j of the form k/Q with $k \in \{0, \dots, J \times Q - 1\}$. The parameter J is the total number of octaves (doublings in scale) taken into account for the analysis, and Q is a parameter called *quality factor* that controls the number of scales per octave. The angular indices ϑ and φ are chosen to be the 20 vertices of a regular dodecahedron.⁴ Initially orienting the regular dodecahedron such that one vertex coincides with the pole \mathbf{e}_z ($\vartheta = 0$), we only keep the 10 vertices with $\vartheta \in [0, \pi/2]$. This provides a good trade-off between computation speed and constraining power of the statistic.

We set $k_c = 4\pi k_{\max}/3$, where k_{\max} denotes the largest spatial frequency we want to probe. We also choose $\sigma_\vartheta = \pi/4$, as well as $J = 6$ and $Q = 2$. These parameters lead to a family of 120 wavelets. We visualize a subset of the wavelets in Fig. 2.

2. Wavelet scattering moments

The definition of the WST coefficients involves operations that relate to the structure of convolutional neural

³Note that the 2^{-3j} prefactor used to normalize the wavelets has no impact on the rest of this work.

⁴Or equivalently, the surface normals of a regular icosahedron. This choice corresponds to an invariant action space of the largest discrete nonplanar transformation subgroup of SO(3), the icosahedral group.

networks, namely cascades of convolutions, pointwise nonlinearities, and average pooling operators.⁵ We define a generalized version of the wavelet scattering moments associated with a given real-valued field F as follows:

$$S_0[p] = \langle |F|^p \rangle, \quad (4)$$

$$S_1[\lambda, p] = \langle |F * \psi_\lambda|^p \rangle, \quad (5)$$

$$S_2[\lambda_1, \lambda_2, p] = \langle ||F * \psi_{\lambda_1} * \psi_{\lambda_2}|^p \rangle, \quad (6)$$

where $*$ denotes the convolution operation, and $\langle \cdot \rangle$ is an averaging operator. These moments thus depend on wavelet indices λ_i and exponents p . We refer to the S_0 , S_1 , and S_2 moments as the zeroth-, first-, and second-order moments, respectively. S_0 coefficients are moments estimates of the one-point function. S_1 coefficients measure moments of x filtered at a given oriented scale λ .⁶ Finally, S_2 coefficients quantify in an analogous way interactions between oriented scales λ_1 and λ_2 . Note that S_2 coefficients with $j_2 \leq j_1$ are discarded since they usually become either noninformative or negligible [47].

A galaxy catalog takes the form of a point cloud, where the density field can be formally written as follows:

$$n(\mathbf{r}) = \sum_i w_{c,i} \delta_D(\mathbf{r} - \mathbf{r}_i), \quad (7)$$

where $w_{c,i}$ are galaxy completeness weights. The weights $w_{c,i}$ only play a role when measuring statistics on the BOSS data to compensate for imaging systematics or failures in measuring accurate redshifts (for a further discussion, see [27]). For the mock catalogs produced by our forward model, these are uniformly set to one. We also introduce the field $\bar{n}(\mathbf{r})$ corresponding to the expected mean space density of galaxies, see [48]. However, as it is usually done in the galaxy clustering analysis literature, and in order to mitigate edge effects due to the BOSS survey geometry in the computation of the WST coefficients, we focus on the density fluctuation field F defined as

$$F(\mathbf{r}) = \frac{1}{\sqrt{I}} [n(\mathbf{r}) - \alpha n_r(\mathbf{r})], \quad (8)$$

where $I = \int \bar{n}^2(\mathbf{r}) d\mathbf{r}$ is a normalization factor,⁷ n_r is a random catalog of more than 4×10^6 objects with same

⁵One of the main motivations of the WST from the perspective of machine learning was to provide a mathematical framework to understand the properties of convolutional neural networks and the origin of their success for visual object classification problems [46].

⁶ S_1 coefficients with $p = 2$ directly relate to the power spectrum [29,42].

⁷In practice, this term is estimated as described in [49].

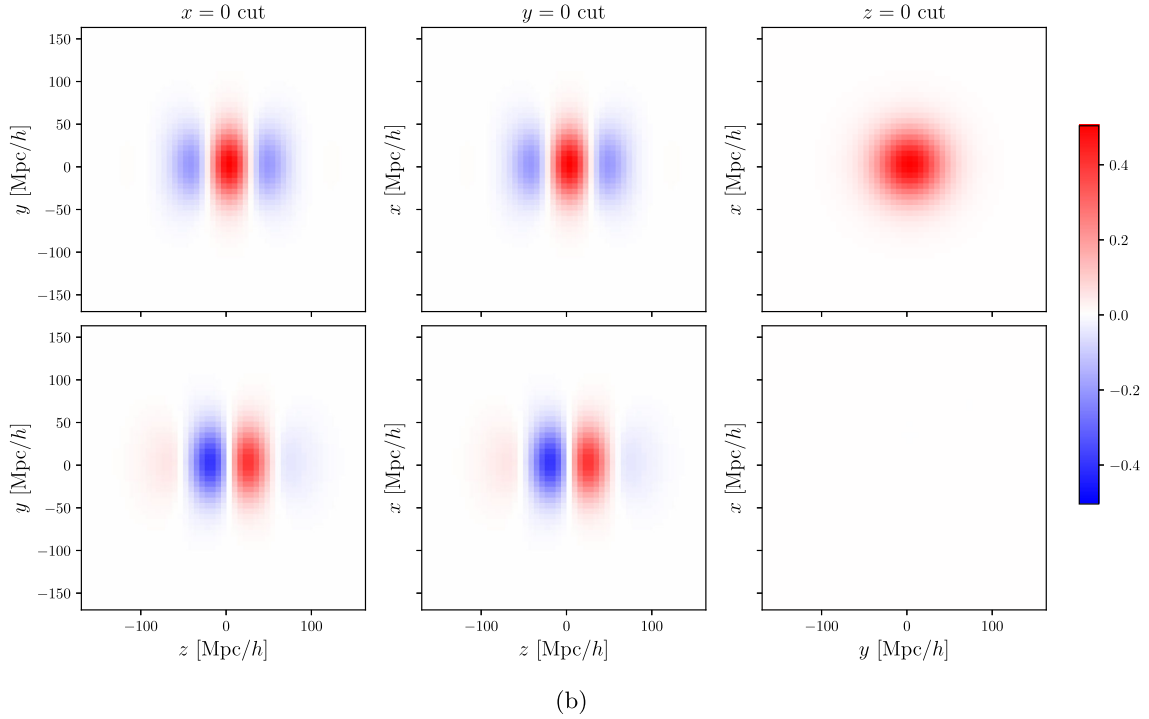
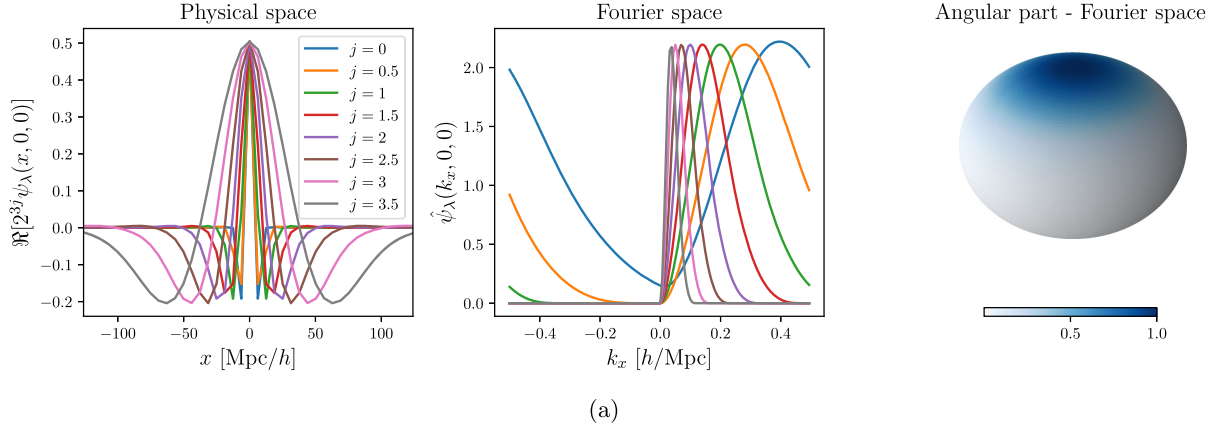


FIG. 2. Visualization of the 3D oriented wavelets designed for this work. (a) Radial (left and center) and angular (right) profiles for a subset of our 3D wavelets with fixed orientation. We show the radial profiles in both physical (left) and Fourier (center) space. The angular part is shown in Fourier space only. Our wavelets cover Fourier space with their passbands. (b) Planar cuts of the real (top row) and imaginary (bottom row) parts of $2^{3j}\psi_\lambda$ for $\lambda = (j, \vartheta, \varphi) = (3, 0, 0)$ (the wavelet is oriented along the z axis). The planar cuts are shown for $x, y, z = 0$ from left to right, respectively.

survey geometry as the galaxy catalog, and α is a normalizing factor selected to make $F(\mathbf{r})$ have a spatial mean of zero. Finally, we create a mesh representing $F(\mathbf{r})$ by interpolating the point cloud onto a 3D Cartesian grid using a triangular shape cloud interpolation scheme from NBODYKIT [50]. The grid spacing is chosen to be $\lambda_{\max}/2$, where $\lambda_{\max} = 2\pi/k_{\max}$, which thus satisfies the Nyquist criterion. The wavelets are also discretized on the same grid. Similarly to [27], we restrict our analysis to spatial frequencies lower than $k_{\max} = 0.5 h/\text{Mpc}$.

3. Reduced moments

With simple invariance assumptions on F that reflect the cosmological principle, it is possible to define so-called *reduced* WST moments which lead to a set of descriptive statistics for F of lower dimension, reduced variance, and which are often easier to interpret [30,31]. We first normalize the WST coefficients as follows:

$$\bar{S}_0[p] = \log(S_0[p]), \quad (9)$$

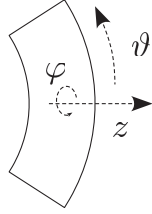


FIG. 3. Definition of the spherical coordinates with respect to the survey geometry. The redshift axis is z . The elevation angle ϑ measures deviation from z . The azimuthal angle φ_c measures rotation around z . The rounded rectangle represents a cut through the survey geometry.

$$\bar{S}_1[\lambda, p] = \log\left(\frac{S_1[\lambda, p]}{S_0[p]}\right), \quad (10)$$

$$\bar{S}_2[\lambda_1, \lambda_2, p] = \log\left(\frac{S_2[\lambda_1, \lambda_2, p]}{S_1[\lambda_1, p]}\right). \quad (11)$$

This procedure helps to disentangle the information measured by these different coefficients⁸ and evens their magnitudes.

Let us first mention that, by defining globally averaged WST coefficients in Sec. II C 2, we had already ignored the potential dependency of the coefficients with cosmological redshift, as if F were statistically homogeneous. In light of the cosmological principle, this assumption is only rigorously justified on surfaces of constant cosmological redshift. However, given the narrow redshift range of our data ($0.45 < z < 0.6$), we consider this approximation reasonable and leave the investigation of local WST statistics for cosmological inference to further work. Similarly, we expect statistical invariance of F with respect to rotations around the redshift axis. By neglecting the curvature of the survey domain and choosing e_z for the redshift axis, rotational invariance around e_z corresponds to an invariance with respect to the azimuthal angle φ (see Fig. 3 for an illustration).

We leverage this rotational invariance by introducing the following first- and second-order reduced moments:

$$\hat{S}_1[j, \vartheta, p] = \langle \bar{S}_1[\lambda, p] \rangle_\varphi, \quad (12)$$

$$\hat{S}_2[j_1, \vartheta_1, j_2, \vartheta_2, \delta\varphi, p] = \langle \bar{S}_2[\lambda_1, \lambda_2, p] \rangle_{|\varphi_1 - \varphi_2| = \delta\varphi}, \quad (13)$$

where $\langle \cdot \rangle_\varphi$ and $\langle \cdot \rangle_{|\varphi_1 - \varphi_2| = \delta\varphi}$ stand for averages over φ and the set of (φ_1, φ_2) pairs with fixed distance $|\varphi_1 - \varphi_2| = \delta\varphi$, respectively. The dependence of these coefficients on the angular variables quantifies the impact of redshift-space distortions.

In the following, we define the vector of summary statistics \mathbf{x} as the concatenation of zeroth-, first-, and second-order

⁸This makes \bar{S}_1 coefficients invariant to global scaling of F , and \bar{S}_2 coefficients invariant to linear filtering of F with top-hat passbands including that of ψ_{λ_1} .

reduced coefficients, that is 3675 coefficients in total. On a secondary level, we will also include in that vector the total number of objects in the input catalog n_{gal} following [27], as well as the shot noise power P_{shot} computed as in [49]. The impact of shot noise on the WST coefficients has not been studied yet, this is the reason why we include this P_{shot} coefficient in the (idealistic) idea that this impact will be directly learned from the data. Future works will address this issue in a more straightforward way.

D. Simulation-based inference

We use WST measurements from 20,000 mock galaxy catalogs constructed using the SimBIG forward model. Formally, these measurements correspond to samples $(\boldsymbol{\theta}_1, \mathbf{x}_1), \dots, (\boldsymbol{\theta}_n, \mathbf{x}_n) \sim p(\boldsymbol{\theta}, \mathbf{x})$ drawn from the joint distribution, where $n = 20,000$. With SBI, we estimate the posterior, $p(\boldsymbol{\theta}|\mathbf{x})$ solely based on these samples.

To do that, we train a conditional normalizing flow [for reviews on normalizing flows, see [51,52]] using the Python package SBI [53]. We use a masked autoregressive flow (MAF) architecture [54], which consists of a sequence of autoregressive affine transforms whose coefficients are given by the output of a multilayer perceptron (MLP) neural network. With $q_\phi(\boldsymbol{\theta}|\mathbf{x})$, a MAF architecture with unknown parameters ϕ , the training procedure consists in minimizing the following loss function:

$$\mathcal{L}(\phi) = D_{\text{KL}}[p(\boldsymbol{\theta}, \mathbf{x}) || q_\phi(\boldsymbol{\theta}|\mathbf{x})p(\mathbf{x})], \quad (14)$$

$$\approx \sum_{(\boldsymbol{\theta}_i, \mathbf{x}_i) \in T} \log p(\boldsymbol{\theta}_i|\mathbf{x}_i) - \log q_\phi(\boldsymbol{\theta}_i|\mathbf{x}_i), \quad (15)$$

where D_{KL} is the Kullback-Leibler divergence, and T is our training set, which is selected to be a subsample of the full data set $\{(\boldsymbol{\theta}_i, \mathbf{x}_i)\}_{1 \leq i \leq n}$.⁹ This minimization is equivalent to the maximization of the training score:

$$S(\phi) = \sum_{i \in T} \log q_\phi(\boldsymbol{\theta}_i|\mathbf{x}_i), \quad (16)$$

which we use in practice.

We perform this optimization using Adam [55]. It is regularized with an early stopping procedure,¹⁰ with a train/validation split of 90%/10%. Hyperparameters, including the number of affine transforms used in the MAF, number of blocks and hidden units of the MLP, dropout probability of the MLP dropout layers (for further regularization), learning rate, and batch size, are all tuned to achieve the best validation score. In practice, we explore the hyperparameter space with a random search monitored by WandB [56] for ~ 2000 models. We provide details on the

⁹For further details on this derivation, we refer the reader to [26].

¹⁰We stop the optimization when the validation score does not increase after 20 consecutive epochs.

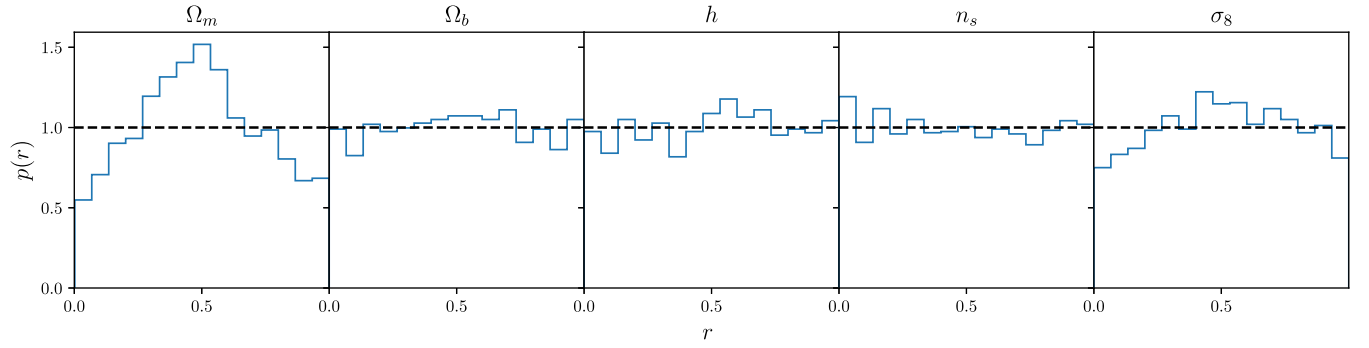


FIG. 4. Posterior accuracy test using SBC. We present the distributions of the rank statistics (Sec. III A) for each of the cosmological parameters. These are derived from the validation set used during posterior estimation. The rank statistics should be uniformly distributed for an ideal posterior estimate (black dashed). Our results indicate accurate estimates for Ω_b , h , and n_s but underconfidence for Ω_m and σ_8 .

“sweep” configuration used for this random search in Appendix B.

To further increase the score $\mathcal{S}(\phi)$, we define an ensemble model from a linear mixture of the ten best models previously optimized as follows [25,57]:

$$q_{\text{ens}}(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{10} w_i q_{\phi_i}(\boldsymbol{\theta}|\mathbf{x}). \quad (17)$$

The weights $\{w_i\}$ are positive numbers such that $\sum_{i=1}^{10} w_i = 1$ (thus ensuring correct normalization of q_{ens}). For this work, a uniform weighting, $w_i = 1/10$, has been sufficient to significantly improve the training and validation scores. Choosing a uniform weighting is suggested only after carefully narrowing down the hyperparameter space through successive random searches (see Appendix B for further details). Note that a more general strategy to optimize these weights without such costly successive random searches has been very recently investigated in this context [58]. This technique will be used in future work.

Let us finally mention that we also have explored a more expressive alternative to the MAF, namely a neural spline flow [59]. However, we faced overfitting issues that we were not able to resolve with standard regularization approaches. Further exploration is left for future work.

III. RESULTS AND DISCUSSION

We present our results in this section. In Sec. III A, we validate our inference pipeline, and study its robustness to the change of forward models following [27]. In Sec. III B, we present the results of our inference on the BOSS data.

A. Validation

We validate our posterior estimation in two stages. First, we assess whether $q_{\text{ens}}(\boldsymbol{\theta}|\mathbf{x})$ accurately estimates the true posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$ using simulation-based calibration (SBC) [60]. Second, we conduct the SimBIG “mock challenge,” where we assess the robustness of our inference pipeline using a suite of test simulations.

1. Posterior accuracy test

We perform SBC to check the validity of our posterior estimation following [60]. For each sample $(\boldsymbol{\theta}, \tilde{\mathbf{x}})$ of the validation set used during the training procedure¹¹ (Sec. II D), we generate $m = 2500$ independent samples from the posterior distribution $q_{\text{ens}}(\boldsymbol{\theta}|\tilde{\mathbf{x}})$, and then compute the (normalized) rank of each component of the vector $\tilde{\boldsymbol{\theta}}$ among the corresponding components of these samples.

We present the distributions of these ranks for each cosmological parameter in Fig. 4. For an ideal posterior estimation, the rank statistics should be uniformly distributed over $[0, 1]$. For Ω_b , h , and n_s , these are close to uniform, which indicates accurate predictions for these parameters. However, for Ω_m and σ_8 , we observe significant \cap -like patterns, which suggest underconfident predictions on these parameters. This is likely due to the limited size of the training set and underlines the challenges of posterior estimation in this high-dimensional context. Nevertheless, we consider that underconfidence does not undermine our analysis, but rather suggests that the constraining power of the WST statistics may be underexploited. Let us finally point out that the ensembling procedure [Eq. (17)] plays a significant role in avoiding overconfidence on Ω_b , h , and n_s . Indeed, separate SBC analyses of the ten best models have shown a slight tendency for overconfidence on these parameters, which has been corrected by the ensembling.

2. SimBIG mock challenge

Next, we assess the generalization capability and robustness of our inference pipeline by applying it to the SimBIG mock challenge data introduced in [27]. This data consists of the following test sets:

¹¹The limited amount of data prevents us from using an independent test set. However, since this validation set only served for the early stopping procedure mentioned in Sec. II D, we do not expect significantly different results even with a separate test set.

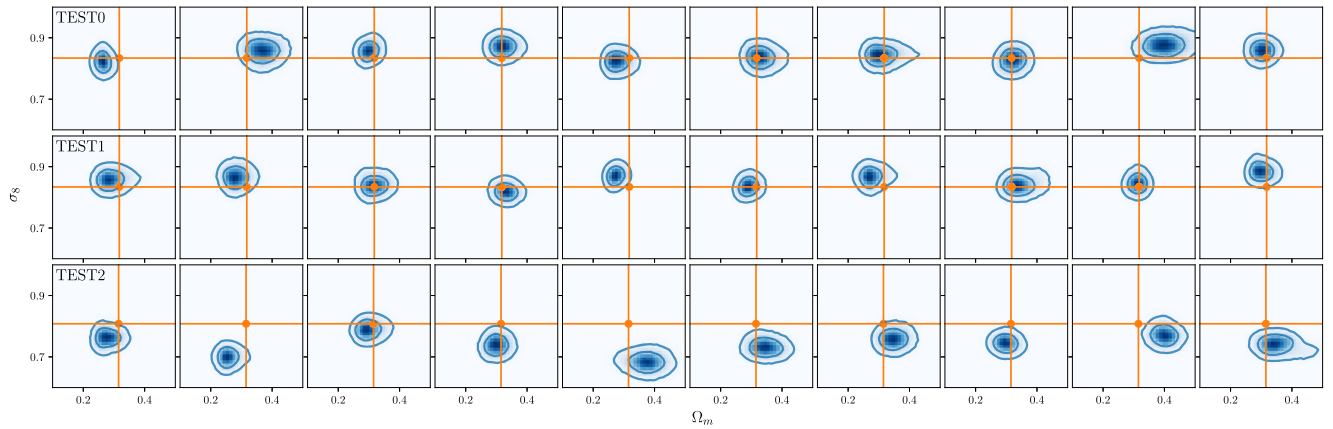


FIG. 5. Posterior distributions for (Ω_m, σ_8) inferred from WST measurements of 10 randomly selected test simulations from: TEST0 (top), TEST1 (middle), and TEST2 (bottom). The true parameters are marked in orange, and the contours represent the 68 and 95 percentiles. For TEST0 and TEST1, the true parameters lie expectedly within the percentiles of the posteriors. However, for TEST2, the true σ_8 tend to lie systematically higher than the inferred posteriors.

- (1) TEST0: 500 mock catalogs built from the SimBIG forward model described in Sec. II B, but using a set of 100 Quijote simulations at a fiducial cosmology (see Table I) and with HOD parameters drawn from a narrower distribution than the prior distribution.
- (2) TEST1: 500 mock catalogs built from a similar forward model with the same fiducial Quijote simulations, but a different friend-of-friend halo finder [61] and a simplified HOD model involving six parameters instead of nine.
- (3) TEST2: 1000 mock catalogs built from the AbacusSummit N -body simulations [62,63] and the CompaSo halo finder [64]. The simulations are constructed at a slightly different fiducial cosmology from Quijote (Table I). Otherwise, we use the same HOD model as the SimBIG forward model.

Since TEST0 uses the SimBIG forward model, it serves as a means to quantify the ability of our pipeline to generalize to new unseen data. On the other hand, TEST1 and TEST2, which use different forward models, help us to quantify the robustness of our pipeline.

We show in Fig. 5, for each of the test sets, examples of posterior distributions inferred from randomly selected test simulations. The top, center, and bottom rows present simulations from the TEST0, TEST1, and TEST2, respectively. We focus on (Ω_m, σ_8) since these are the parameters mostly significantly constrained by galaxy clustering. On the TEST0 and TEST1 examples, the true parameters are well within the 95% contours on almost all examples. The TEST0 results visually demonstrate the ability of our pipeline to generalize to new unseen data, while the TEST1 results indicate first elements of robustness to the change of the forward model. However, in comparison, the TEST2 results show contours that are almost always below the true parameters. This indicates a bias in the σ_8 predictions, suggesting that our WST analysis is not fully robust to changes in the forward model.

We now complement the visual validation results with a quantitative assessment. For each test simulation, we compute the posterior mean μ and standard deviation σ , and focus on the difference between μ and the true parameter θ^{fid} in units of σ for each cosmological parameter. Although we do not expect the posterior mean μ to be an unbiased estimate of the true parameter θ^{fid} , for a robust pipeline, we do expect a form of statistical consistency of these estimates across the test sets. We show in Fig. 6 the distributions of $(\mu - \theta^{\text{fid}})/\sigma$ for each cosmological parameter. We consider that our pipeline is robust if these distributions are invariant to the change of forward model.¹² These results indicate robustness for all parameters except σ_8 . The σ_8 histograms confirm what we find in Fig. 5: our posterior estimate systematically infers a lower value of σ_8 for TEST2. We also find that our posterior estimate systematically infers in average a higher value of σ_8 for TEST1. This demonstrates that the σ_8 posteriors are not robust to the alternative forward models considered in the SimBIG mock challenge. The results of the mock challenge imply the following: (1) the change of forward model has a significant impact on galaxy clustering, (2) the WST is sensitive to this impact, and (3) our inference pipeline erroneously attributes this to variations of σ_8 .

3. Addressing nonrobustness

We make a first attempt to address the robustness issues on σ_8 . Since we expect galaxy clustering to better agree across different forward models on larger scales closer to the linear regime, we introduce stringent scale cuts on the WST statistic. Introducing the scale index j_{min} , we define

¹²One interpretation of these distributions is that they are the likelihoods of a compression of the WST statistics. Thus, the comparison examines whether the compressed WST likelihoods are robust to change in the forward model.

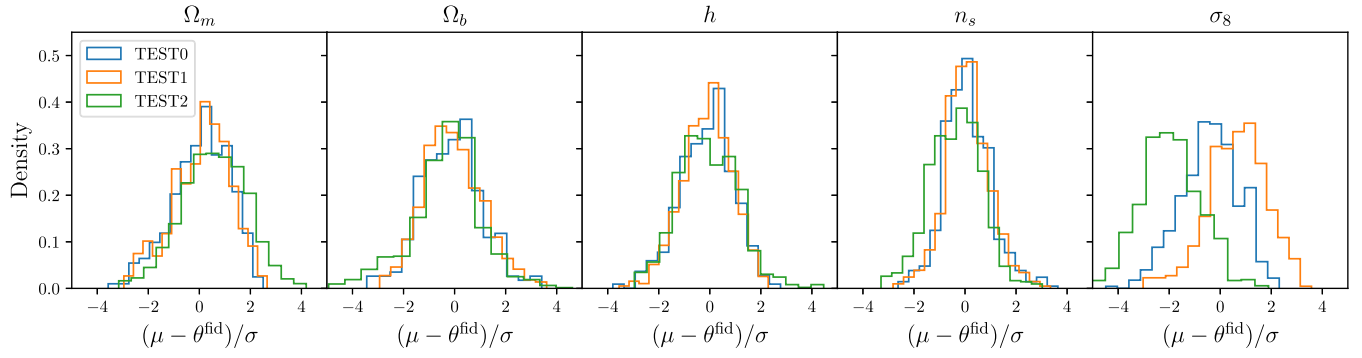


FIG. 6. Distributions of the differences between the posterior mean μ and the true parameter θ^{fid} normalized by the posterior standard deviation σ , for each cosmological parameter and test set. The distributions agree across the test sets for Ω_m , Ω_b , h , and n_s , but not for σ_8 . This indicates that our predictions for σ_8 are not robust to variations of the forward model.

new summary statistics where WST coefficients involving wavelets with $j < j_{\min}$ are discarded. We progressively increase j_{\min} and repeat our validation test. We are able to resolve the robustness issues raised by the mock challenge for $j_{\min} = 2$. This discards any information from modes with $k \gtrsim 0.3$ h/Mpc [see Fig. 2(a)]. Figure 7 shows that the $(\mu - \theta^{\text{fid}})/\sigma$ distributions obtained for $j_{\min} = 2$ now agree for σ_8 . Interestingly, the equivalent distributions for the other cosmological parameters are not impacted by this scale cut. This suggests that small-scale clustering was mostly the origin of the point (3) raised above.

B. BOSS WST analysis

We now analyze the BOSS CMASS galaxy sample with our inference pipeline. We show in Fig. 8 the posterior distribution of the cosmological parameters obtained from both the full WST (left) and $j_{\min} = 2$ WST statistics (right). For the $j_{\min} = 2$ results, we use an ensemble of 50 models instead of 10 (see Sec. II D) for reasons explained below. In both cases, we compare our results to two distinct power spectrum analyses on the same data: the SimBIG P_ℓ analysis [26,27] and the [15] P_ℓ analysis based on PT. The SimBIG and [15] P_ℓ analyses include scales down to

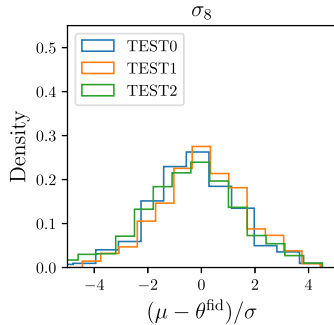


FIG. 7. Same as Fig. 6 but using summary statistics where WST coefficients with $j < j_{\min} = 2$ are discarded. This removes galaxy clustering information for scales smaller than $k_{\max} \approx 0.3$ h/Mpc . We focus on σ_8 . By imposing a stringent scale cut we improve the robustness of our WST analysis.

$k_{\max} = 0.5$ and 0.25 h/Mpc , respectively. We also show for reference in Figs. 10 and 11 equivalent plots that include the HOD parameters. We finally summarize in Table II all the posterior mean, standard deviation, median, and 68% credible interval. The $j_{\min} = 2$ WST analysis yields significantly tighter constraints on almost all parameters while being consistent with the P_ℓ analyses. Compared to the standard PT-based $P_\ell(k < 0.25$ $h/\text{Mpc})$ analysis, we tighten the 68% CI constraints for Ω_m , Ω_b , h , n_s , and σ_8 by a factor 1.2, 1.5, 0.8, 1.3, and 1.8, respectively. These are clear improvements for all parameters except h . The deterioration of the constraints for h could be explained by the smaller BOSS CMASS sample size that we use compared to [15].

However, we raise a notable concern on these results. For both the $j_{\min} = 0$ and $j_{\min} = 2$ analyses, we find that the separate models included in the ensemble [Eq. (17)] infer highly variable posteriors from one another (see Fig. 9). This is the case even for models with similar validation scores and that pass the posterior accuracy test. This could be a sign of model misspecification. In other words, the WST statistics of the BOSS sample could be an outlier compared to the training set. For $j_{\min} = 2$, this variability is higher than the $j_{\min} = 0$ case. As a result, we increased the number of models included in the ensemble from 10 to 50. We believe that this higher variability for $j_{\min} = 2$ is the consequence of the higher variance of large-scale information, which makes the training procedure more difficult. We leave a precise quantification of model misspecification and strategies to resolve it in this context for future work.

Finally, we note that our results have significant discrepancies with the WST analysis of [41]. We point out however that these two analyses are hardly comparable. Besides the fact that the variant of the WST statistics used in [41] significantly differ from the one we designed in this paper, the [41] analysis is conducted with a Gaussian likelihood assumption following the approach of standard clustering analyses. Furthermore, the mean function of their Gaussian likelihood is approximated by linearly extrapolating WST measurements from a fiducial cosmology using

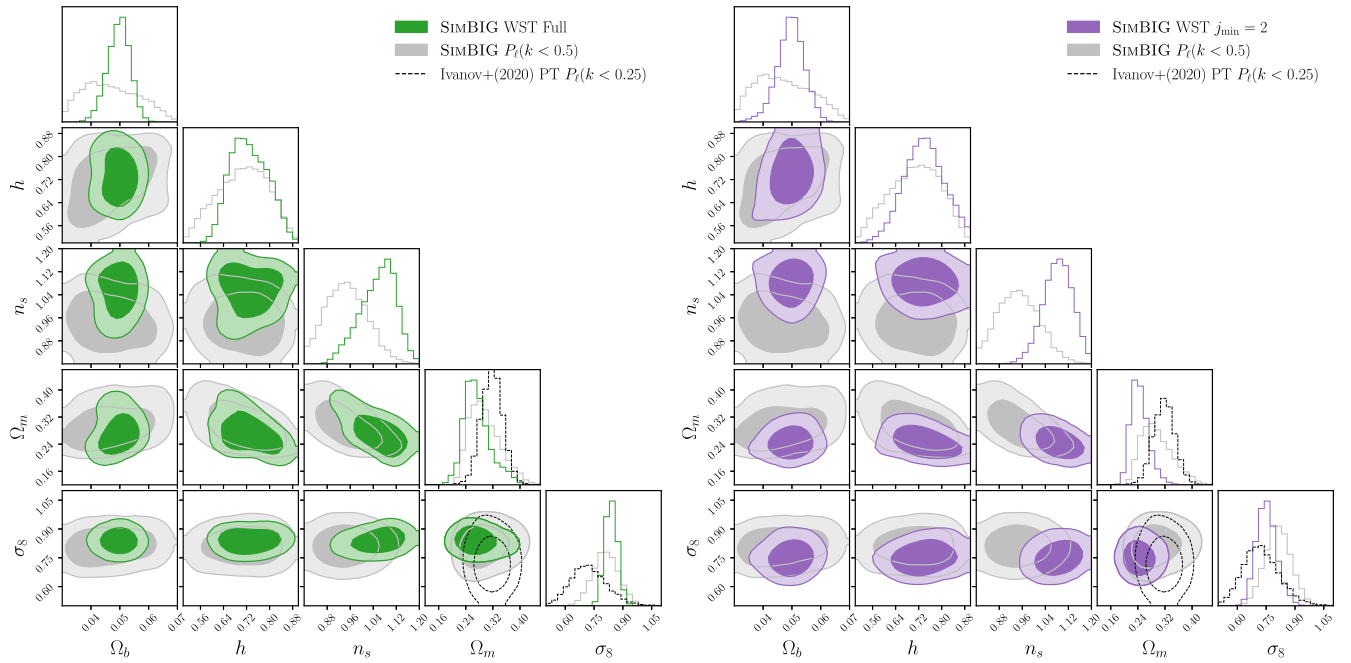


FIG. 8. Posterior distributions for the Λ CDM cosmological parameters inferred from the BOSS CMASS SGC sample using WST with SimBIG. On the left, we present constraints using the full WST statistics, which probe scales down to $k_{\max} = 0.5 h/\text{Mpc}$. On the right we present constraints using the $j_{\min} = 2$ WST statistics, where modes with $k \gtrsim 0.3 h/\text{Mpc}$ are discarded. The contours represent the 68 and 95 percentiles. We also show for comparison posteriors from the SimBIG P_ℓ analysis (gray, [26,27]) and the [15] PT based P_ℓ analysis (black dashed).

numerical derivatives, all estimated with the AbacusSummit simulations. They also assume that the covariance matrix of the likelihood is independent of cosmology and use approximate Patchy N body that were calibrated using only P_ℓ [65]. Our SBI analysis relaxes all of these assumptions and we rely only on full N -body simulations.

Overall, WST analyses present significant challenges in robustness compared to P_ℓ analyses. We show that the WST is more sensitive to the model-specific imprints on galaxy clustering than the power spectrum. In that sense, the challenges we face with robustness is in fact a demonstration of both the success of the WST and the failure of current forward models of galaxy clustering. If forward models of galaxy clustering converged with each

other at successfully describing the observed galaxy clustering, then we would be able to robustly exploit the sensitivity of WST and precisely constrain the cosmological parameters. In subsequent work, we will explore extensions of the SimBIG forward model that add more flexibility to mitigate the impact of model misspecification.

IV. CONCLUSION

We present cosmological constraints on the Λ CDM parameters Ω_m , Ω_b , h , n_s , and σ_8 from analyzing the BOSS CMASS SGC sample using WST statistics. Our analysis relies on a new variant of WST statistics which leverage symmetries of redshift-space data. Assuming

TABLE II. Constraints on the Λ CDM cosmological parameters inferred from the BOSS CMASS SGC using the WST with SimBIG. For each parameter, we report the posterior mean and standard deviation, as well as the 68% credible interval centered on the median. We also show for comparison the constraints from the SimBIG P_ℓ analysis [26,27] and the [15] PT-based P_ℓ analysis.

| Analysis | Quantity | Ω_M | Ω_B | H | N_S | σ_8 |
|--|-------------------|------------------------|---------------------------|------------------------|------------------------|------------------------|
| SimBIG WST (Full) | Mean $\pm \sigma$ | 0.28 ± 0.04 | 0.050 ± 0.004 | 0.73 ± 0.06 | 1.05 ± 0.06 | 0.84 ± 0.03 |
| | Median & 68% CI | $0.27^{+0.04}_{-0.03}$ | $0.051^{+0.004}_{-0.004}$ | $0.73^{+0.06}_{-0.06}$ | $1.06^{+0.06}_{-0.07}$ | $0.83^{+0.03}_{-0.03}$ |
| SimBIG WST $J_{\text{MIN}} = 2$ | Mean $\pm \sigma$ | 0.25 ± 0.03 | 0.050 ± 0.005 | 0.74 ± 0.07 | 1.08 ± 0.05 | 0.75 ± 0.05 |
| | Median & 68% CI | $0.24^{+0.03}_{-0.02}$ | $0.050^{+0.004}_{-0.005}$ | $0.74^{+0.07}_{-0.07}$ | $1.08^{+0.05}_{-0.05}$ | $0.75^{+0.05}_{-0.05}$ |
| SimBIG $P_\ell(K < 0.5 H/\text{MPC})$ | Mean $\pm \sigma$ | 0.29 ± 0.04 | 0.046 ± 0.008 | 0.70 ± 0.09 | 0.94 ± 0.07 | 0.81 ± 0.07 |
| | Median & 68% CI | $0.28^{+0.05}_{-0.04}$ | $0.046^{+0.009}_{-0.009}$ | $0.70^{+0.11}_{-0.10}$ | $0.93^{+0.08}_{-0.07}$ | $0.81^{+0.07}_{-0.08}$ |
| PT $P_\ell(K < 0.25 H/\text{MPC})$ | Mean $\pm \sigma$ | 0.32 ± 0.03 | 0.052 ± 0.007 | 0.68 ± 0.06 | 0.96 ± 0.05 | 0.73 ± 0.10 |
| | Median & 68% CI | $0.32^{+0.03}_{-0.03}$ | $0.052^{+0.007}_{-0.007}$ | $0.68^{+0.06}_{-0.06}$ | $0.95^{+0.07}_{-0.05}$ | $0.72^{+0.10}_{-0.09}$ |

conservative priors, we compute posterior distributions using the SimBIG SBI inference framework. These are learned by training MAF normalizing flows on 20,000 forward modeled mock galaxy catalogs with full survey realism. The hyperparameters are optimized to maximize the best validation score, and we use an ensembling procedure of the best models to further improve this score.

We validate our posterior estimates in terms of accuracy using SBC, and in terms of robustness using the SimBIG mock challenge, which relies on three different test sets constructed with different forward models. Our analysis passes the accuracy test but reveals robustness challenges. These take the form of a bias on σ_8 constraints when using test simulations generated from different forward models. This demonstrates that the WST statistics identify differences between the clustering of galaxies in these test sets, which are misinterpreted as variations of σ_8 . We are able to improve robustness by imposing a $j_{\min} = 2$ cut that discards small-scale clustering information beyond $k \sim 0.3$ h/Mpc.

Compared to the standard PT-based $P_\ell(k < 0.25$ h/Mpc) analysis, our WST $j_{\min} = 2$ analysis tightens the constraints on Ω_m , Ω_b , h , n_s , and σ_8 by factors 1.2, 1.5, 0.8, 1.3, and 1.8, respectively. However, despite the improved robustness of our WST $j_{\min} = 2$ analysis, we raise concerns on these results. We find significant variability of the posteriors across the separate models of our ensemble model. We attribute this to model misspecification. A more thorough quantification of model misspecification is still needed to further support the legitimacy of these results. We, therefore, refrain from interpreting the cosmological implication of our results.

Our analysis underlines key challenges in leveraging new summary statistics that can extract non-Gaussian information from galaxy clustering for cosmological inference. As our WST analysis demonstrates, highly informative summary statistics are more likely to reveal model misspecification. This is a particularly important consideration for galaxy clustering where our models have not converged on small, nonlinear scales and may yet be limited in accurately describing observations. This motivates the development of even more accurate and flexible forward models as well as strategies to account for model misspecification. In subsequent work, we will analyze the WST with an extended SimBIG forward model with more flexibility and present more reliable cosmological constraints. In accompanying papers [66,67], we present parallel SimBIG analyses of galaxy clustering using the galaxy bispectrum and a field-level approach involving convolutional neural networks, respectively.

ACKNOWLEDGMENTS

We thank Miles Cranmer, Mikhail Ivanov, Stéphane Mallat, Oliver Philcox, and Benjamin D. Wandelt for valuable discussions. We also acknowledge Mikhail Ivanov for providing us with posteriors used for comparison. J. H. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 101025187. Software used: Matplotlib [68], NBODYKIT [50], NumPy [69], PyTORCH [70], SBI [53], SciPy [71].

APPENDIX A: ISOTROPIC SPLINE WAVELET

We define the spherical Fourier-Bessel (SFB) transform of a square integrable scalar function $f(r, \theta, \varphi)$ as

$$\hat{f}_{lm}(k) = \sqrt{\frac{2}{\pi}} \int f(r, \theta, \varphi) j_l(kr) \overline{Y_l^m(\theta, \varphi)} r^2 \sin(\theta) dr d\theta d\varphi, \quad (\text{A1})$$

where Y_l^m functions are spherical harmonics and j_l are spherical Bessel functions. The SFB transform can be inverted with the following inversion formula:

$$f(r, \theta, \varphi) = \sqrt{\frac{2}{\pi}} \sum_{l=0}^{+\infty} \sum_{m=-l}^l \int \hat{f}_{lm}(k) k^2 j_l(kr) dk Y_l^m(\theta, \varphi). \quad (\text{A2})$$

Following [45], we introduce an isotropic scaling function Φ from its SFB coefficients:

$$\hat{\Phi}_{00}^{k_c}(k) = \frac{3}{2} B_3\left(\frac{2k}{k_c}\right) \quad \text{and} \quad \hat{\Phi}_{lm}^{k_c}(k) = 0 \quad \text{for } l, m \neq 0, \quad (\text{A3})$$

where k_c is a cutoff frequency, and B_3 a B-spline function of order 3 defined by

$$B_3(x) = \frac{1}{12} (|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3). \quad (\text{A4})$$

With $j_0(x) = \sin(x)/x$ and $Y_0^0(\theta, \varphi) = (2\sqrt{\pi})^{-1}$, the previous inversion formula leads to

$$\Phi^{k_c}(r) = \begin{cases} \frac{768}{8\sqrt{2\pi k_c^3} r^6} (4 \sin(k_c r/4) - k_c r \cos(k_c r/4)) \times \sin(k_c r/4)^3, & \text{if } r \neq 0 \\ \frac{k_c^3}{32\sqrt{2\pi}}, & \text{if } r = 0 \end{cases}. \quad (\text{A5})$$

This function allows us to define the following isotropic spline wavelet ψ^{iso} :

$$\psi^{\text{iso}}(r) = \Phi^{2k_c}(r) - \Phi^{k_c}(r). \quad (\text{A6})$$

TABLE III. wandb sweep configuration used for the final random search in the hyperparameter space during posterior estimation.

| Hyperparameter | Minimum value | Maximum value | Distribution | Quantization step |
|------------------------|--------------------|--------------------|----------------------|--------------------|
| Number of transforms | 5 | 11 | int_uniform | |
| Number of hidden units | 256 | 1024 | q_log_uniform_values | 32 |
| Number of blocks | 2 | 4 | int_uniform | |
| Dropout probability | 0.1 | 0.2 | q_uniform | 0.1 |
| Batch size | 20 | 100 | q_uniform | 5 |
| Learning rate | 5×10^{-6} | 5×10^{-5} | q_log_uniform_values | 1×10^{-6} |

TABLE IV. Hyperparameters of the ten best architectures (for the $j_{\min} = 0$ analysis).

| Rank | Best validation score | Number of transforms | Number of hidden units | Number of blocks | Dropout probability | Batch size | Learning rate |
|------|-----------------------|----------------------|------------------------|------------------|---------------------|------------|----------------------|
| 1 | 15.58 | 10 | 288 | 2 | 0.1 | 25 | 7×10^{-6} |
| 2 | 15.56 | 6 | 864 | 2 | 0.1 | 40 | 6×10^{-6} |
| 3 | 15.48 | 8 | 864 | 2 | 0.1 | 75 | 5×10^{-6} |
| 4 | 15.47 | 11 | 544 | 2 | 0.1 | 40 | 5×10^{-6} |
| 5 | 15.46 | 8 | 416 | 3 | 0.1 | 30 | 7×10^{-6} |
| 6 | 15.44 | 6 | 704 | 3 | 0.2 | 50 | 5×10^{-6} |
| 7 | 15.44 | 7 | 672 | 3 | 0.1 | 35 | 7×10^{-6} |
| 8 | 15.43 | 9 | 608 | 3 | 0.1 | 50 | 7×10^{-6} |
| 9 | 15.39 | 6 | 416 | 2 | 0.1 | 60 | 6×10^{-6} |
| 10 | 15.38 | 6 | 416 | 2 | 0.1 | 75 | 1.2×10^{-5} |

We report for reference its central frequency:

$$k_0 = \int_0^{+\infty} k^3 |\hat{\psi}^{\text{iso}}(k)|^2 dk / \int_0^{+\infty} k^2 |\hat{\psi}^{\text{iso}}(k)|^2 dk, \quad (\text{A7})$$

$$= \frac{10647k_c}{13394} \approx 0.8k_c. \quad (\text{A8})$$

APPENDIX B: POSTERIOR ESTIMATION HYPERPARAMETER OPTIMIZATION

We give further details on the hyperparameter optimization described in Sec. II D.

The hyperparameters of the MAF architecture were optimized through a succession of random searches

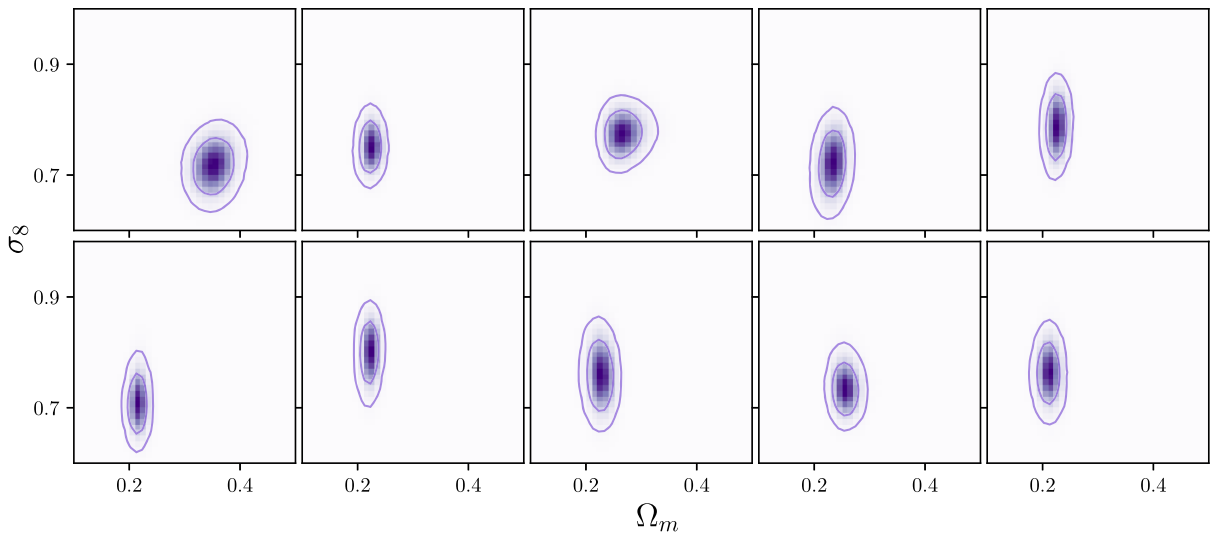


FIG. 9. Observational constraints on σ_8 and Ω_m obtained across the 10 best models included in the ensemble model with $j_{\min} = 2$. This important variability suggests a form of model misspecification.

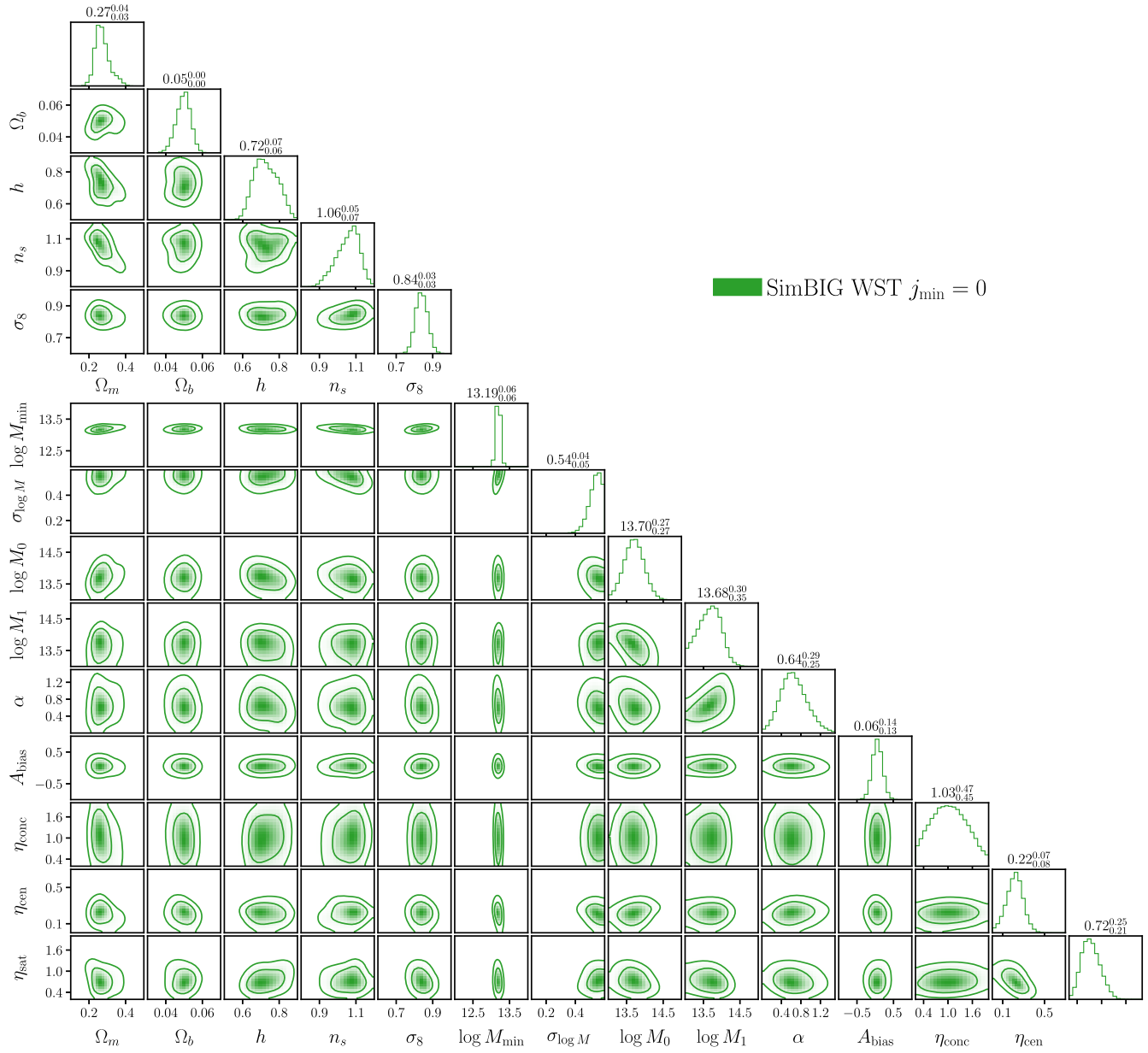


FIG. 10. Same as Fig. 8 (left) but also including predictions of the HOD parameters.

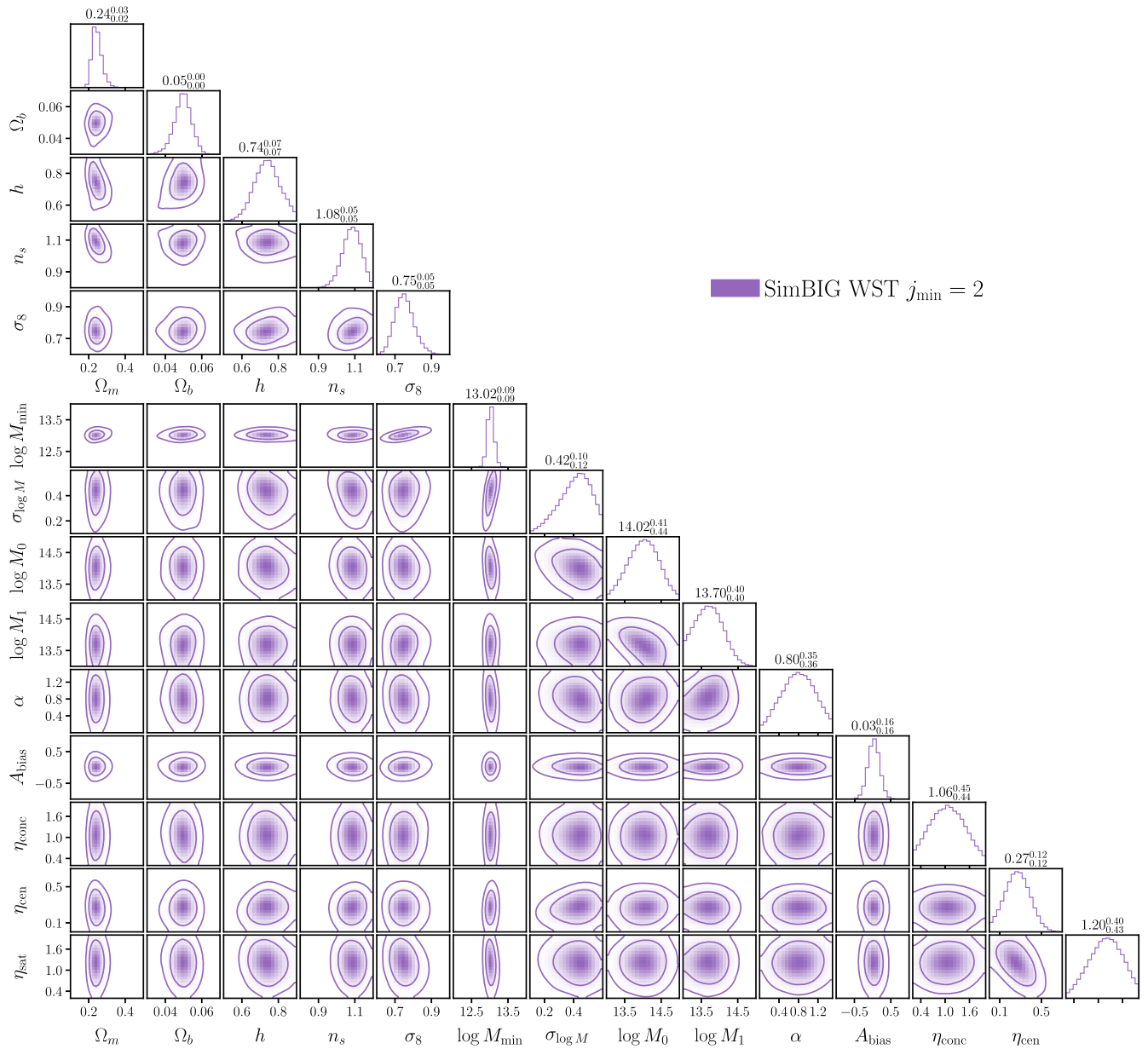


FIG. 11. Same as Fig. 8 (right) but also including predictions of the HOD parameters.

monitored by `wandB` [56]. To guarantee reproducibility of this work, we give in Table III the “sweep” configuration used in our last random search, and in Table IV the hyperparameters of the ten best models (in terms of best validation score) used to define our ensemble posterior distribution. Note that these architectures were all trained using a consistent train-validation split.

We also give details on the MLP network that constitutes the conditioner of the MAF. It uses ReLU activation functions and includes batch normalization layers. These batch normalization layers significantly improved the best validation score.

APPENDIX C: ADDITIONAL RESULTS

- [1] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot *et al.*, *Astron. J.* **142**, 72 (2011).
- [2] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey, R. H. Barkhouser, J. E. Bautista, A. Beifiori, A. A. Berlind *et al.*, *Astron. J.* **145**, 10 (2013).
- [3] A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland, O. Ballester *et al.* (DESI Collaboration), [arXiv:1611.00036](https://arxiv.org/abs/1611.00036).
- [4] M. Takada, R. S. Ellis, M. Chiba, J. E. Greene, H. Aihara, N. Arimoto, K. Bundy, J. Cohen, O. Doré, G. Graves *et al.*, *Publ. Astron. Soc. Jpn.* **66**, R1 (2014).
- [5] R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, [arXiv:1110.3193](https://arxiv.org/abs/1110.3193).
- [6] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon *et al.*, [arXiv:1503.03757](https://arxiv.org/abs/1503.03757).
- [7] C. Hahn, F. Beutler, M. Sinha, A. Berlind, S. Ho, and D. W. Hogg, *Mon. Not. R. Astron. Soc.* **485**, 2956 (2019).
- [8] C. F. Park, E. Allys, F. Villaescusa-Navarro, and D. Finkbeiner, *Astrophys. J.* **946**, 107 (2023).
- [9] N. Kaiser, *Mon. Not. R. Astron. Soc.* **227**, 1 (1987).
- [10] A. J. S. Hamilton, *Mon. Not. R. Astron. Soc.* **289**, 285 (1997).
- [11] J. A. Peacock, S. Cole, P. Norberg, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. D. Cannon, M. Colless, C. Collins, W. Couch *et al.*, *Nature (London)* **410**, 169 (2001).
- [12] M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, A. J. S. Hamilton, G. R. Knapp *et al.*, *Phys. Rev. D* **74**, 123507 (2006).
- [13] L. Guzzo, M. Pierleoni, B. Meneux, E. Branchini, O. Le Fèvre, C. Marinoni, B. Garilli, J. Blaizot, G. De Lucia, A. Pollo *et al.*, *Nature (London)* **451**, 541 (2008).
- [14] F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, H. Gil-Marín, J. N. Grieb, N. Hand, F.-S. Kitaura *et al.*, *Mon. Not. R. Astron. Soc.* **466**, 2242 (2016).
- [15] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, *J. Cosmol. Astropart. Phys.* **05** (2020) 042.
- [16] C. Chen, Y. Li, F. Villaescusa-Navarro, S. Ho, and A. Pullen, [arXiv:2012.05472](https://arxiv.org/abs/2012.05472).
- [17] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, *Phys. Rep.* **367**, 1 (2002).
- [18] V. Desjacques, D. Jeong, and F. Schmidt, *Phys. Rep.* **733**, 1 (2018).
- [19] C. Hahn and F. Villaescusa-Navarro, *J. Cosmol. Astropart. Phys.* **04** (2021) 029.
- [20] G. D'Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, [arXiv:2206.08327](https://arxiv.org/abs/2206.08327).
- [21] O. H. E. Philcox and M. M. Ivanov, *Phys. Rev. D* **105**, 043517 (2022).
- [22] E. Massara, F. Villaescusa-Navarro, C. Hahn, M. M. Abidi, M. Eickenberg, S. Ho, P. Lemos, A. M. Dizgah, and B. Régaldo-Saint Blancard, *Astrophys. J.* **951**, 70 (2023).
- [23] J. Hou, A. Moradinezhad Dizgah, C. Hahn, and E. Massara, *J. Cosmol. Astropart. Phys.* **03** (2023) 045.
- [24] K. Cranmer, J. Brehmer, and G. Louppe, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055 (2020).
- [25] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, *Mon. Not. R. Astron. Soc.* **488**, 4440 (2019).
- [26] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régaldo-Saint Blancard, and M. M. Abidi, *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2218810120 (2023).
- [27] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. M. Dizgah, B. R.-S. Blancard, and M. M. Abidi, *J. Cosmol. Astropart. Phys.* **04** (2023) 010.
- [28] S. Mallat, in *Proceedings of the 2010 18th European Signal Processing Conference* (2010), pp. 716–720.
- [29] S. Mallat, *Commun. Pure Appl. Math.* **65**, 1331 (2012).
- [30] E. Allys, F. Levrier, S. Zhang, C. Colling, B. Regaldo-Saint Blancard, F. Boulanger, P. Hennebelle, and S. Mallat, *Astron. Astrophys.* **629**, A115 (2019).
- [31] B. Regaldo-Saint Blancard, F. Levrier, E. Allys, E. Bellomi, and F. Boulanger, *Astron. Astrophys.* **642**, A217 (2020).
- [32] A. K. Saydjari, S. K. N. Portillo, Z. Slepian, S. Kahraman, B. Burkhart, and D. P. Finkbeiner, *Astrophys. J.* **910**, 122 (2021).
- [33] S. Cheng, Y.-S. Ting, B. Ménard, and J. Bruna, *Mon. Not. R. Astron. Soc.* **499**, 5902 (2020).
- [34] S. Cheng and B. Ménard, *Mon. Not. R. Astron. Soc.* **507**, 1012 (2021).
- [35] D. T. Chung, *Mon. Not. R. Astron. Soc.* **517**, 1625 (2022).
- [36] B. Greig, Y.-S. Ting, and A. A. Kaurov, *Mon. Not. R. Astron. Soc.* **513**, 1719 (2022).
- [37] B. Greig, Y.-S. Ting, and A. A. Kaurov, *Mon. Not. R. Astron. Soc.* **519**, 5288 (2023).
- [38] M. Eickenberg, G. Exarchakis, M. Hirn, and S. Mallat, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg *et al.* (Curran Associates, Inc., 2017), Vol. 30.
- [39] M. Eickenberg, E. Allys, A. Moradinezhad Dizgah, P. Lemos, E. Massara, M. Abidi, C. Hahn, S. Hassan, B. Regaldo-Saint Blancard, S. Ho, S. Mallat, J. Andén, and F. Villaescusa-Navarro, [arXiv:2204.07646](https://arxiv.org/abs/2204.07646).
- [40] G. Valogiannis and C. Dvorkin, *Phys. Rev. D* **105**, 103534 (2022).
- [41] G. Valogiannis and C. Dvorkin, *Phys. Rev. D* **106**, 103509 (2022).
- [42] E. Allys, T. Marchand, J.-F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. Mallat, *Phys. Rev. D* **102**, 103506 (2020).
- [43] F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys *et al.*, *Astrophys. J. Suppl. Ser.* **250**, 2 (2020).
- [44] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, *Astrophys. J.* **762**, 109 (2013).
- [45] F. Lanusse, A. Rassat, and J. L. Starck, *Astron. Astrophys.* **540**, A92 (2012).
- [46] S. Mallat, *Phil. Trans. R. Soc. A* **374**, 20150203 (2016).
- [47] J. Bruna and S. Mallat, *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1872 (2013).
- [48] H. A. Feldman, N. Kaiser, and J. A. Peacock, *Astrophys. J.* **426**, 23 (1994).
- [49] R. Scoccimarro, *Phys. Rev. D* **92**, 083532 (2015).
- [50] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, *Astron. J.* **156**, 160 (2018).

- [51] I. Kobyzev, S. D. Prince, and M. A. Brubaker, *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3964 (2021).
- [52] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, *J. Mach. Learn. Res.* **22**, 1 (2022).
- [53] 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).
- [54] G. Papamakarios, T. Pavlakou, and I. Murray, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30.
- [55] D. P. Kingma and J. Ba, in *International Conference on Learning Representations*, edited by Y. Bengio and Y. LeCun (2015).
- [56] L. Biewald, Experiment tracking with weights and biases (2020).
- [57] B. Lakshminarayanan, A. Pritzel, and C. Blundell, [arXiv:1612.01474](https://arxiv.org/abs/1612.01474).
- [58] Y. Yao and B. Régaldo-Saint Blancard (to be published).
- [59] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), Vol. 32.
- [60] S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman, [arXiv:1804.06788](https://arxiv.org/abs/1804.06788).
- [61] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **292**, 371 (1985).
- [62] N. A. Maksimova, L. H. Garrison, D. J. Eisenstein, B. Hadzhiyska, S. Bose, and T. P. Satterthwaite, *Mon. Not. R. Astron. Soc.* **508**, 4017 (2021).
- [63] L. H. Garrison, D. J. Eisenstein, D. Ferrer, N. A. Maksimova, and P. A. Pinto, *Mon. Not. R. Astron. Soc.* **508**, 575 (2021).
- [64] B. Hadzhiyska, D. Eisenstein, S. Bose, L. H. Garrison, and N. Maksimova, *Mon. Not. R. Astron. Soc.* **509**, 501 (2022).
- [65] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín, H. Guo, G. Yepes, A. Klypin, C. G. Scóccola *et al.*, *Mon. Not. R. Astron. Soc.* **456**, 4156 (2016).
- [66] C. Hahn *et al.* (SimBIG Collaboration), preceding paper, *Phys. Rev. D* **109**, 083534 (2024).
- [67] P. Lemos *et al.* (SimBIG Collaboration), following paper, *Phys. Rev. D* **109**, 083536 (2024).
- [68] J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
- [69] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, *Nature (London)* **585**, 357 (2020).
- [70] A. Paszke, S. Gross, F. Massa *et al.*, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle *et al.* (Curran Associates, Inc., 2019), Vol. 32.
- [71] P. Virtanen, R. Gommers, T. E. Oliphant *et al.*, *Nat. Methods* **17**, 261 (2020).