# Cosmological constraints from the redshift-space galaxy skew spectra

Jiamin Hou[*]

*Department of Astronomy, University of Florida, 211 Bryant Space Science Center,*
*Gainesville, Florida 32611, USA and*
*Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstrasse 1,*
*85748 Garching bei München, Germany*

Azadeh Moradinezhad Dizgah[†]

*Laboratoire d'Annecy de Physique Théorique (CNRS/USMB), F-74940 Annecy, France*
*and Département de Physique Théorique, Université de Genève,*
*24 quai Ernest Ansermet, 1211 Genève 4, Switzerland*

ChangHoon Hahn

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

Michael Eickenberg

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

Shirley Ho

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

Pablo Lemos

*Department of Physics, Université de Montréal, Montréal, 1375 Avenue Thérèse-Lavoie-Roux,*
*Quebec City H2V 0B3, Canada;*
*Mila—Quebec Artificial Intelligence Institute, Montréal, 6666 Rue Saint-Urbain,*
*Quebec City H2S 3H1, Canada and Center for Computational Mathematics, Flatiron Institute,*
*162 5th Avenue, New York, New York 10010, USA*

Elena Massara

*Waterloo Centre for Astrophysics, University of Waterloo,*
*200 University Ave W, Waterloo, Ontario N2L 3G1, Canada*
*and Department of Physics and Astronomy, University of Waterloo,*
*200 University Ave W, Waterloo, Ontario N2L 3G1, Canada*

Chirag Modi

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

Liam Parker

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

Bruno Régaldo-Saint Blancard

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

[*]jiamin.hou@ufl.edu
[†]azadeh.moradinezhad@lapth.cnrs.fr

Extracting the non-Gaussian information of the cosmic large-scale structure (LSS) is vital in unlocking the full potential of the rich datasets from the upcoming stage-IV galaxy surveys. Galaxy skew spectra serve as efficient beyond-two-point statistics, encapsulating essential bispectrum information with computational efficiency akin to power spectrum analysis. This paper presents the first cosmological constraints from analyzing the full set of redshift-space galaxy skew spectra of the data from the SDSS-III BOSS, accessing cosmological information down to nonlinear scales. Employing the SIMBIG forward modeling framework and simulation-based inference via normalizing flows, we analyze the CMASS-SGC subsample, which constitute approximately 10% of the full BOSS data. Analyzing the scales up to $k_{\max} = 0.5\ h^{-1}$ Mpc, we find that the skew spectra improve the constraints on $\Omega_{\mathrm{m}}, \Omega_{\mathrm{b}}, h$, and $n_s$ by 34%, 35%, 18%, 10%, respectively, compared to constraints from previous SIMBIG power spectrum multipoles analysis, yielding $\Omega_{\mathrm{m}} = 0.288^{+0.024}_{-0.034}, \Omega_{\mathrm{b}} = 0.043^{+0.005}_{-0.007}, h = 0.759^{+0.104}_{-0.050}, n_{\mathrm{s}} = 0.918^{+0.041}_{-0.090}$ (at 68% confidence limit). On the other hand, the constraints on $\sigma_8$ are weaker than from the power spectrum. Including the big bang nucleosynthesis (BBN) prior on baryon density reduces the uncertainty on the Hubble parameter further, achieving $h = 0.750^{+0.034}_{-0.032}$, which is a 38% improvement over the constraint from the power spectrum with the same prior. Compared to the SIMBIG bispectrum (monopole) analysis, skew spectra offer comparable constraints on larger scales ($k_{\max} < 0.3\ h^{-1}$ Mpc) for most parameters except for $\sigma_8$.

## I. INTRODUCTION

The LSS holds valuable clues about the Universe's origin, composition, and evolution, making it a crucial arena for testing fundamental physics. In the coming years, the next-generation galaxy surveys including Dark Energy Spectroscopic Survey [DESI; [1]], *Euclid* [2], Nancy Roman Space Telescope [Roman; [3]], SPHEREx [4], and Subaru Prime Focus Spectrograph [PFS; [5]] will provide a detailed spectroscopic view of the three-dimensional distribution of galaxies across significantly larger cosmic volumes than current surveys. This expanded scope promises a marked reduction in statistical uncertainties in the measured clustering statistics of galaxies, yielding unparalleled constraints on the standard ΛCDM model and opening up exciting new avenues for exploring new physics.

The nonlinear nature of the LSS implies that the commonly applied two-point correlation functions are insufficient in capturing the full cosmological information of galaxy clustering. Therefore, to exploit the upcoming galaxy surveys to their full potential, constructing summary statistics that capture the non-Gaussian clustering information is crucial. This has spurred a significant amount of work in the literature focusing on constructing optimal summary statistics, including bispectrum [6–9], skew spectra [10–13], marked 2-point functions [14–16], density split statistics [17], and wavelet scattering transforms [18,19]. More ambitious approaches to extract information at the field level without relying on any summary statistics [20–24] have also been a topic of growing interest. The potential of these approaches has been extensively explored using Fisher forecasts (e.g., Hou *et al.* [13], Eickenberg *et al.* [18], Agarwal *et al.* [25], Hahn and Villaescusa-Navarro [26], Valogiannis and Dvorkin [27], Massara *et al.* [28], Paillas *et al.* [29]), and very recently, several of them (e.g.,

Paillas *et al.* [30], Hahn *et al.* [31], Régaldo-Saint Blancard *et al.* [32], Lemos *et al.* [33], Valogiannis, Yuan, and Dvorkin [34]) have been applied to the existing data from Sloan Digital Sky Survey (SDSS)-III Baryon Oscillation Spectroscopic Survey Data Release 12 [BOSS DR12; [35,36]].

Among various proposed summary statistics, the galaxy skew spectra [10–12,37,38] are optimal, yet computationally efficient proxy statistics for the galaxy bispectrum. They are constructed from the cross-correlations of the observed galaxy density field with appropriately weighted quadratic galaxy fields. For parameters such as galaxy biases, the growth rate of structure, and the amplitude of primordial non-Gaussianity (PNG), by construction, the skew spectra fully capture the information of the bispectrum in the linear regime [11]. Despite these encouraging results from Fisher forecasts [11,13], the applicability of the full redshift space galaxy skew spectra to observational data, in particular down to nonlinear scales, has not been studied. This investigation is the focus of the current article.

The standard approach to cosmological inference from the LSS is based on a Bayesian likelihood analysis and requires: (1) a theoretical model of the observable, (2) an explicit form of the likelihood, and (3) schemes for correcting the observational systematics in the measurements. These requirements can hinder the maximal exploitation of upcoming LSS data.

First, if using perturbation theory (PT) to describe the observables, the limited validity of the perturbative description restricts the theory-based analyses to relatively large scales ($k_{\max} \simeq 0.25$ Mpc$^{-1}\ h$) [39]. Despite the considerable recent developments in perturbative description of the LSS [40–48], the PT-based analyses are intrinsically limited to the weakly nonlinear regime. Moreover, apart from the galaxy bispectrum, for some of the powerful

recently proposed statistics such as wavelet scattering transforms, a perturbative description is missing altogether.

The second limitation of the standard analysis lies in the assumption of a Gaussian likelihood for the data vector. This assumption, which leans on the central limit theorem, does not hold on large scales with low signal-to-noise ratio and on small scales where modes are highly correlated. Thus an explicit assumption of the Gaussian likelihood, can potentially bias the parameter constraints [49,50].

Lastly, the standard analysis accounts for observational systematics such as targeting, imaging, and spectroscopic incompleteness by applying weights designed to correct their effects on measured observables [51–53]. However, even at the precision level of current surveys, the weights are insufficient at correcting for effects such as fiber collisions, which can bias the clustering measurements on weakly nonlinear scales [54–56]. Moreover, it is important to note that these correction techniques are currently optimized only for two-point statistics and their validity for other summary statistics has not been thoroughly tested.

To overcome these challenges, simulation-based inference [SBI; [57,58]] is emerging as a promising avenue [59]. Without a strong assumption for the form of the likelihood, SBI enables estimation of the parameter posteriors (or the likelihood function) by leveraging high-fidelity numerical simulations to forward model the observables and employing deep generative models from machine learning for efficient parameter inference. In the context of galaxy clustering, several recent studies have explored the potential of SBI, employing either a set of summary statistics or at the field-level (e.g., de Santi *et al.* [60,61], Tucci and Schmidt [62], Modi *et al.* [63]). In terms of application of the SBI to observational data, [64,65] introduced Simulation-Based Inference of Galaxies (SIMBIG), a forward-modeling framework that enables performing SBI on a spectroscopic galaxy sample, employing either summary statistics or at the field-level. The SIMBIG forward model capitalizes on high-fidelity $N$-body simulations, coupled with the state-of-the-art halo occupation distribution (HOD) model to accurately simulate nonlinear galaxy clustering. Notably, SIMBIG integrates pertinent observational effects such as survey masks and fiber collisions. In contrast to the common approach of accounting for the fiber collisions via correction weights, SIMBIG forward models their effect. In a series of recent papers, SIMBIG has been used to obtain cosmological constraints from SDSS-III BOSS CMASS galaxies. By employing several statistics (including the power spectrum [64,65], the bispectrum [31], and the wavelet scattering transforms [32]), and by performing field-level analysis using convolutions neural networks [33], these works have highlighted the substantial potential of SBI in obtaining constraints that surpass the standard analysis of galaxy clustering [66].

This is the first paper using the SIMBIG framework to obtain cosmological constraints from analyzing the full set of redshift-space galaxy skew spectra, accessing information down to nonlinear scales. As part of this analysis, we also present a new estimator for the skew spectra that extends the previously developed estimators for periodic-box simulations [12,13] to account for survey mask and angular systematic weights so that it can be used on observations.

The structure of the paper is as follows: Sec. II provides an overview of the galaxy skew spectra and the new estimator for measuring the skew spectra on a realistic galaxy sample. In Sec. III, we briefly describe the observational data used in our analysis, and in Sec. IV we describe the SIMBIG pipeline (the forward modeling, the simulation-based inference, and the validation tests). We show the results in Sec. V, including the validation tests on synthetic data and the inferred posterior distribution of cosmological parameters from the BOSS-CMASS sample. Finally, we draw our conclusions in Sec. VI.

We provide additional details and complementary results in a series of appendices. In Appendix A, we provide explicit forms of the skew spectra, in Appendix C, we discuss the constraints on HOD parameters, and in Appendix E we investigate the dependence of the inferred constraints on several analysis choices including the outlier removal scheme, the removal of shot noise, the scale cuts, and the smoothing scale.

## II. SUMMARY STATISTICS: GALAXY SKEW SPECTRA

The galaxy skew spectra are efficient proxy statistics for the bispectrum and offer several advantages over the bispectrum; first, they are simple to interpret since they are functions of a single wavenumber and not triangle shapes. Second, the computational cost of measuring them from the data is comparable to that of the power spectrum. While accounting for all bispectrum triangles requires $\mathcal{O}(N^2)$ operations capturing the full information of the bispectrum using the skew spectra requires $\mathcal{O}(N \log N)$ operations, where $N = (k_{\max}/\Delta k)^3$ is the number of 3D Fourier-space grid points at which the fields are evaluated. Third, for standard Bayesian inference, if estimating the covariance matrices from mocks, the skew spectra requires a significantly smaller number of mocks given the significantly smaller size of the data vector. Lastly, in comparison to the bispectrum [67], accounting for the survey window function is expected to be considerably simpler for the skew spectra [68].

### A. Theoretical construction

The skew spectra are constructed by cross-correlating the observed galaxy density field, $\delta_g$, with an appropriately weighted square of it, $\mathcal{S}_n$ [10,12];

$$\mathcal{P}_{\mathcal{S}_n\delta}(k) = \int \frac{d\hat{k}}{4\pi} \langle \delta_{\mathrm{g}}(-\mathbf{k}) \mathcal{S}_n(\mathbf{k}) \rangle. \tag{1}$$

The quadratic field, $\mathcal{S}_n$, is a product of two fields in configuration space (evaluated at the same spacial positions), thus, a convolution in Fourier space [69];

$$\mathcal{S}_n(\mathbf{k}) = \int_{\mathbf{q}} \mathcal{D}_n(\mathbf{q}, \mathbf{k} - \mathbf{q}) \delta_g^R(\mathbf{q}) \delta_g^R(\mathbf{k} - \mathbf{q}), \qquad (2)$$

with $\int_{\mathbf{q}} \equiv \int d^3q/(2\pi)^3$. The fields involved in the convolution are smoothed, $\delta_g^R(\mathbf{k}) = \delta_g(\mathbf{k}) W_R(\mathbf{k})$, to limit the contribution of small-scale fluctuations to skew spectra. While a top-hat filter in Fourier space amounts to imposing a sharp $k$-cut on contribution of small-scale modes to the convolution integral, to avoid edge effects in the measurements of the skew spectra, we use a Gaussian smoothing kernel $W_R(\mathbf{k}) = \exp(-k^2R^2/2)$. In perturbative analysis of the skew spectra [10–12], smoothing the field (with relatively large $R$) ensures that only the Fourier modes on the semi-linear regime are included. For SBI performed in this paper, the smoothing serves the purpose of limiting the analysis to scales where the forward model is most trusted [70].

The weights, $\mathcal{D}_n(\mathbf{q}, \mathbf{k} - \mathbf{q})$, are constructed such that for a theoretical bispectrum that can be decomposed as a sum of product-separable contributions (in terms of the three Fourier modes forming a triangle), the skew spectra correspond to the maximum likelihood estimators for parameters appearing as overall amplitudes of individual contributions in the sum. Explicit forms of the kernels $\mathcal{D}_n(\mathbf{q}, \mathbf{k} - \mathbf{q})$ are given in Appendix A. This implies that, by construction, considering full set of 14 galaxy skew spectra in redshift space (3 in real space) fully capture the information of the bispectrum on large scales for parameters such as galaxy biases, the amplitude of primordial power spectrum and bispectrum, and the growth rate of structure [10,11] [71]. For other cosmological parameters (including $\Lambda$CDM parameters and the sum of neutrino masses), by performing numerical Fisher forecasts on QUIJOTE simulations [72,13] demonstrated that the skew spectra can considerably improve the constraints from the power spectrum, at the level comparable to the bispectrum.

In addition to the clustering component in Eq. (1), each of the skew spectra receives a shot noise contribution, which in the Poisson limit is given by

$$\mathcal{P}_{\mathcal{S}_n\delta}^{\text{shot}}(k) = \frac{1}{2} \int d\mu_k \left\{ \left[ \frac{1}{\bar{n}^2} + \frac{P_g(\mathbf{k})}{\bar{n}} \right] J_{\mathcal{D}_n}(\mathbf{k}) + \frac{2}{\bar{n}} \tilde{J}_{\mathcal{D}_n}(\mathbf{k}) \right\},$$

$$(3)$$

where $P_g(\mathbf{k})$ is the full 3D galaxy power spectrum and the kernels $J_{\mathcal{D}_n}$ and $\tilde{J}_{\mathcal{D}_n}$ are given by

$$J_{\mathcal{D}_n}(\mathbf{k}) = \int_{\mathbf{q}} W_R(\mathbf{q}) W_R(\mathbf{k} - \mathbf{q}) \mathcal{D}_n(\mathbf{q}, \mathbf{k} - \mathbf{q}), \quad (4)$$

$$\tilde{J}_{\mathcal{D}_n}(\mathbf{k}) = \int_{\mathbf{q}} W_R(\mathbf{q}) W_R(\mathbf{k} - \mathbf{q}) D_n(\mathbf{q}, \mathbf{k} - \mathbf{q}) P_g^R(\mathbf{q}). \quad (5)$$

## B. Estimators for Survey Data

Previous works have focused on the analysis of the skew spectra on simulated dark matter distribution and halo catalogs in *N*-body simulations with periodic boundary conditions [12,13]. In order to analyze the data from BOSS survey, we extend the previous estimators to measure the skew spectra using on Fast Fourier Transforms (FFT) [12,13] to incorporate the survey geometry and observational systematics. We thus construct an equivalent of Feldman-Kaiser-Peacock [FKP; [73]] estimators for the power spectrum and bispectrum, where the galaxy sample is characterized by $n_g$ galaxies at positions $\mathbf{x}$ and a randoms catalog with $n_r$ objects to incorporate the radial and angular selection functions,

$$\delta_g(\mathbf{x}) = I_{33}^{-1/3} w_{\text{fkp}}(\mathbf{x}) [n'_g(\mathbf{x}) - \alpha n'_r(\mathbf{x})]. \qquad (6)$$

Here, $n'_g(\mathbf{x}) = w_c(\mathbf{x}) n_g(\mathbf{x})$ and $n'_r(\mathbf{x}) = w_{c,r}(\mathbf{x}) n_r(\mathbf{x})$ are the weighted number densities of the observed and random galaxies, and $\alpha(\ll 1)$ is their ratio. As in previous SIMBIG analyses, for each observed galaxy, we assign the completeness weight of $w_c = w_{\text{sys}} w_{\text{noz}}$, where $w_{\text{sys}}$ is an angular systematic weight based on stellar density and seeing conditions, and $w_{\text{noz}}$ is a redshift failure weight. Since for the fiber collisions, the standard weighting scheme has been shown to be inaccurate [74], we forward model them in our pipeline instead of using weights as is commonly done in other analyses of BOSS data (for further details, see Hahn *et al.* [65]). We have also included the FKP weight $w_{\text{fkp}}(\mathbf{x}) = [1 + \bar{n}(g)P_0]^{-1}$ (assuming $P_0 = 10^4$ [$h^{-1}$ Mpc]$^3$) to reduce the variance of the estimator [75].[76] The normalization factor of $I_{33}$ in Eq. (6) is defined in analogy to that of the bispectrum [6] since both skew spectra and bispectrum have three powers of galaxy density fields,

$$I_{33} = \int d^3\mathbf{x} \bar{n}_g^3(\mathbf{x}) w_c^3(\mathbf{x}) w_{\text{fkp}}^3(\mathbf{x})$$

$$\equiv \sum_{i=1}^{N_g} \bar{n}_g^2(\mathbf{x}_i) w_c^2(\mathbf{x}_i) w_{\text{fkp}}^3(\mathbf{x}_i), \qquad (7)$$

where $\bar{n}_g$ is the mean number density per redshift bin for the observed galaxy galaxies.

As a simplifying assumption, we consider a fixed line-of-sight (LoS) direction when constructing the quadratic fields. While in computing the cross-correlation between the linear and the quadratic field the varying LoS is accounted for as in the [77] estimator. The skew spectra correspond to the monopole of the cross-correlation between the input galaxy field and its square.

For implementation, we used the NBODYKIT [78] https://github.com/bccp/nbodykit [79]. We apply the weights $\mathcal{D}_n(\mathbf{q}, \mathbf{k} - \mathbf{q})$ [see Eq. (1)] to the first two linear density fields to obtain the quadratic field. We then prepare both the quadratic field and the third linear density field as

FKPCATALOG objects, an NBODYKIT-based interface for simultaneous modeling of a data and a random catalog. These two FKPCATALOG objects are then passed to the CONVOLVEDFFTPOWER module for cross-correlating the quadratic field and the linear field.

In analogy to the power spectrum and the bispectrum FKP estimators, we subtract the Poisson shot noise from the measured skew spectra. The power-spectrum-independent contribution to shot noise in Eq. (3) is estimated by inserting two unity fields, whereas the kernels that depend on the power spectrum are computed by correlating a smoothed 3D power spectrum $P_g^R(\mathbf{q}) = W_R(\mathbf{q})P_g(\mathbf{q})$ and a unity field. Due to the survey geometry, the full 3D power spectrum is non-Hermitian, whereas the unity fields are real-valued Hermitian fields and the third dimension of the field is halved after the FFT. In practice, we concatenate two Hermitian unity fields to match the dimensionality.

As in the previous SIMBIG analysis of the galaxy power spectrum and bispectrum, in addition to the skew spectra, we also include the average galaxy number density, $\bar{n}_g$. Therefore, in total, the data vector that we pass to the SBI pipeline has 1401 elements, with $k_{\min} = 0.003$ Mpc$^{-1}h$ that corresponds to the fundamental frequency of the BOSS SGC, and $\Delta k = 0.005$ Mpc$^{-1}h$ [80]. We explore the impact of using a different $k_{\min}$ and its implications on our analysis.

## III. OBSERVATIONAL DATA: BOSS CMASS GALAXY SAMPLE

As in previous SIMBIG analyses [31–33,64–66], we consider the CMASS galaxy sample (consisting mainly of luminous red galaxies) of BOSS DR12 [81,82]. Since the SIMBIG forward model employs QUIJOTE $N$-body simulations with the box size of $L_{\text{box}} = 1$ Gpc$^{-1}h$ to model the nonlinear clustering of dark matter, we only use the data in the southern galactic cap (SGC) and further apply

angular cuts of DEC $> -6°$ and $-25° <$ RA $< 28°$ and a radial cut corresponding to redshift range of $0.45 < z < 0.6$. The resulting galaxy sample spans $\sim$10% of the full BOSS data and about 50% of the CMASS-SGC sample, with 109,636 galaxies and a number density of $\sim 4.2 \times 10^{-4}$ [Mpc$^{-1}h$]$^3$.

## IV. SIMBIG PIPELINE

SIMBIG is a framework for performing SBI to infer cosmological constraints from galaxy clustering data, leveraging high-fidelity simulations to forward model the observations and deep generative models to perform inference by estimating the posterior of model parameters. Figure 1 shows a schematic sketch of the pipeline. In this section, we describe in more detail the different components of the pipeline, including the forward model, the SBI, and the validation tests.

### A. Forward model

The current implementation of the SIMBIG forward model is specifically tailored to produce galaxy catalogs that match the statistical properties of the BOSS-CMASS sample. This involves four integral components: nonlinear dark matter distributions from $N$-body simulations, a halo finder to construct halo catalogs, an HOD model to populate halos with galaxies, and survey realism (including fiber collisions and survey geometry).

SIMBIG uses $N$-body simulations from QUIJOTE (high-resolution) suite, featuring 2,000 $\Lambda$CDM cosmologies run in a Latin-hypercube configuration for parameters $\{\Omega_m, \Omega_b, h, n_s, \sigma_8\}$ [72]. The simulations are run with Gadget-III TreePM + SPH code [83], evolving $1024^3$ dark matter particles in a periodic box of $L_{\text{box}} = 1$ $h^{-1}$ Gpc on the side. The particles are evolved from redshift $z = 127$ to $z = 0$, with the initial condition generated from the second-order Lagrangian perturbation theory (2LPT). Then, we
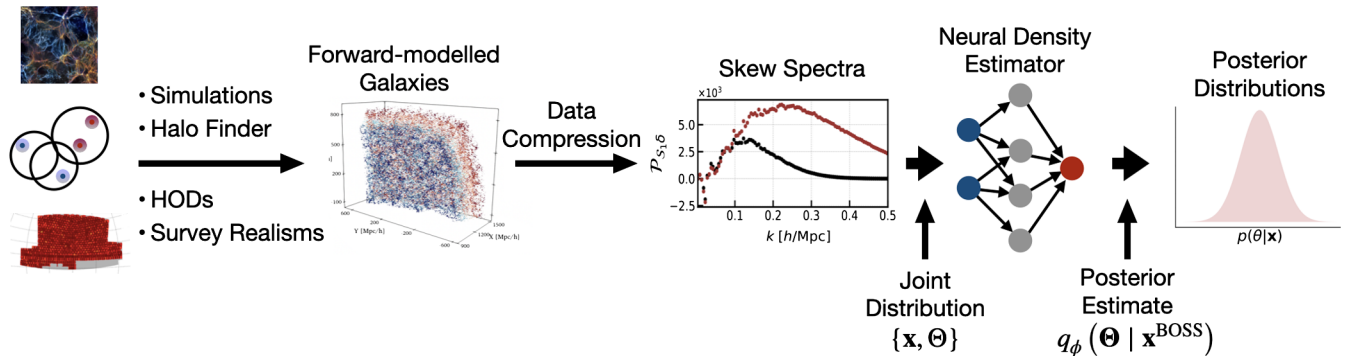


FIG. 1. Schematic view of the SIMBIG pipeline; to generate synthetic data that is statistically indistinguishable from BOSS observations, we utilize 2,000 $N$-body simulations from the QUIJOTE suite run at different cosmologies, apply Rockstar halo finder to identify halos, populate halos with galaxies using an extended (9-parameter) HOD model, and finally apply the survey realism. The resulting 20,000 synthetic mocks are then used to train normalizing flows as neural density estimators to obtain approximate posterior distributions, $q_\phi$, for cosmological and HOD parameters from given summary statistics $\mathbf{x}$, which in this paper refer to the skew spectra.

construct halo catalogs using the ROCKSTAR halo finder [84], which identifies halos using phase space information of the dark matter particles.

To populate halos with galaxies, SIMBIG employs the HOD framework [85]. We consider a state-of-the-art HOD model [86] that includes the standard parameters for characteristic mass scale for halos to host a central galaxy, $\log M_{\mathrm{min}}$, the scatter of halo mass at fixed galaxy luminosity, $\sigma_{\log M}$, the minimum halo mass for halos to host a satellite galaxy, $\log M_0$, the characteristic mass scale for halos to host a satellite galaxy, $\log M_1$, and power-law index for the mass dependence of satellite occupation $\alpha$. In addition, it also incorporates assembly bias, $A_{\mathrm{bias}}$, concentration bias of satellite, $\eta_{\mathrm{conc}}$, velocity biases for central, $\eta_{\mathrm{cen}}$, and satellite, $\eta_{\mathrm{sat}}$, galaxies. For each halo catalog, we generate 10 galaxy catalog with different values of the HOD parameters. In total, our training dataset consists of 20,000 galaxy mock catalogs.

The final step is to apply the survey realism. This includes imposing the angular footprint of the survey and the veto masks (including masking for bright stars, centerpost, bad fields, and collision priority). We also model the fiber collision effect for galaxy pairs within an angular resolution $< 62''$ [87]. In the radial direction, we apply the redshift cut $0.45 < z < 0.6$.

In total, the forward model involves 5 cosmological parameters and 9 HOD parameters. We chose uniform prior on all parameters except for the assembly bias, where a Gaussian prior centered at zero was implemented. The exact prior range can be found in Table 1 in [65]. The wide range of the HOD parameter priors results in a large variation in the galaxy number density and clustering amplitude.

While the SIMBIG pipeline offers a random catalog that is 40 times the size of the real BOSS data sample, to avoid noise induced by the random catalog, we generated an additional random catalog that is 35 times the size of the training set with the highest total galaxy counts. We use the MAKE SURVEY toolkit [88] to trim the angular footprint of the random.

In order to improve the training and reduce the dynamic range of the data vector, we preprocess the skew spectra of the training catalogs, which exhibit substantial variation in mean number density and clustering amplitude. We remove outliers and normalize the skew spectra in order to ensure that all features contribute meaningfully (Appendix E).

### B. Simulation-based inference

The 20,000 forward modeled galaxy mocks are formally samples drawn from the joint probability distribution $p(\mathbf{\Theta}, \mathbf{x})$ of the model parameters, $\mathbf{\Theta}$, and a summary statistic, $\mathbf{x}$, measured from observational data. Using this training dataset, we estimate the full posterior distribution over the parameters conditional on the observations, $p(\mathbf{\Theta}|\mathbf{x})$. In SIMBIG, the SBI is performed using neural density estimation with normalizing flows (NFs) [89–91], which enable efficient estimation of the posteriors with a limited number of simulations.

NFs are powerful tools for constructing expressive probability distributions describing the data using a simple base probability distribution, $\pi(\mathbf{z})$, and a chain of trainable smooth bijective transformations (diffeomorphisms), $f$, to map the base distribution to the target one. The base distribution is often chosen to be a multivariate Gaussian, which is easy to sample from. The transformations $f$ are chosen to have a tractable Jacobian so that the target distribution can be computed from the base distribution via a change of variables. A neural network is trained to obtain the flow (find the transformation $f$) that best approximate the posterior, $q_\phi(\mathbf{\Theta}, \mathbf{x}) \approx p(\mathbf{\Theta}, \mathbf{x})$, by minimizing the Kullback–Leibler (KL) divergence between $p(\mathbf{\Theta}, \mathbf{x}) = p(\mathbf{\Theta}|\mathbf{x})p(\mathbf{x})$ and $q_\phi(\mathbf{\Theta}|\mathbf{x})p(\mathbf{x})$. In practice, this is achieved by maximizing the log-likelihood over the training set.

For the analysis of the skew spectra [92], we use a masked autoregressive flow [MAF; [93]] architecture implemented in the SBI PYTHON package [94].[95] MAF uses masked autoencoder for distribution estimation [MADE; [96]] as a building block and by stacking several of them combines normalizing flows with an autoregressive design, which is well-suited for estimating conditional probabilities such as the posteriors. We split the data into training and validation sets with a 90/10 split, and use only the 90% for training, leaving the rest for validation test presented in Sec. V B 1. For the split, we randomly select from 20,000 simulations, irrespective of their cosmologies and HOD parameters.

We perform the optimization using ADAM [97] and tune the hyperparameters (number of blocks and hidden layers, dropout probability, learning rate, batch size, and number of transformations for the normalizing flows) via experimentation using the OPTUNA hyperparameter optimization framework [98] to obtain the best validation score. The details of the configurations for the OPTUNA search are provided in table I. We train 2,771 models for our baseline configuration with $R = 5\ h^{-1}$ Mpc and $k_{\mathrm{max}} = 0.5$ Mpc$^{-1}h$. For each additional configurations that we use to study the dependence of our results on various analysis choices, we train 2,000–3,000 models. To ensure stability and robustness of our results against model variations [99,100], we ensemble average (with equal weights) the top 10 models in each configuration, $q_\phi(\mathbf{\Theta}|\mathbf{x}) = \sum_{i=1}^{10} q_\phi^i(\mathbf{\Theta}|\mathbf{x})/10$. Overall, we do not observe a large variability in the top selected models and they provide consistent results.

### C. SIMBIG mock challenge

As in the previous SIMBIG analyses, we perform a mock challenge using additional 2,000 simulated galaxy catalogs to validate the posterior estimates, $q_\phi$. The test simulations

TABLE I.   Hyperparameter configuration used for OPTUNA optimization.

| Hyperparameter | Min | Max | Type | Distribution | Step |
| --- | --- | --- | --- | --- | --- |
| Number of transforms | 5 | 11 | int | Uniform | N/A |
| Number of hidden units | 256 | 1024 | int | Log uniform | N/A |
| Number of blocks | 2 | 4 | int | Uniform | N/A |
| Dropout probability | 0.1 | 0.3 | Float | Uniform | 0.1 |
| Batch size | 20 | 100 | int | Uniform | 5 |
| Learning rate | $5 \times 10^{-6}$ | $5 \times 10^{-5}$ | Float | Log uniform | $10^{-6}$ |

are organized in three datasets as introduced in [65]. The first two test sets (TEST0, TEST1) are constructed from the QUIJOTE $N$-body suite, while the third set (TEST2) uses ABACUSSUMMIT $N$-body simulations [101,102]. Below we list a few important differences between the three test simulations;

  (i)  TEST0: 500 galaxy mocks generated with the same SIMBIG forward model as the training data, but using a set of 100 QUIJOTE simulations at a fixed fiducial cosmology [103], and the HOD model with parameters sampled from a narrower distribution compared to the prior distribution. There are 5 galaxy catalogs per cosmology with 9-parameter HOD models.

 (ii)  TEST1: 500 galaxy mocks built from the same QUIJOTE simulation as in TEST0 but applying a different halo finder and a simpler HOD model. Halos are identified using friend-of-friend [FoF; [104]] halo finder 5 galaxy catalogs are constructed per cosmology with 5-parameter HOD models [105].

(iii)  TEST2: 1,000 mocks built from the ABACUSSUMMIT base simulations. The simulations evolve $6,912^3$ particles in a periodic box of $L_{\rm box} = 2\ h^{-1}$ Gpc. We use 25 realizations of base simulations, which have the same cosmology and differ in their phase [106]. Each simulation is divided into 8 subvolumes to obtain effectively 200 independent boxes. Halos are identified using COMPASO halo finder [107]. For each box, 5 galaxy catalogs are constructed using the standard 5-parameter HOD models.

## V. RESULTS

In this section, we begin by presenting the measured skew spectra of the CMASS-SGC galaxy sample. Subsequently, we detail the conducted validation tests and finally present the first cosmological constraints from the analysis of the full set of galaxy skew spectra on an observational dataset including nonlinear scales. We compare the cosmological and HOD constraints with the previous SIMBIG power spectrum (monopole and quadrupole) and bispectrum (monopole) results, and describe the impact of a number of analysis choices on the inferred constraints.

### A. Measured skew spectra on BOSS-CMASS-SGC sample

Figure 2 shows the measured skew spectra (including the clustering and shot noise components), applying smoothing scales of $R = 5\ h^{-1}$ Mpc in red, and $R = 10\ h^{-1}$ Mpc in black. Figure 3 shows the Poisson shot-noise contribution, consisting of three contributions in Eq. (3), which have not been previously measured.

The shapes of the measured spectra are consistent with those on synthetic halo and galaxy catalogs used in [13]. As identified before, the skew spectra shapes fall into three categories: (i) "constant type" involving square of galaxy density field $\delta^2$, and characterized by a large-scale peak and a sharp decline toward smaller scales. (ii) "displacement type" involving the operator $\partial_i \partial_j \nabla^2$, and distinguished by a positive bump and a zero-crossing feature on large scales for LoS-dependent contributions. (iii) "tidal type," involving the tidal operator $S^2$, and marked by a negative dip and anti-correlation with other skew spectra on smaller scales. The skew spectra $\mathcal{P}_{S_n \delta}$ with $n > 3$ optimally capture the LoS dependence of the bispectrum (see Appendix A for specific forms of the kernels $\mathcal{S}_n$).

On small scales, the skew spectra all approach zero due to smoothing, as shown in Fig. 2. Comparing Figs. 2 and 3 we also notice that the clustering component is smaller than the Poisson shot noise around $k = 0.4\ h^{-1}$ Mpc, coincide with the averaged inter-particle separation $\Delta s \sim 15\ h^{-1}$ Mpc for the BOSS galaxy sample.

### B. Posterior validation

Before applying the estimated posterior, $q_\phi$, on the skew spectra of the observed CMASS galaxies, we perform two validation tests to asses whether $q_\phi$ can *robustly* infer *unbiased* posteriors of the cosmological parameters. We describe these tests in this section.

#### 1. Accuracy test: Simulation-based calibration

Our first test aims to assess the accuracy of the approximate posterior, $q_\phi$. Although, theoretically, a sufficiently large number of simulations and optimally trained normalizing flows would inherently guarantee this accuracy by design (as we minimized the KL divergence between $q_\phi$ and the true posterior), SIMBIG employs a
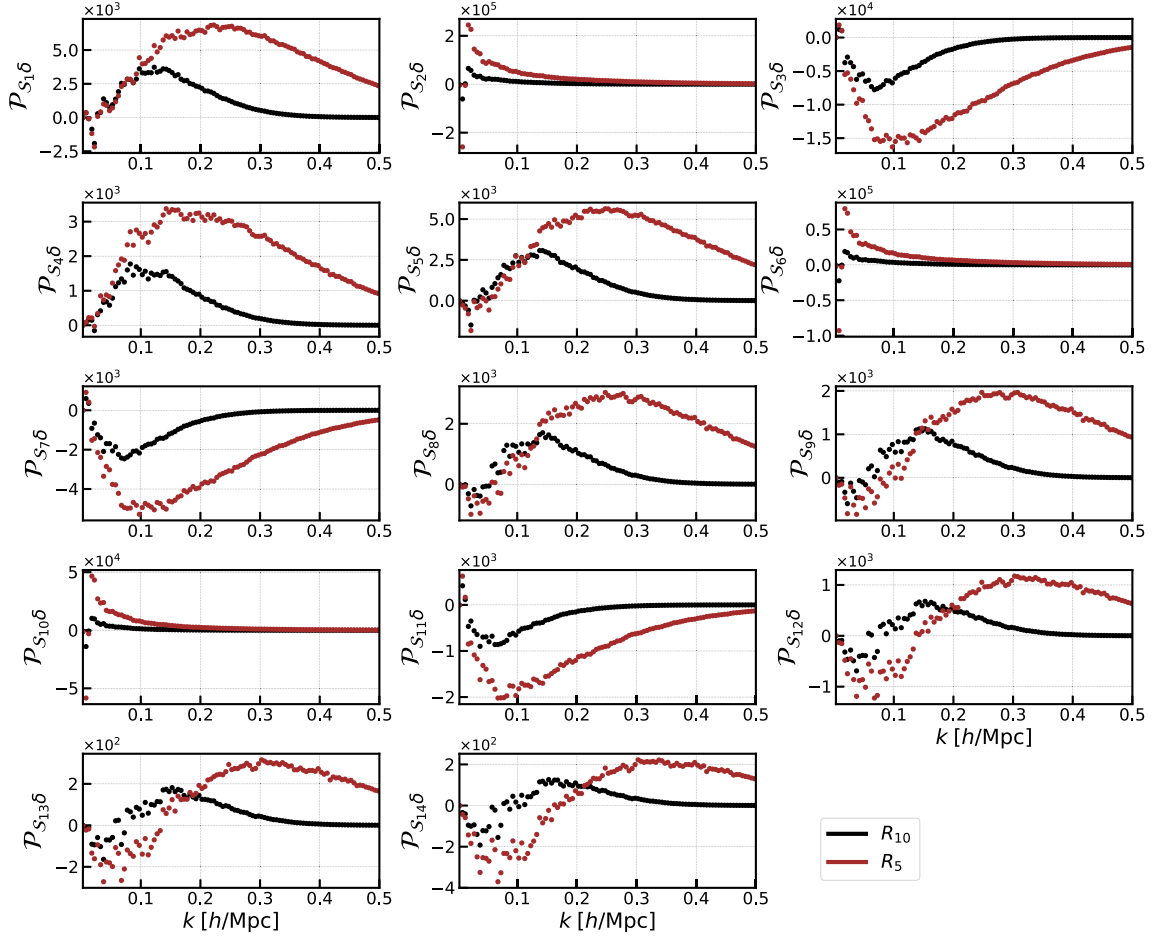
FIG. 2.    Measurements of the 14 skew spectra from the subsample of the BOSS DR12-SGC galaxies before the shot noise subtraction. Different colors correspond to two smoothing scales of $R = 5\ h^{-1}$ Mpc (in red) and $R = 10\ h^{-1}$ Mpc (in black). Due to smoothing, skew spectra approach zero at small scales.

relatively small number of simulations for the size of the data vector and dimensionality of the parameter space. Hence, assessing the accuracy of the estimated posterior is pivotal. To conduct this evaluation, we employ simulation-based calibration [SBC; [108]] on the validation dataset that was excluded from the training of our posterior estimate (see Sec. IV B).

For each test simulation with skew spectra, $\mathbf{x}_{\text{test},i}$, we draw $N_{\text{sample}} = 10,000$ samples from $q_\phi(\boldsymbol{\Theta}|\mathbf{x}_{\text{test},i})$. We then, calculate the fraction of the parameters that are below the true cosmological parameter of the simulation:
$$\text{rank}(\boldsymbol{\Theta}_{\mathrm{x}}) \equiv \sum_{i=1}^{N_{\text{sample}}} [1|_{\boldsymbol{\Theta}_{\mathrm{x},i} < \boldsymbol{\Theta}_{\mathrm{x}}^{\text{fid}}}]/N_{\text{sample}}.$$

Figure 4 shows the distribution of the rank statistics for $k_{\text{max}} = 0.5$ Mpc$^{-1}h$ and $R = 5\ h^{-1}$ Mpc. In the case of independent samples from the true posterior, we should expect uniformly distributed rank statistics. We see that the rank statistics for all parameters are nearly flat. For $\Omega_{\mathrm{m}}$, we see a slight ∩-shape distributed, which is an indication that the estimated posterior is broader than the true posteriors (i.e., underconfident). Therefore, our constraint should be considered conservative. In Appendix E we also show the

validation tests for a larger smoothing scale of $R = 10\ h^{-1}$ Mpc and a smaller cutoff of $k_{\text{max}} = 0.25$ Mpc$^{-1}h$.

### *2. Robustness test: Mock challenge*

Next, we perform the SIMBIG "mock challenge" using the test simulations TEST0, TEST1, and TEST2 described in Sec. IV C. Among the three test simulations, TEST0 is the closest to the training dataset as they share the same $N$-body simulation, halo finder, and HOD model parameters. In contrast, TEST1 used a different halo finder (FoF, which does not resolve substructure as well as the ROCKSTAR). It includes only standard HOD parameters and assumes that only the halo mass governs the galaxy–halo connection. Lastly, TEST2 is the most different one from the training dataset; It is built from a different $N$-body simulation (ABACUS) with halos identified with a different halo finder (COMPASO) and populated with galaxies using a standard 5-parameter HOD model.

To test the robustness of our trained model to changes in the forward model, we compute the mean and the standard
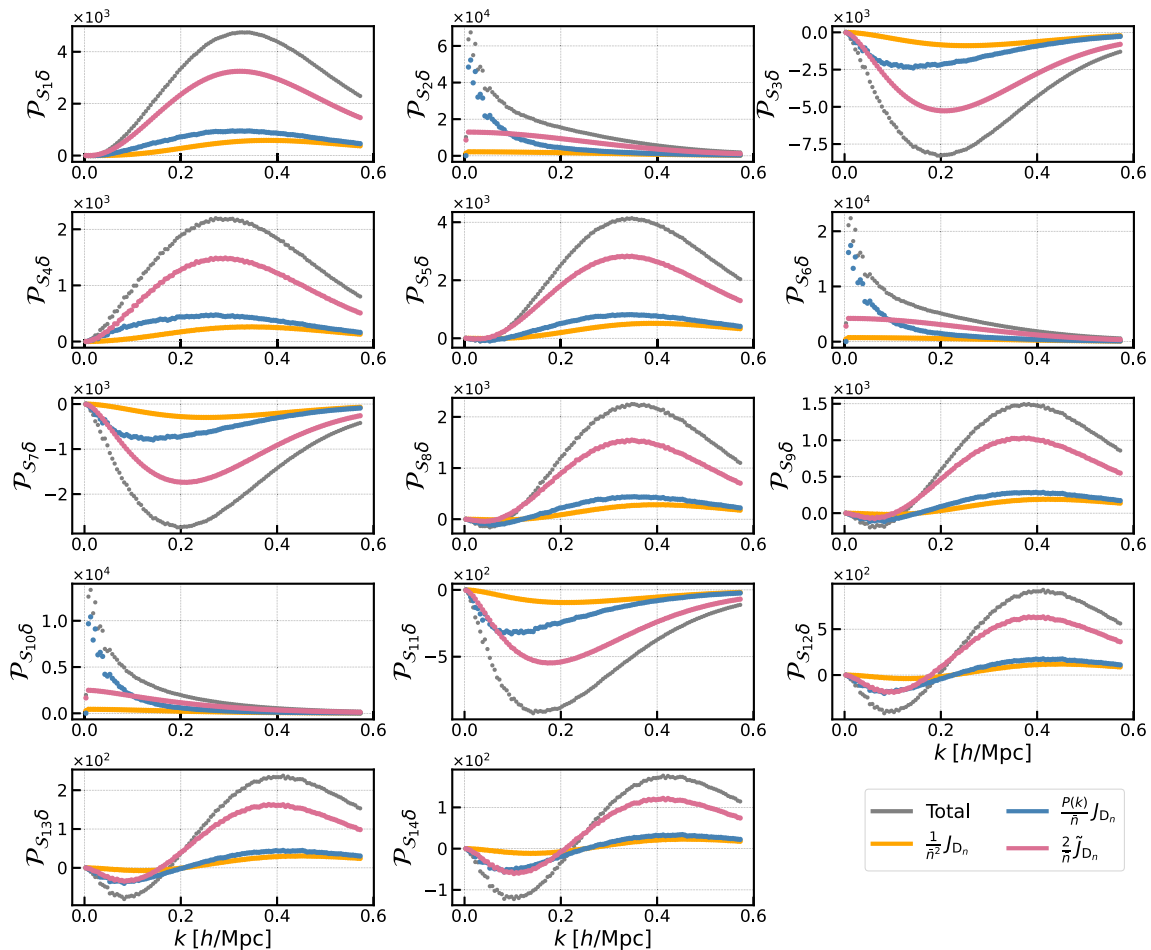
FIG. 3. The three contributions to the Poisson shot noise of individual galaxy skew spectra given in Eq. (3). The smoothing scale is set to $R = 5 \ h^{-1}$ Mpc. Grey curves denote the total shot noise contribution.

deviation of the parameter posteriors for each test simulation and plot the distribution of the difference between the posterior mean $(\overline{\theta_i})$ and the fiducial value, normalized by the standard deviation $\sigma_{\theta_i}$ for each of the parameters and show the three test sets in Fig. 5. If the analysis is robust to variation of the forward model, we expect to obtain a

statistical consistency for the distributions across the three test sets, while the posterior mean is not expected to be an unbiased estimate of the true parameter. Given the good consistency for the five cosmological parameters for the three test sets, we conclude that our pipeline is robust and can be applied to observational data.
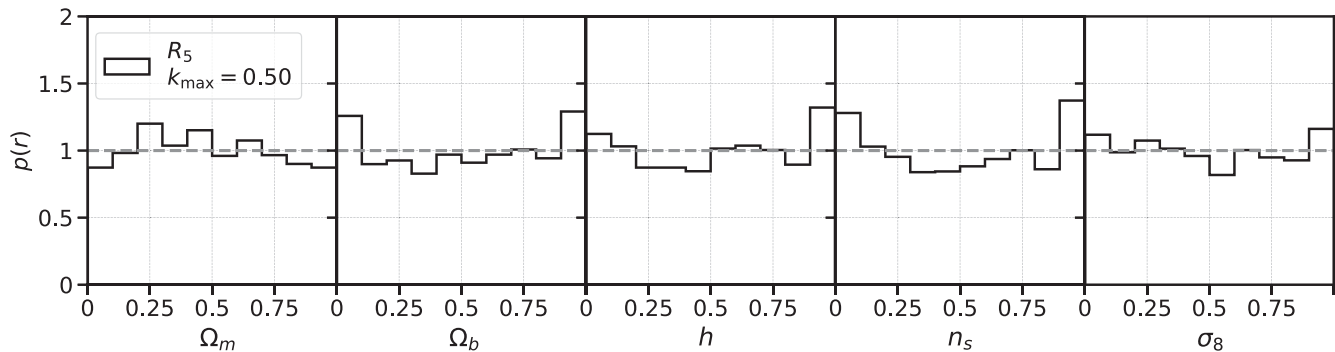


FIG. 4. Accuracy test: shown are the rank statistics of the validation dataset for the configuration $R = 5 \ h^{-1}$ Mpc and $k_{\max} = 0.5$ Mpc$^{-1}h$. For most of the parameters, the rank statistics are uniformly distributed. The slight ∩-shape in the $\Omega_{\mathrm{m}}$ plot is an indication of an under-confident (conservative) error bar.
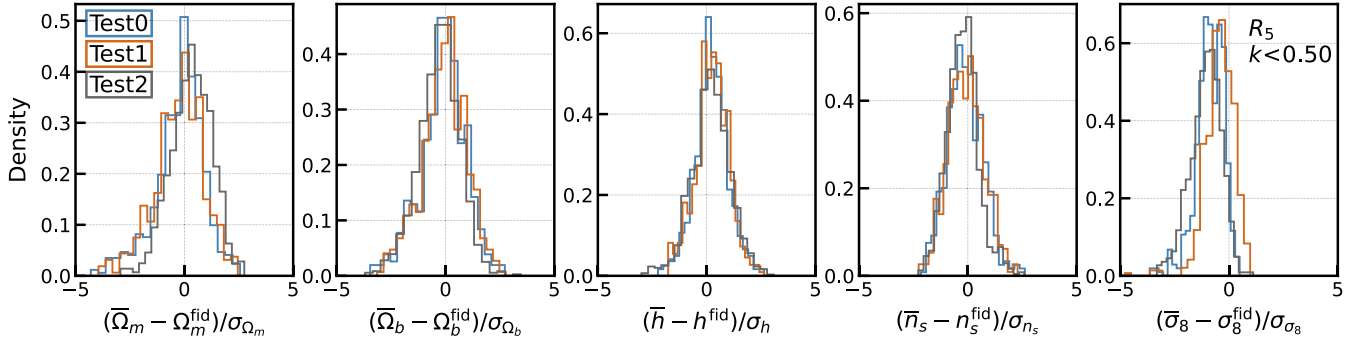
FIG. 5.　Robustness test: the distribution of the difference between the inferred posterior mean $\overline{\theta}_i$ and the fiducial value $\theta_i^{\text{fid}}$ normalized by the standard deviation of the posterior $\sigma_{\theta_i}$ for each cosmological parameter on the three sets of test simulations.

## C. Inferred posterior on BOSS CMASS-SGC data

Having validated the trained models on various test simulations, we now move on to applying $q_\phi$ to observational data. To illustrate the importance of extracting non-Gaussian information of the LSS beyond what is encoded on the power spectrum, we first compare the cosmological constraints from the skew spectra with those from previous SIMBIG analysis of the power spectrum multipoles [64]. Next, we compare the results from the skew spectra with those from previous SIMBIG analysis of the bispectrum monopole [31] to compare their information content. Lastly, we discuss the dependence of the inferred constraints on several of the analysis choices including the smoothing scale, the minimum and maximum scale cuts, the shot noise subtraction. Table I of Appendix C presents the $1\sigma$ uncertainties on cosmological parameters for each of these studies.

### 1. Comparison with power spectrum multipoles

For this comparison, we select an optimal scale in both the maximum wave number $k_{\text{max}} = 0.5$ Mpc$^{-1}h$ and a smoothing scale of $R = 5$ $h^{-1}$ Mpc. This scale choice should, in principle, allow us to extract information from the galaxy field maximally. We notice that, the $k_{\text{max}}$ cut for the skew spectra and the power spectrum (or the bispectrum) is not strictly speaking equivalent. As we will allude below, on the one hand, smoothing of the observed field washes out small-scale information and on the other hand, the convolution of two density fields, mixes small- and large-scale information.

Figure 6 shows the inferred posteriors on cosmological parameters from the skew spectra (blue) and the power spectrum multipoles (gray). The full posterior distributions including the HOD parameters are shown in Fig. 8 of Appendix C. Without applying the BBN prior, we find the posterior mean and the 68% CF to be $\Omega_{\text{m}} = 0.288^{+0.024}_{-0.034}$, $\Omega_{\text{b}} = 0.043^{+0.005}_{-0.007}$, $h = 0.759^{+0.104}_{-0.050}$, $n_{\text{s}} = 0.918^{+0.041}_{-0.090}$, and $\sigma_8 = 0.778^{+0.066}_{-0.093}$. Compared to the power spectrum, the skew spectra improve the constraints in terms of the 68% CL by 34%, 35%, and 18% in matter density $\Omega_{\text{m}}$, baryon density $\Omega_{\text{b}}$, and $h$, respectively. However, the

constraint in $\sigma_8$ are weaker than that from the power spectrum. As mentioned earlier, the larger uncertainty on $\sigma_8$ can be associated with two facts. First, smoothing the observed field in the skew spectra washes out some information on small scales, which plays a role in constraints on $\sigma_8$. Second, the convolutional nature of skew spectra kernels causes a redistribution of information across various scales. Consequently, the influence of $\sigma_8$ is not exclusively related to small scales. Imposing the BBN prior on baryon density using importance sampling with $\omega_b = \Omega_b/h^2 = 0.02268 \pm 0.00038$, the skew spectra improve the constraints on matter density $\Omega_{\text{m}}$, baryon density $\Omega_{\text{b}}$, and $h$, by 34%, 49%, and 38%, respectively.

Despite the reported improvements over the SIMBIG power spectrum, it is crucial to recognize two points; first,
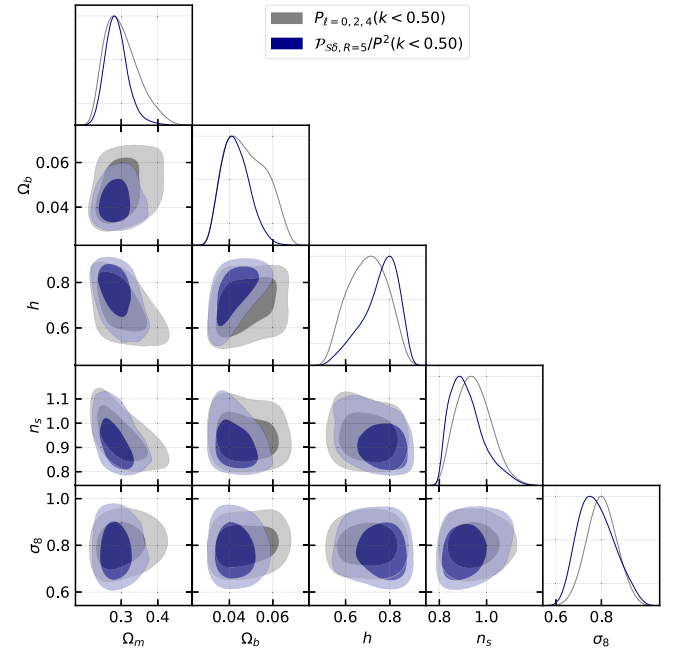


FIG. 6.　Posterior distribution of cosmological parameters from the power spectrum multipoles (gray) and the skew spectra (blue) setting $k_{\text{max}} = 0.5$ $h^{-1}$ Mpc for both. The smoothing scale for the skew spectra is set to $R = 5$ $h^{-1}$ Mpc.

the inferred posteriors from the power spectrum on most parameters are underconfident, thus, the reported relative improvement by the skew spectra can potentially change with a better-trained model for the SIMBIG power spectrum. Second, the skew spectra utilized in this study are not fully optimized for constraining cosmological parameters. Instead, they represent maximum-likelihood estimators of amplitudelike parameters. Consequently, the reported constraints could potentially be enhanced by strategically weighting the quadratic field using either the Fisher information matrix [109] or the score function [110]. The anticipated improvement arises from up-weighting the Fourier modes that contribute the most to the constraints on a given cosmological parameter. We defer further investigation into this possibility to future work.

### 2. Comparison with bispectrum monopole

Given that the construction of skew spectra inevitably mixes scales and washes out the information below the smoothing scale, we compare the skew spectra and the bispectrum using a more conservative choice of scales: applying a smoothing scale of $R = 10 \ h^{-1}$ Mpc and a lower cutoff scale of $k_{\max} = 0.25$ Mpc$^{-1}h$. The comparison of the constraints from the three statistics at higher $k_{\max}$ is presented in Appendix E.

Figure 7 shows the constraints from the skew spectra (blue), the power spectrum multipoles (gray) [64], and the
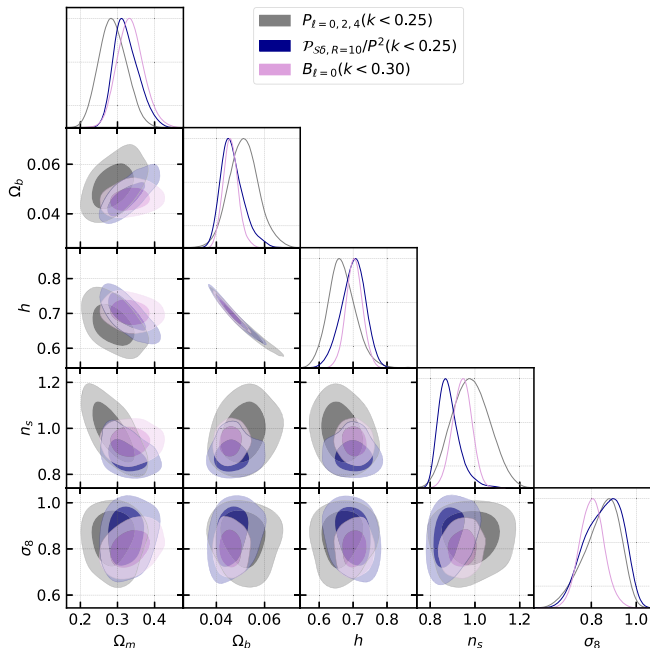


FIG. 7.   Posterior distribution of cosmological parameters from the power spectrum (gray), the skew spectra (blue), and the bispectrum monopole (red) imposing the BBN prior, and setting $k_{\max} = 0.25 \ h^{-1}$ Mpc for the first two and $k_{\max} = 0.3$Mpc$^{-1}h$ for the third. The smoothing scale is set to $R = 10 \ h^{-1}$ Mpc for the skew spectra.

bispectrum monopole (magenta) [31]. Here the BBN prior is imposed for all analyses. The constraints from the bispectrum monopole are obtained using $k_{\max} = 0.3$ Mpc$^{-1}h$, while those from the power spectrum assume $k_{\max} = 0.25$ Mpc$^{-1}h$. The three summary statistics provide consistent cosmological constraints. The width of the posteriors from $\mathcal{P}_{\mathcal{S}_n}$ and $B_0$, except for $\sigma_8$, are nearly identical. Again, the reduced constraining ability of the skew spectra on $\sigma_8$ can be attributed to its inherent smoothing and convolutional structure, as previously discussed. Without BBN prior, the constraints from the bispectrum monopole on most of the parameters are tighter. We note that since for $R = 10 \ h^{-1}$ Mpc the inferred posteriors of all parameters from skew spectra show under-confidence (see Fig. 9), the reported comparison with the bispectrum monopole may be altered with future improvements of SIMBIG framework.

Although the above configuration offers a more meaningful comparison between the skew spectra and the bispectrum, we also compare the results of the two statistics for a smaller value of the smoothing scale of $R = 5 \ h^{-1}$ Mpc and considering $k_{\max} = \{0.25, 0.5\}$ Mpc$^{-1}h$. With the reduced smoothing scale, the constraints by the skew spectra is weaker when contrasted with the bispectrum. Additionally, we observed a $\sim 1\sigma$ discrepancy in $\Omega_b$ between the skew spectra and the bispectrum monopole under these scale cuts. A similar discrepancy in $h$ emerges when applying the BBN prior. The fact that skew spectra and bispectrum results align only with conservative scale cuts suggests model misspecification may be affecting each statistic differently. This could be due to the skew spectra's sensitivity to redshift-space distortions, which are not entirely accounted for in the bispectrum monopole. In particular, the HOD model's inaccuracies in characterizing satellite galaxy kinematics might affect the results from the skew spectra. A thorough understanding of this model misspecification requires more detailed future research.

### 3. Constraints on HOD parameters

In addition to the cosmological parameters, we also show the full posterior distribution including the HOD parameters with BBN priors, in Figs. 8 of Appendix C, respectively. Our analysis reveals no significant enhancement in HOD constraints using the skew spectra compared to the power spectrum multipoles and bispectrum monopole. We found that the posterior distribution of $\sigma_{\log M}$ touches the prior range from the right, indicating that we need to extend the HOD simulations to a wider prior [111]. However, this expansion is unlikely to substantially affect the posterior of other parameters, considering the weak degeneracies between $\sigma_{\log M}$ and other parameters. Additionally, the degeneracy directions in HOD parameters derived from the skew spectra are similar to those from the power spectrum multipoles, as predicted by [13]. This similarity is likely attributable to both estimators accessing redshift-

space information. It is worth noting that among cosmological parameters, $\sigma_8$ is the one most degenerate with HOD parameters. Given the limited constraints on HOD parameters from skew spectra, this degeneracy can be partially responsible for the rather weak constraints on $\sigma_8$. More discussion in the implication of posterior distributions of HODs can be found in Appendix C.

### 4. Assessment of alternative analysis configurations

To gain further insight into how alternative analysis choices impact the inferred posteriors, we trained additional sets of normalizing flow models for five additional configurations. For each case, we trained 2,000–3,000 models and considered the ensemble average of the top 10 models with lowest validation loss. All analyses are performed setting the smoothing scale of $R = 5 \ h^{-1}$ Mpc and including scales up to $k_{\max} = 0.5$ Mpc$^{-1}h$. We summarize our observations here and refer the interested reader to Appendix E for further details and visualizations of the results. We also present the $1\sigma$ constraints for these analyses in Table II

(i)   Removal of the outliers in the training data: we remove training galaxy catalogs with extreme number densities and skew spectra amplitudes before conducting SBI. Removing these outliers changes the constraining power of the inferred posteriors by $<15\%$. This consistency holds both with and without the BBN prior (see Appendix B for further details on different outlier removal schemes).

(ii)  Subtraction of the Poisson shot noise: in the standard analysis based on an explicit form of the likelihood, the Poisson shot noise is subtracted in the estimator of a given summary statistics to reduce the covariance. With our SBI approach, the inferred posteriors with or without shot noise subtraction in the skew spectra estimator yields similar constraints (see Fig. 11). This suggests that the SBI method effectively distinguishes the shot noise and clustering components of the skew spectra. Further validation of this interpretation for galaxy samples of varying number densities is left for a future work.

(iii) Choice of the large-scale cutoff, $k_{\min}$: given that the systematic weights for BOSS data are not well tested on large scales, we investigate the impact of excluding the scales close to the survey's fundamental modes $k_f$ from the analysis. We find that setting $k_{\min} = 0.01$ [Mpc$^{-1}h$], as done in the standard BOSS analyses (e.g., [112]), has only a percent level impact on the overall constraints on $\Lambda$CDM parameters. This is somewhat expected since imposing this cutoff only excludes a single $k$-bin given our binning scheme.

(iv)  Choice of the small-scale cutoff, $k_{\max}$: to quantify the dependence of the results on the

small-scale cutoff, we experimented with $k_{\max} = \{0.15, 0.25, 0.5\}$ [Mpc$^{-1}h$]. The most significant degradation, especially in $\Omega_m$ (30%), was observed with a cutoff $k < 0.15$ Mpc$^{-1}h$. Even with such an aggressive cutoff, the overall constraints remained comparable to those from the power spectrum with $k_{\max} = 0.5$ Mpc$^{-1}h$. This conclusion holds with or without BBN priors (see Fig. 11).

(v)   Choice of smoothing scale: we compared the results with the posterior inferred from skew spectra with smoothing scale $R = 10 \ h^{-1}$ Mpc. Larger smoothing scales leads to expected degradation in constraints, most notably a 40% decrease in $\Omega_b$ precision. Other parameters were less affected (see Fig. 12).

## VI. DISCUSSIONS AND CONCLUSIONS

This paper presents the first cosmological constraints from analysing the full set of galaxy skew spectra obtained from the BOSS DR12 CMASS-SGC dataset. This work is part of a series of articles by the SIMBIG collaboration, in which several summary statistics of CMASS galaxy sample were analyzed [66], including the power spectrum [64,65], the bispectrum [31], the wavelet scattering transform [32], and the field-level analysis using the CNN [33].

In order to apply the skew spectra to observational data from spectroscopic galaxy surveys, we developed a new FFT-based estimators for the skew spectra, akin to power spectrum and bispectrum FKP estimators. This estimator integrates essential components such as the survey mask, systematic weights, and a subtraction of the Poisson shot noise.

We use the SIMBIG framework to perform a simulation-based inference using normalizing flows for neural density estimation to obtain cosmological constraints. The SIMBIG pipeline incorporates accurate forward modeling of the observed galaxy distribution for the BOSS CMASS-SGC sample, enabling SBI to infer posteriors for cosmological and HOD parameters from summary statistics of choice. Before applying the SIMBIG framework to CMASS data, as in previous SIMBIG analyses, we perform two validation tests to ensure that the inferred posteriors are robust and unbiased. Our validation tests, which include simulation-based calibration and the SIMBIG mock challenge, confirms that the inferred posterior is robust and unbiased for both parameters.

Setting $k_{\max} = 0.5$ Mpc$^{-1}h$ and $R = 5 \ h^{-1}$ Mpc, and without introducing any informative priors, the skew spectra enhance constraints on $\Omega_m$, $\Omega_b$, and $h$ by 34%, 35%, and 18%, respectively, over constraints from the power spectrum multipoles. Imposing the BBN prior, we improve the constraint on $h$ by a factor of 2.3, enabling us to constrain $h$ at 3% level with only 10% of the full BOSS survey. These relative improvement can potentially be modified given the current underconfident posterior

distribution inferred from the SIMBIG power spectrum. The inferred value of $n_s$ tends to be lower than that of Planck2018 [113], while the mean of the posterior of $h$ tends to be higher. We attribute these tendencies to either statistical fluctuation or potential model misspecification on small scales. A more detailed investigation of this matter is reserved for future work.

The skew spectra did not yield a reduction in uncertainties on $\sigma_8$ and HOD parameters compared to the power spectrum. This lack of improvement can be attributed to the smoothing and convolutional nature of the skew spectra. The smoothing tends to diminish small-scale modes, and the convolution mixes scales, transferring information from small to larger scales. Consequently, the distinctive information typically associated with $\sigma_8$ and HOD parameters, often tied to small scales, becomes dispersed across various scales, complicating interpretation. To validate this intuition, one could explore an alternative parameterization of the galaxy-dark matter relation, such as using a perturbative forward model [62]. Additionally, the skew spectra's limited constraining power may be influenced by the low mean number density and volume of the CMASS-SGC galaxy sample. Further investigation into these issues is deferred to future research.

Our results confirm the expectation that on the semi-linear scales (setting $R = 10 \, h^{-1} \, \mathrm{Mpc}$ and imposing $k_{\mathrm{max}} = 0.25 \, \mathrm{Mpc}^{-1} h$), the skew spectra capture the majority of the information of the bispectrum. The SIMBIG analyses of the bispectrum monopole and the skew spectra yield comparable constraints on all parameters, except for $\sigma_8$, which is better constrained by the bispectrum monopole. Extending the analysis to the smaller scales (setting $R = 5 \, h^{-1} \, \mathrm{Mpc}$ and imposing $k_{\mathrm{max}} = 0.5 \, \mathrm{Mpc}^{-1} h$), we observe the inferred constraints from the skew spectra are weaker relative to the bispectrum. In addition, we report a $\sim 1\sigma$ discrepancy in the mean of the posterior of $\Omega_b$ and $h$ from the bispectrum monopole and skew spectra. This discrepancy can hint at a potential model misspecification on small scales, possibly due to inaccuracies in HOD descriptions. We leave a more in-depth investigation of this issue to future works.

Beyond the main result, we performed a series of tests to explore whether our cosmological constraints are impacted by the removal of outliers from training data, the subtraction of the Poisson shot noise, the choice of large- and small-scale cutoff scales, and the choice of the smoothing scale. We found that removal of the outlier can affect the constraints mildly, less than 15%. Regarding the shot noise subtraction, we observe a marginal impact on the posterior distributions. We found that applying a highly conservative scale cut ($k_{\mathrm{max}} = 0.15 \, \mathrm{Mpc}^{-1} \, h$) and using a larger smoothing scale lead to degradation of the constraints on $\Omega_m$ and $\Omega_b$, respectively, leaving the uncertainties on other parameters largely unaffected.

This work can be extended in several directions. First, there is room for improvement in the optimality of the skew spectra. The current kernels defining the quadratic fields were designed for optimal constraint on the amplitude of the primordial power spectrum and bispectrum, the growth rate of structure, and galaxy biases. However, they may not fully capture all the shape information. A future extension can refine the kernels to be "shape-sensitive" and optimally capable of capturing cosmological information. Additionally, the 14 galaxy skew spectra can be broadly classified into three families (see Fig. 2). Consequently, the size of the data vector may be potentially reduced by constructing optimal combinations.

Given that the skew spectra are optimally designed for constraining PNG, another natural direction for future work is to analyze BOSS data to constrain the Gaussianity of initial conditions using the full set of galaxy skew spectra. Within SIMBIG framework, this analysis requires incorporating non-Gaussian initial conditions into the forward model and investigating whether the HOD model is affected by PNG.

Combining the power spectrum and the skew spectra should improve the constraints from the two statistics individually. Hence, an investigation of an optimal strategy for effectively combining multiple summary statistics is left for future work.

Lastly, considering the observed limited sensitivity of the skew spectra to HOD parameters in our analysis, an interesting avenue for exploration would be to investigate the constraining power of skew spectra when performing SBI using an alternative forward model, e.g., based on perturbation theory [62].

While finalizing this draft, a related work, [114], appeared on arXiv, in which an application of a subset of the skew spectra considered here (neglecting the line-of-sight-dependent quadratic kernels) to BOSS data was presented. In contrast to this work, in which we focused on constraining $\Lambda$CDM using SBI, [114] employed a subset of three skew spectra to constrain primordial non-Gaussianity of equilateral and orthogonal shape using perturbation theory and assuming a Gaussian likelihood.

## APPENDIX A: EXPLICIT FORMS OF SKEW SPECTRA KERNELS IN REDSHIFT SPACE

A set of 14 skew spectra, capture the clustering information of the tree-level galaxy bispectrum in redshift space. Each of the skew spectra corresponds to the maximum likelihood estimator of a specific combination of bias parameters and the growth rate $f$. The explicit forms of the quadratic fields and the parameter combination they are most sensitive to are given below [12],

$$b_1^3: \mathcal{S}_1 = F_2[\delta, \delta], \tag{A1}$$

$$b_1^2 b_2: \mathcal{S}_2 = \delta^2, \tag{A2}$$

$$b_1^2 b_{\mathcal{G}_2}: \mathcal{S}_3 = S^2[\delta, \delta], \tag{A3}$$

$$b_1^3 f: \mathcal{S}_4 = \hat{z}_i \hat{z}_j \partial_i \left( \delta \frac{\partial_j}{\nabla^2} \delta \right), \tag{A4}$$

$$b_1^2 f: \mathcal{S}_5 = 2F_2[\delta^\parallel, \delta] + G_2^\parallel[\delta, \delta], \tag{A5}$$

$$b_1 b_2 f: \mathcal{S}_6 = \delta \delta^\parallel, \tag{A6}$$

$$b_1 b_{\mathcal{G}_2} f: \mathcal{S}_7 = \mathcal{S}^2[\delta, \delta^\parallel], \tag{A7}$$

$$b_1^2 f^2: \mathcal{S}_8 = \hat{z}_i \hat{z}_j \partial_i \left( \delta \frac{\partial_j}{\nabla^2} \delta^\parallel + 2\delta^\parallel \frac{\partial_j}{\nabla^2} \delta \right), \tag{A8}$$

$$b_1 f^2: \mathcal{S}_9 = F_2[\delta^\parallel, \delta^\parallel] + 2G_2^\parallel[\delta^\parallel, \delta], \tag{A9}$$

$$b_2 f^2: \mathcal{S}_{10} = (\delta^\parallel)^2, \tag{A10}$$

$$b_{\mathcal{G}_2} f^2: \mathcal{S}_{11} = S^2(\delta^\parallel, \delta^\parallel), \tag{A11}$$

$$b_1 f^3: \mathcal{S}_{12} = \hat{z}_i \hat{z}_j \partial_i \left( \delta^\parallel \frac{\partial_j}{\nabla^2} \delta + 2\delta^\parallel \frac{\partial_j}{\nabla^2} \delta^\parallel \right), \tag{A12}$$

$$f^3: \mathcal{S}_{13} = G_2^\parallel[\delta^\parallel, \delta^\parallel], \tag{A13}$$

$$f^4: \mathcal{S}_{14} = \hat{z}_i \hat{z}_j \partial_i \left( \delta^{\parallel\parallel} \frac{\partial_j}{\nabla^2} \delta^\parallel \right). \tag{A14}$$

The $F_2$ and $G_2$ functions are the standard perturbation theory kernels for matter density and velocity (see, e.g., Eq. (B.13) in [115]), while $S^2$ is the Fourier transform of the Galileon operator, capturing the effect of the tidal field,

$$S^2(\mathbf{k}_1, \mathbf{k}_2) \equiv \left( \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \right)^2 - 1. \tag{A15}$$

We defined the redshift-space operators as

$$\mathcal{O}^\parallel = \hat{z}_i \hat{z}_j \frac{\partial_i \partial_j}{\nabla^2} \mathcal{O}, \tag{A16}$$

$$\mathcal{O}^{\parallel\parallel} = \hat{z}_i \hat{z}_j \hat{z}_m \hat{z}_n \frac{\partial_i \partial_j \partial_m \partial_n}{\nabla^4} \mathcal{O}, \tag{A17}$$

and the operators $\mathcal{O}[a, b]$ that act on arbitrary fields $a$ and $b$ as

$$\mathcal{O}[a, b](\mathbf{k}) \equiv \int_{\mathbf{q}} \frac{1}{2} [a(\mathbf{q})b(\mathbf{k} - \mathbf{q}) + b(\mathbf{q})a(\mathbf{k} - \mathbf{q})]$$
$$\times \mathcal{O}(\mathbf{q}, \mathbf{k} - \mathbf{q}). \tag{A18}$$

## APPENDIX B: PREPROCESSING OF THE MEASUREMENTS TO REMOVE OUTLIERS

The SIMBIG forward-modeled galaxy mocks display wide variations in the mean number density and the clustering amplitude due to the broad HOD parameter priors. To enhance the stability of our training process and ensure that all features meaningfully contribute to the neural networks, we preprocess the training dataset, by removing "outlier" mock catalogs. Furthermore, we normalize the measured skew spectra by dividing them by square of the power spectrum to obtain features of order unity. We compared several algorithms for removing the outliers, which we describe below.

We remove the outliers based on the amplitude of the skew spectra, and compare the performance of three algorithms; (1) elimination based on the z-score: this is a simple prescription to remove the data points that lie on the tails of a (Gaussian) distribution. The data points with a z-score (the number of standard deviations from the mean) greater than a threshold are declared to be outliers. (2) Local outlier factor [LOF [116]] is a machine-learning-based anomaly detection algorithm. It hinges on the idea that anomalies will have a significantly lower density of neighbors than on average, indicating that they are relatively more isolated. It assesses the local density of data points by comparing the density of a data point to its neighbors. (3) Isolation forest [IF [117]] is also machine-learning based; the IsoForest isolates anomalies by randomly partitioning the data and creating isolation trees (a decision tree for anomaly isolation). It exploits the fact that anomalies are fewer and more isolated than normal instances, making them easier to separate within the isolation trees.

In the main analyses presented in this paper, we apply the local outlier algorithms using its implementation in SKLEARN package [118]. We also tested the impact of not removing the outliers and found that it can change the constraints on the 5 cosmological parameters less than 15%. The results are summarized in Table II.

TABLE II. Table summarizes the posteriors of $\Lambda$CDM cosmological parameters inferred from the skew spectra, and the power spectrum. We present the median and 68% uncertainties of the parameters.

| No priors | Scale cuts | $\Omega_m$ | $\Omega_b$ | $h$ | $n_s$ | $\sigma_8$ |
|---|---|---|---|---|---|---|
| $\mathcal{P}_{R=5}$ | $k \in (k_f, 0.5)$ | $0.288^{+0.024}_{-0.034}$ | $0.043^{+0.005}_{-0.007}$ | $0.756^{+0.104}_{-0.050}$ | $0.918^{+0.041}_{-0.090}$ | $0.778^{+0.066}_{-0.093}$ |
| | $k \in (0.01, 0.5)$ | $0.297^{+0.027}_{-0.040}$ | $0.042^{+0.005}_{-0.009}$ | $0.735^{+0.107}_{-0.072}$ | $0.924^{+0.046}_{-0.087}$ | $0.794^{+0.072}_{-0.098}$ |
| | $k \in (0.01, 0.25)$ | $0.292^{+0.024}_{-0.037}$ | $0.044^{+0.005}_{-0.008}$ | $0.761^{+0.102}_{-0.051}$ | $0.936^{+0.050}_{-0.085}$ | $0.798^{+0.078}_{-0.086}$ |
| | $k \in (0.01, 0.15)$ | $0.322^{+0.042}_{-0.055}$ | $0.043^{+0.005}_{-0.009}$ | $0.729^{+0.118}_{-0.076}$ | $0.930^{+0.047}_{-0.091}$ | $0.837^{+0.091}_{-0.066}$ |
| $\mathcal{P}_{R=10}$ | $k \in (0.01, 0.5)$ | $0.309^{+0.030}_{-0.037}$ | $0.047^{+0.007}_{-0.011}$ | $0.716^{+0.090}_{-0.079}$ | $0.902^{+0.032}_{-0.083}$ | $0.848^{+0.107}_{-0.068}$ |
| | $k \in (0.01, 0.25)$ | $0.308^{+0.029}_{-0.036}$ | $0.052^{+0.007}_{-0.007}$ | $0.753^{+0.080}_{-0.063}$ | $0.909^{+0.043}_{-0.079}$ | $0.846^{+0.081}_{-0.079}$ |
| $\mathcal{P}_{R=5}$, shot noise incl. | $k \in (0.01, 0.5)$ | $0.275^{+0.026}_{-0.036}$ | $0.042^{+0.005}_{-0.008}$ | $0.738^{+0.101}_{-0.068}$ | $0.953^{+0.056}_{-0.111}$ | $0.778^{+0.065}_{-0.087}$ |
| $\mathcal{P}_{R=5}$, no outlier removal | $k \in (0.01, 0.5)$ | $0.288^{+0.027}_{-0.034}$ | $0.043^{+0.005}_{-0.008}$ | $0.760^{+0.103}_{-0.053}$ | $0.924^{+0.041}_{-0.089}$ | $0.798^{+0.080}_{-0.097}$ |
| $P_{\ell=0,2}$ [64] | $k \in (k_f, 0.5)$ | $0.302^{+0.032}_{-0.057}$ | $0.048^{+0.008}_{-0.012}$ | $0.700^{+0.093}_{-0.095}$ | $0.947^{+0.063}_{-0.082}$ | $0.802^{+0.067}_{-0.066}$ |
| | $k \in (k_f, 0.25)$ | $0.283^{+0.034}_{-0.041}$ | $0.052^{+0.008}_{-0.008}$ | $0.695^{+0.105}_{-0.105}$ | $0.981^{+0.082}_{-0.094}$ | $0.845^{+0.098}_{-0.059}$ |
| $B_0$ [31] | $k \in (k_f, 0.5)$ | $0.294^{+0.027}_{-0.026}$ | $0.059^{+0.005}_{-0.005}$ | $0.756^{+0.040}_{-0.040}$ | $0.952^{+0.039}_{-0.034}$ | $0.784^{+0.037}_{-0.042}$ |
| | $k \in (k_f, 0.3)$ | $0.336^{+0.034}_{-0.038}$ | $0.045^{+0.004}_{-0.004}$ | $0.696^{+0.041}_{-0.043}$ | $0.944^{+0.043}_{-0.043}$ | $0.802^{+0.051}_{-0.056}$ |
| **BBN prior** | **scale cuts** | $\Omega_m$ | $\Omega_b$ | $h$ | $n_s$ | $\sigma_8$ |
| $\mathcal{P}_{R=5}$ | $k \in (k_f, 0.5)$ | $0.285^{+0.028}_{-0.033}$ | $0.041^{+0.003}_{-0.004}$ | $0.750^{+0.034}_{-0.032}$ | $0.915^{+0.047}_{-0.082}$ | $0.781^{+0.071}_{-0.095}$ |
| | $k \in (0.01, 0.5)$ | $0.291^{+0.028}_{-0.034}$ | $0.041^{+0.003}_{-0.004}$ | $0.744^{+0.032}_{-0.033}$ | $0.920^{+0.051}_{-0.081}$ | $0.798^{+0.082}_{-0.093}$ |
| | $k \in (0.01, 0.25)$ | $0.289^{+0.026}_{-0.034}$ | $0.041^{+0.003}_{-0.004}$ | $0.747^{+0.033}_{-0.035}$ | $0.940^{+0.054}_{-0.083}$ | $0.803^{+0.088}_{-0.085}$ |
| | $k \in (0.01, 0.15)$ | $0.316^{+0.040}_{-0.053}$ | $0.042^{+0.004}_{-0.007}$ | $0.741^{+0.054}_{-0.044}$ | $0.926^{+0.047}_{-0.079}$ | $0.834^{+0.093}_{-0.072}$ |
| $\mathcal{P}_{R=10}$ | $k \in (0.01, 0.5)$ | $0.306^{+0.030}_{-0.038}$ | $0.045^{+0.004}_{-0.006}$ | $0.712^{+0.044}_{-0.042}$ | $0.898^{+0.040}_{-0.070}$ | $0.853^{+0.098}_{-0.072}$ |
| | $k \in (0.01, 0.25)$ | $0.320^{+0.032}_{-0.036}$ | $0.048^{+0.005}_{-0.005}$ | $0.692^{+0.033}_{-0.036}$ | $0.921^{+0.055}_{-0.072}$ | $0.839^{+0.075}_{-0.078}$ |
| $\mathcal{P}_{R=5}$, shot noise incl. | $k \in (0.01, 0.5)$ | $0.273^{+0.025}_{-0.033}$ | $0.041^{+0.003}_{-0.003}$ | $0.745^{+0.028}_{-0.031}$ | $0.941^{+0.060}_{-0.089}$ | $0.779^{+0.070}_{-0.081}$ |
| $\mathcal{P}_{R=5}$, no outlier removal | $k \in (0.01, 0.5)$ | $0.285^{+0.026}_{-0.032}$ | $0.040^{+0.003}_{-0.004}$ | $0.751^{+0.033}_{-0.032}$ | $0.924^{+0.050}_{-0.077}$ | $0.802^{+0.098}_{-0.094}$ |
| $P_{\ell=0,2}$ [64] | $k \in (k_f, 0.5)$ | $0.301^{+0.034}_{-0.057}$ | $0.047^{+0.005}_{-0.008}$ | $0.701^{+0.055}_{-0.052}$ | $0.941^{+0.062}_{-0.078}$ | $0.804^{+0.062}_{-0.070}$ |
| | $k \in (k_f, 0.25)$ | $0.286^{+0.037}_{-0.040}$ | $0.051^{+0.007}_{-0.006}$ | $0.669^{+0.032}_{-0.050}$ | $0.985^{+0.086}_{-0.086}$ | $0.849^{+0.092}_{-0.056}$ |
| $B_0$ [31] | $k \in (k_f, 0.5)$ | $0.291^{+0.032}_{-0.026}$ | $0.050^{+0.002}_{-0.003}$ | $0.676^{+0.020}_{-0.018}$ | $0.949^{+0.046}_{-0.032}$ | $0.768^{+0.029}_{-0.050}$ |
| | $k \in (k_f, 0.3)$ | $0.335^{+0.032}_{-0.038}$ | $0.046^{+0.003}_{-0.003}$ | $0.704^{+0.024}_{-0.024}$ | $0.943^{+0.043}_{-0.041}$ | $0.803^{+0.051}_{-0.052}$ |

## APPENDIX C: FULL POSTERIOR DISTRIBUTION: CONSTRAINTS ON THE HOD PARAMETERS

To illustrate the constraining power of different summary statistics ($P_\ell$, $B_0$, $\mathcal{S}_R$) and the degeneracies between the nuisance and cosmological parameters, in Fig. 8, we show the full posterior distributions on model parameters from analyses of the BOSS CMASS-SGC sample. For all three statistics, the BBN prior is imposed, and the small-scale cutoff scale is set to $k_{max} = 0.5$ Mpc$^{-1}h$. For the skew specra, the smoothing scale of $R = 5\ h^{-1}$ Mpc is imposed.

Generally, except for $\log M_1$ (the characteristic halo mass scale to host satellite galaxies), the skew spectra do not improve the constraints on any of the HOD parameters over the power spectrum. The inferred posterior on $\sigma_{logM}$ (the scatter in the halo mass–galaxy luminosity relation) hits the prior range, indicating that additional mock catalogs sampling a wider prior range are necessary. Similar to $\sigma_8$, the

failure of the skew spectra in improving the constraints on the HOD parameters can potentially be associated smoothing of the input field and the convolutional nature of the skew spectra. The former washes out small-scale modes, while the latter mixes different scales and transfers information from small to larger scales. This interplay potentially disperses the information usually linked to $\sigma_8$ and the HOD parameters, which are typically associated with small scales, thereby complicating their interpretation.

The comparison between the skew spectra and the bispectrum monopole reveals that their constraining power on HOD parameters is quite similar. Notably, certain HOD parameters, like $\log M_{min}$ and $\eta_{cent}$, even show enhancements in constraint accuracy. This observation suggests that the velocity information encapsulated by the RSD could be crucial for more effectively constraining HOD parameters.

Worth noting that among cosmological parameters, $\sigma_8$ shows the strongest degeneracy with several HOD
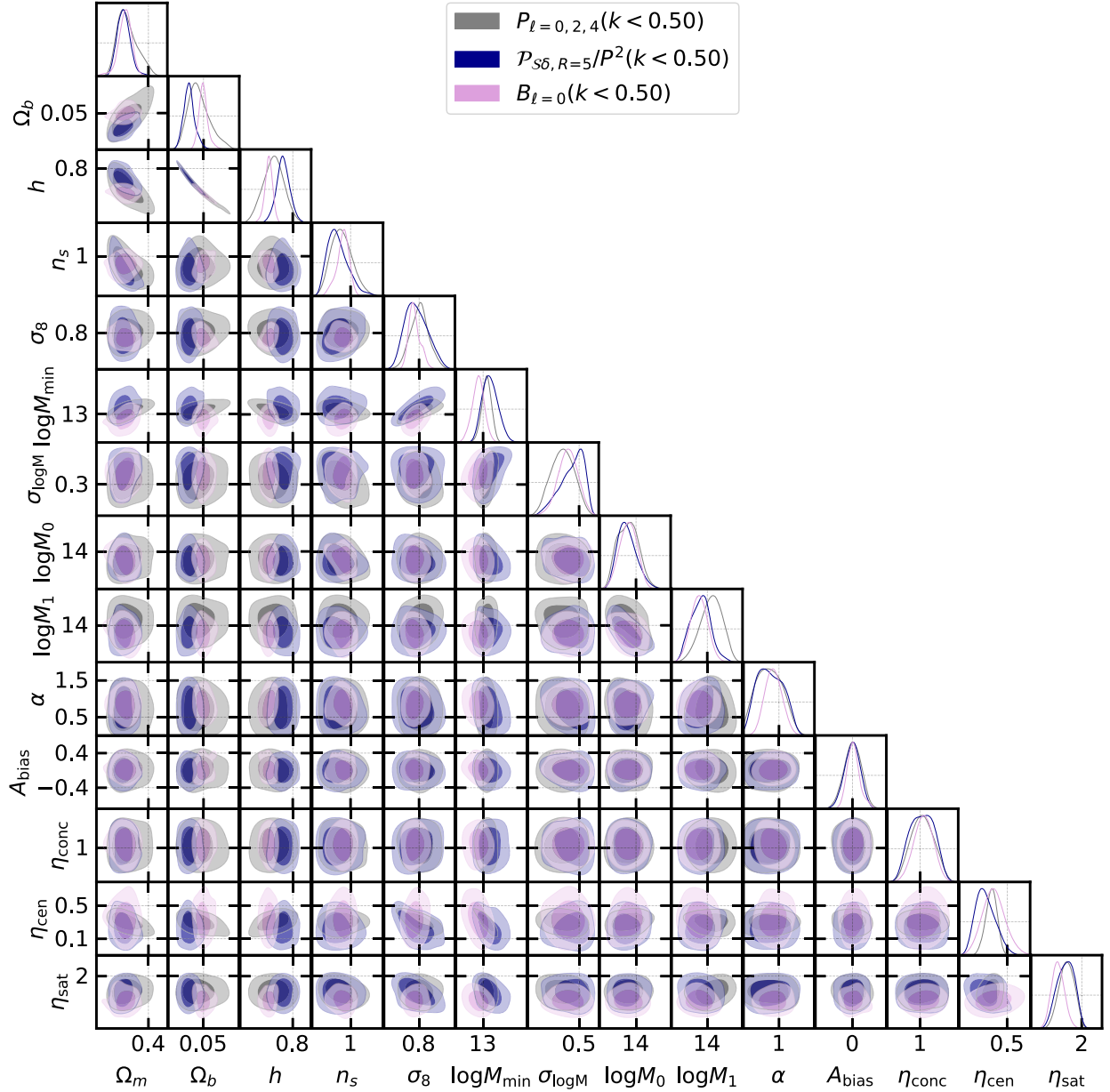
FIG. 8. Posterior distribution of cosmological parameters inferred the skew spectra (blue) and power spectrum (gray) with a cutoff scale $k_{\mathrm{max}} = 0.5 \, h^{-1}$ Mpc including the BBN prior and smoothing scale $R = 5 \, h^{-1}$ Mpc.

parameters, especially, $\log M_{\mathrm{min}}, \eta_{\mathrm{cent}}, \eta_{\mathrm{sat}}$, which are poorly constraints by the skew spectra. These degeneracies play a part in the limited constraining power of the skew spectra on $\sigma_8$. Therefore, analysis of skew spectra using a different model of dark matter–galaxy relation, e.g., the perturbative biasing description, could potentially provide better constraints on $\sigma_8$.

## APPENDIX D: VALIDATION TEST WITH CONSERVATIVE SCALE CUTS

In Sec. V C 2, we presented a comparison of the skew spectra, the power spectrum multipoles, and the bispectrum

monopole for a rather conservative choice of scale cuts ($k_{\mathrm{max}} = 0.25 \, \mathrm{Mpc}^{-1}h$, $R = 10 \, h^{-1}$ Mpc). This section provides the two validation tests to assess the accuracy and robustness of the estimated posteriors from the skew spectra.

Figure 9 shows the distribution of the rank statistics, which exhibits ∩-shaped distributions for all parameters. This behavior indicates that the estimated posteriors are broader than the true posteriors (i.e., underconfident) and implies that the skew spectra constraints should be considered conservative. Figure 10 shows the distributions of the difference between the estimated posterior means and the fiducial values of cosmological parameters across the
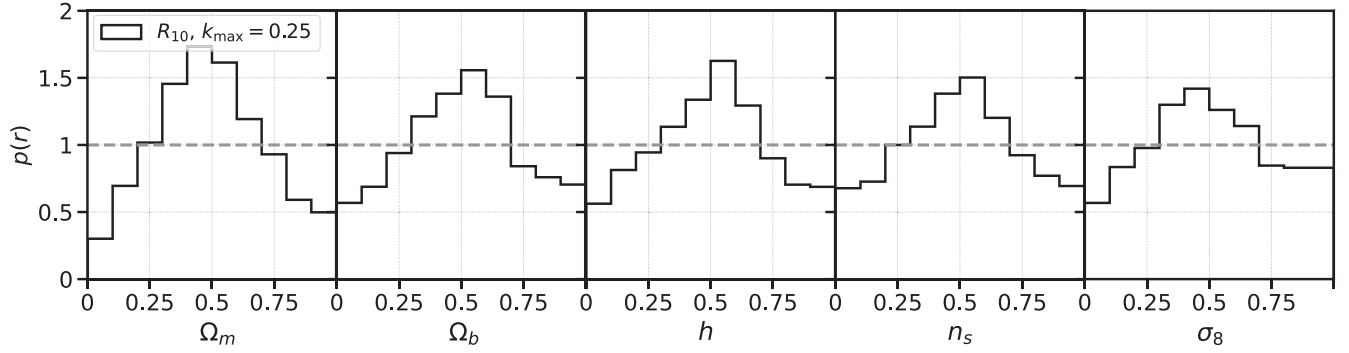
FIG. 9. Rank statistics of the validation data for analysis of the skew spectra with $R = 10\,h^{-1}$ Mpc and $k_{\mathrm{max}} = 0.25$ Mpc$^{-1}\,h$.
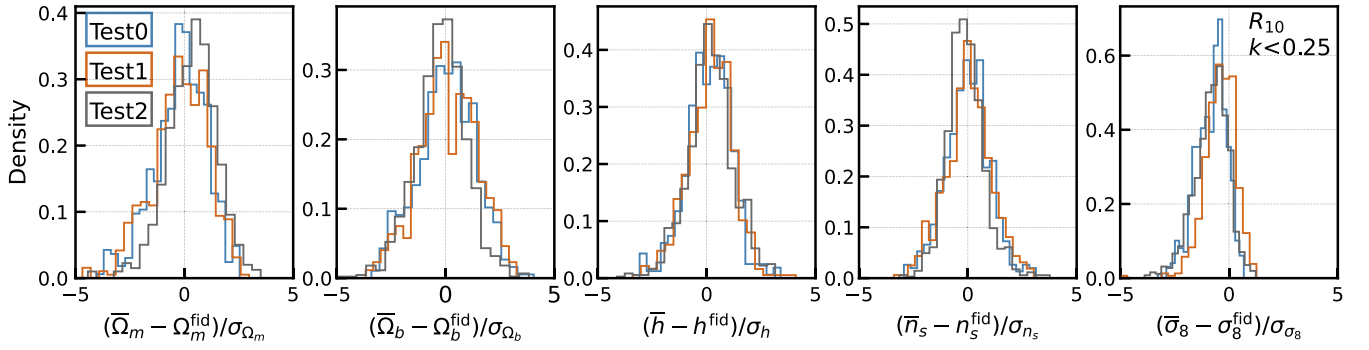


FIG. 10. Distribution of the difference between the inferred posterior mean $\overline{\theta}_i$ and the fiducial value $\theta_i^{\mathrm{fid}}$ normalized by the standard deviation of the posterior $\sigma_{\theta_i}$ for each cosmological parameter on the three sets of test simulations.

three test simulations. The statistical consistency of the distributions across the three test sets demonstrates the robustness of the trained models to details of the forward modeling.

## APPENDIX E: DEPENDENCE OF THE CONSTRAINTS ON ANALYSIS CHOICES

As described in Sec. V C 4, we perform several additional SBI analyses to investigate the dependence on the inferred cosmological constraints on various analysis choices. For each configuration, we retrain 2,000–3,000 neural network models. We report the $1\sigma$ constraints on cosmological parameters for these tests in Table II. We provide further details on these tests here.

First, we study the impact of shot noise. For this test, we opt for $k_{\mathrm{max}} = 0.5$ Mpc$^{-1}h$ and $R = 5$ to enhance the influence at small scales where shot noise effects are non-negligible. The left panel of Fig. 11 shows the comparison with and without shot noise subtraction. We found that the constraints on $\Omega_b$ and $h$ are identical in the two cases, while the widths of the $\sigma_8$ and $n_s$ posteriors are mildly affected. For $\Omega_m$, we observe a slight shift of the peak of the posterior. Despite these marginal differences, we conclude that the constraints are largely unaffected by the subtraction of shot in the estimator. This implies that in

contrast to a conventional likelihood-based inference approach, the SBI approach can effectively isolate the influence of shot noise and extract the cosmological response based on a given summary statistics. However, we note that validating this interpretation requires performing further tests on samples with varying number density and employing different summary statistics.

Second, we investigate the impact of the maximum scale cuts by choosing $k_{\mathrm{max}} = \{0.15, 0.25, 0.5\}$ [Mpc$^{-1}h$]. Here, we set the smoothing scale to $R = 5\,h^{-1}$ Mpc. Due to the convolutional structure of the skew spectra and the smoothing operation, small-scale information is partially transferred to larger scales. As previously shown in the numerical Fisher forecast of [13], while skew spectra can achieve better constraints at relatively large scales compared to the power spectrum or the bispectrum monopole, the constraining power also saturates faster as we push the $k_{\mathrm{max}}$ cuts toward smaller scales. This saturation is also apparent in the fast-dropping power in Fig. 2. As expected, applying an aggressive small-scale cut at $k < 0.15$ Mpc$^{-1}\,h$ degrades the constraints, particularly $\Omega_m$ by 30%. Nevertheless, when imposing the BBN priors, even for this cutoff, the constraints from skew spectra are still comparable to those from the power spectrum multipoles with a scale cut of $k_{\mathrm{max}} = 0.5$ Mpc$^{-1}\,h$.
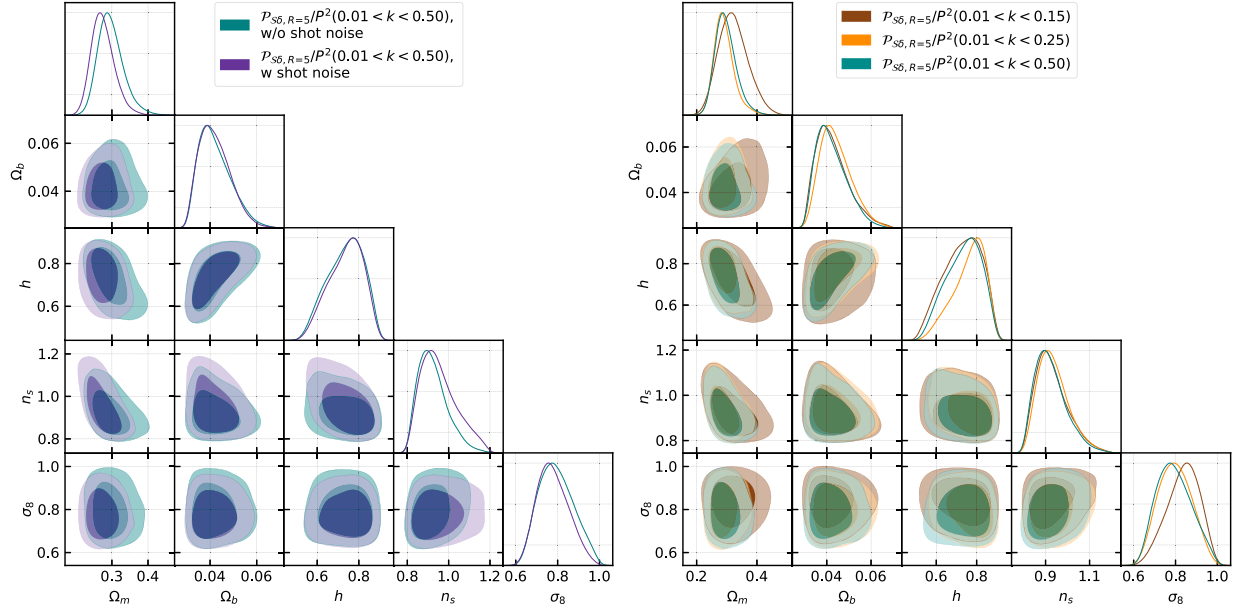
FIG. 11.    Left: posterior distribution of cosmological parameters with (purple) and without (green) subtracting the Poisson shot noise in skew spectra estimators. Right: posterior distribution of cosmological parameters for three choices of small-scale cutoff: $k_{max} = 0.15$ Mpc$^{-1}h$ in red, $k_{max} = 0.25$ Mpc$^{-1}h$ in orange, and $k_{max} = 0.5$ Mpc$^{-1}h$ in green.



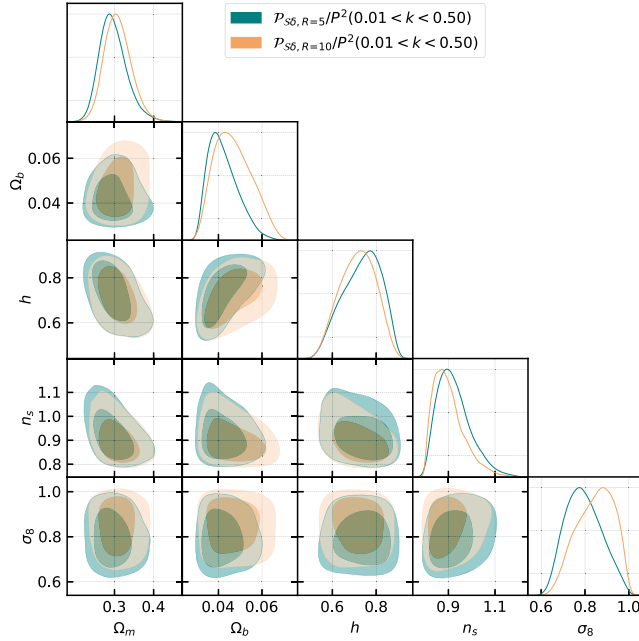FIG. 12.    Posterior distribution of cosmological parameters for two choices of the smoothing scale: $R = 5$ $h^{-1}$ Mpc in green and $R = 10 \, h^{-1}$ Mpc in orange. The small-scale cutoff is set to $k_{max} = 0.5$ Mpc$^{-1}h$ in both cases.

Third, we study the impact of the choice of the smoothing scales by imposing $R = 5 \, h^{-1}$ Mpc and $R = 10 \, h^{-1}$ Mpc. Figure 2 shows that the smoothing scale can shift the featured peak of the skew spectra, which could potentially be sensitive to information at different scales.

Furthermore, a larger smoothing scale washes out the small-scale fluctuations more significantly. We find that, as expected, the constraints degrade when using a larger smoothing scale. The degradation is particularly visible in $\Omega_b$ by $\sim 40\%$. For other parameters, the effect is marginal.

[1] A. Aghamousa *et al.* (DESI Collaboration), arXiv:1611.00036.

[2] R. Laureijs *et al.* (EUCLID Collaboration), arXiv:1110.3193.

[3] Y. Wang, Z. Zhai, A. Alavi, E. Massara, A. Pisani, A. Benson, C. M. Hirata, L. Samushia, D. H. Weinberg, J. Colbert, O. Doré, T. Eifler, C. Heinrich, S. Ho, E. Krause, N. Padmanabhan, D. Spergel, and H. I. Teplitz, Astrophys. J. **928,** 1 (2022).

[4] J. Bock (SPHEREx Science Team), in *American Astronomical Society Meeting Abstracts* (2018), Vol. 231, p. 354.21.

[5] N. Tamura *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **9908,** 99081M (2016).

[6] R. Scoccimarro, Phys. Rev. D **92,** 083532 (2015).

[7] N. S. Sugiyama, S. Saito, F. Beutler, and H.-J. Seo, Mon. Not. R. Astron. Soc. **484,** 364 (2019).

[8] G. D'Amico, Y. Donath, M. Lewandowski, L. Senatore, and P. Zhang, arXiv:2206.08327.

[9] O. H. E. Philcox and M. M. Ivanov, Phys. Rev. D **105,** 043517 (2022).

[10] M. Schmittfull, T. Baldauf, and U. Seljak, Phys. Rev. D **91,** 043530 (2015).

[11] A. Moradinezhad Dizgah, H. Lee, M. Schmittfull, and C. Dvorkin, J. Cosmol. Astropart. Phys. 04 (2020) 011.

[12] M. Schmittfull and A. Moradinezhad Dizgah, J. Cosmol. Astropart. Phys. 03 (2021) 020.

[13] J. Hou, A. Moradinezhad Dizgah, C. Hahn, and E. Massara, J. Cosmol. Astropart. Phys. 03 (2023) 045.

[14] R. K. Sheth, A. J. Connolly, and R. Skibba, arXiv:astro-ph/0511773.

[15] M. White, J. Cosmol. Astropart. Phys. 11 (2016) 057.

[16] E. Massara, F. Villaescusa-Navarro, S. Ho, N. Dalal, and D. N. Spergel, Phys. Rev. Lett. **126,** 011301 (2021).

[17] E. Paillas, Y.-C. Cai, N. Padilla, and A. G. Sánchez, Mon. Not. R. Astron. Soc. **505,** 5731 (2021).

[18] M. Eickenberg *et al.*, arXiv:2204.07646.

[19] G. Valogiannis and C. Dvorkin, Phys. Rev. D **105,** 103534 (2022).

[20] J. Jasche and G. Lavaux, Astron. Astrophys. **625,** A64 (2019).

[21] G. Lavaux, J. Jasche, and F. Leclercq, arXiv:1909.06396.

[22] A. Andrews, J. Jasche, G. Lavaux, and F. Schmidt, Mon. Not. R. Astron. Soc. **520,** 5746 (2023).

[23] A. Kostić, N.-M. Nguyen, F. Schmidt, and M. Reinecke, J. Cosmol. Astropart. Phys. 07 (2023) 063.

[24] J. Stadler, F. Schmidt, and M. Reinecke, J. Cosmol. Astropart. Phys. 10 (2023) 069.

[25] N. Agarwal, V. Desjacques, D. Jeong, and F. Schmidt, J. Cosmol. Astropart. Phys. 03 (2021) 021.

[26] C. Hahn and F. Villaescusa-Navarro, J. Cosmol. Astropart. Phys. 04 (2021) 029.

[27] G. Valogiannis and C. Dvorkin, Phys. Rev. D **105,** 103534 (2022).

[28] E. Massara, F. Villaescusa-Navarro, C. Hahn, M. M. Abidi, M. Eickenberg, S. Ho, P. Lemos, A. Moradinezhad Dizgah, and B. R.-S. Blancard, Astrophys. J. **951,** 70 (2023).

[29] E. Paillas, C. Cuesta-Lazaro, P. Zarrouk, Y.-C. Cai, W. J. Percival, S. Nadathur, M. Pinon, A. de Mattia, and F. Beutler, Mon. Not. R. Astron. Soc. **522,** 606 (2023).

[30] E. Paillas *et al.*, arXiv:2309.16541.

[31] C. Hahn, P. Lemos, L. Parker, B. Régaldo-Saint Blancard, M. Eickenberg, S. Ho, J. Hou, E. Massara, C. Modi, A. Moradinezhad Dizgah, and D. Spergel, arXiv:2310.15246.

[32] B. Régaldo-Saint Blancard, C. Hahn, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, L. Parker, Y. Yao, and M. Eickenberg, arXiv:2310.15250.

[33] P. Lemos, L. H. Parker, C. Hahn, S. Ho, M. Eickenberg, J. Hou, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régaldo-Saint Blancard, and D. Spergel, in *Machine Learning for Astrophysics* (2023), p. 18; arXiv:2310.15256.

[34] G. Valogiannis, S. Yuan, and C. Dvorkin, arXiv:2310.16116.

[35] B. Reid *et al.*, Mon. Not. R. Astron. Soc. **455,** 1553 (2016).

[36] S. Alam *et al.* (BOSS Collaboration), Mon. Not. R. Astron. Soc. **470,** 2617 (2017).

[37] J.-P. Dai, L. Verde, and J.-Q. Xia, J. Cosmol. Astropart. Phys. 08 (2020) 007.

[38] P. Chakraborty, S.-F. Chen, and C. Dvorkin, J. Cosmol. Astropart. Phys. 07 (2022) 038.

[39] For analysis in which simulation-based emulators are used to go beyond the perturbative limit (e.g., [34,119,120]), the emulator error is the limiting factor. The increasing precision of the data imposes more stringent constraint on the required accuracy of both PT- and emulator-based modeling.

[40] P. McDonald, Phys. Rev. D **74,** 103512 (2006); **74,** 129901(E) (2006).

[41] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 07 (2012) 051.

[42] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, J. High Energy Phys. 09 (2012) 082.

[43] L. Senatore, J. Cosmol. Astropart. Phys. 11 (2015) 007.

[44] L. Senatore and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 02 (2015) 013.

[45] Z. Vlah, M. White, and A. Aviles, J. Cosmol. Astropart. Phys. 09 (2015) 014.

[46] D. Blas, M. Garny, M. M. Ivanov, and S. Sibiryakov, J. Cosmol. Astropart. Phys. 07 (2016) 052.

[47] M. M. Ivanov and S. Sibiryakov, J. Cosmol. Astropart. Phys. 07 (2018) 053.

[48] G. Cabass, M. M. Ivanov, M. Lewandowski, M. Mirbabayi, and M. Simonović, Phys. Dark Universe **40,** 101193 (2023).

[49] E. Sellentin and A. F. Heavens, Mon. Not. R. Astron. Soc. **473,** 2355 (2018).

[50] C. Hahn, F. Beutler, M. Sinha, A. Berlind, S. Ho, and D. W. Hogg, Mon. Not. R. Astron. Soc. **485,** 2956 (2019).

[51] A. J. Ross *et al.*, Mon. Not. R. Astron. Soc. **424,** 564 (2012).

[52] A. J. Ross *et al.*, Mon. Not. R. Astron. Soc. **464,** 1168 (2017).

[53] E. Kitanidis, M. White, Y. Feng, D. Schlegel, J. Guy, A. Dey, M. Landriau, D. Brooks, M. Levi, J. Moustakas, F. Prada, G. Tarle, and B. A. Weaver, Mon. Not. R. Astron. Soc. **496,** 2262 (2020).

[54] H. Guo, I. Zehavi, and Z. Zheng, Astrophys. J. **756,** 127 (2012).

[55] C. Hahn, R. Scoccimarro, M. R. Blanton, J. L. Tinker, and S. A. Rodríguez-Torres, Mon. Not. R. Astron. Soc. **467,** 1940 (2017).

[56] D. Bianchi, A. Burden, W. J. Percival, D. Brooks, R. N. Cahn, J. E. Forero-Romero, M. Levi, A. J. Ross, and G. Tarle, Mon. Not. R. Astron. Soc. **481**, 2338 (2018).

[57] K. Cranmer, J. Brehmer, and G. Louppe, Proc. Natl. Acad. Sci. U.S.A. **117**, 30055 (2020).

[58] J.-M. Lueckmann, J. Boelts, D. S. Greenberg, P. J. Gonçalves, and J. H. Macke, arXiv:2101.04653.

[59] The SBI is also commonly referred to in the literature as *likelihood-free* or *implicit likelihood* inference.

[60] N. S. M. de Santi *et al.*, Astrophys. J. **952**, 69 (2023).

[61] N. S. M. de Santi *et al.*, arXiv:2310.15234.

[62] B. Tucci and F. Schmidt, arXiv:2310.03741.

[63] C. Modi, S. Pandey, M. Ho, C. Hahn, B. Régaldo-Saint Blancard, and B. Wandelt, arXiv:2309.15071.

[64] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. R.-S. Blancard, and M. M. Abidi, Proc. Natl. Acad. Sci. U.S.A. **120**, e2218810120 (2023).

[65] C. Hahn, M. Eickenberg, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. R.-S. Blancard, and M. M. Abidi, J. Cosmol. Astropart. Phys. 04 (2023) 010.

[66] C. Hahn *et al.*, arXiv:2310.15246.

[67] K. Pardede, F. Rizzo, M. Biagetti, E. Castorina, E. Sefusatti, and P. Monaco, J. Cosmol. Astropart. Phys. 10 (2022) 066.

[68] J. Hou and A. Moradinezhad Dizgah (to be published).

[69] In this derivation we have dropped out the linear term in the overdensity field, which is only relevant for statistically inhomogeneous fields [10,13].

[70] Similarly, in numerical Fisher forecasts [13] the smoothing ensures that only modes on scales not affected by the resolution of the simulations are included in the analysis.

[71] We define the *full set* of redshift-space galaxy skew spectra by including the 14 kernels that encode the line-of-sight velocity information, in contrast to a *subset* of skew spectra that are constructed from the 3 real-space kernels.

[72] F. Villaescusa-Navarro *et al.*, Astrophys. J. Suppl. Ser. **250**, 2 (2020).

[73] H. A. Feldman, N. Kaiser, and J. A. Peacock, Astrophys. J. **426**, 23 (1994).

[74] C. Hahn, R. Scoccimarro, M. R. Blanton, J. L. Tinker, and S. A. Rodríguez-Torres, Mon. Not. R. Astron. Soc. **467**, 1940 (2017).

[75] H. A. Feldman, N. Kaiser, and J. A. Peacock, Astrophys. J. **426**, 23 (1994).

[76] We adopted the FKP weights minimizing the variance for the power spectrum. This may not be optimal for skew spectra. We leave the extension for future work.

[77] K. Yamamoto, M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka, Publ. Astron. Soc. Jpn. **58**, 93 (2006).

[78] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, Astron. J. **156**, 160 (2018).

[79] https://github.com/bccp/nbodykit.

[80] We note that the choice of the fundamental frequency could be susceptible to systematic effects approaching the survey boundary as we discuss in Appendix E.

[81] D. J. Eisenstein *et al.* (SDSS Collaboration), Astron. J. **142**, 72 (2011).

[82] K. S. Dawson *et al.* (BOSS Collaboration), Astron. J. **145**, 10 (2013).

[83] V. Springel, Mon. Not. R. Astron. Soc. **364**, 1105 (2005).

[84] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, Astrophys. J. **762**, 109 (2012).

[85] A. A. Berlind and D. H. Weinberg, Astrophys. J. **575**, 587 (2002).

[86] Z. Zheng, A. L. Coil, and I. Zehavi, Astrophys. J. **667**, 760 (2007).

[87] The overlapping tiling is approximated by randomly downsampling the collided galaxies by 40%.

[88] https://github.com/mockFactory/make_survey.

[89] E. G. Tabak and E. Vanden-Eijnden, Commun. Math. Sci. **8**, 217 (2010).

[90] E. G. Tabak and C. V. Turner, Commun. Pure Appl. Math. **66** (2013).

[91] D. Jimenez Rezende and S. Mohamed, arXiv:1505.05770.

[92] In SIMBIG analyses of the power spectrum, bispectrum and wavelet scattering transforms MAF was used, while in the field-level analysis using CNNs, Neural Spline Flows [NSF;[121]] were utilized. For skew spectra, we found that using the NSF has a significantly slower convergence to stable results.

[93] G. Papamakarios, T. Pavlakou, and I. Murray, arXiv:1705.07057.

[94] 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 Softwaare **5**, 2505 (2020).

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

[96] M. Germain, K. Gregor, I. Murray, and H. Larochelle, Made: Masked autoencoder for distribution estimation, *Proceedings of the 32nd International Conference on Machine Learning* (2015); arXiv:1502.03509.

[97] D. P. Kingma and J. Ba, arXiv:1412.6980.

[98] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, arXiv:1907.10902.

[99] B. Lakshminarayanan, A. Pritzel, and C. Blundell, arXiv:1612.01474.

[100] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, Mon. Not. R. Astron. Soc. **488**, 4440 (2019).

[101] 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).

[102] L. H. Garrison, D. J. Eisenstein, D. Ferrer, N. A. Maksimova, and P. A. Pinto, Mon. Not. R. Astron. Soc. **508**, 575 (2021).

[103] QUIJOTE fiducial cosmology is set to $\{\Omega_m, \Omega_b, h, n_s, \sigma_8\} = \{0.3175, 0.049, 0.6711, 0.9624, 0.834\}$.

[104] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. White, Astrophys. J. **292**, 371 (1985).

[105] The velocity bias of the central galaxies is fixed to $\eta_{cent} = 0.2$ in order to construct galaxy samples with reasonable power spectrum quadrupole.

[106] The base simulations are run at Planck2018 [113] cosmology corresponding to base_plikHM_TT-TEEE_lowl_lowE_lensing data.

[107] B. Hadzhiyska, D. Eisenstein, S. Bose, L. H. Garrison, and N. Maksimova, Mon. Not. R. Astron. Soc. **509**, 501 (2022).

[108] S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman, arXiv:1804.06788.

[109] A. Heavens, R. Jimenez, and O. Lahav, Mon. Not. R. Astron. Soc. **317,** 965 (2000).

[110] J. Alsing and B. Wandelt, Mon. Not. R. Astron. Soc. **476,** L60 (2018).

[111] This is the case with or without the BBN prior.

[112] F. Beutler *et al.* (BOSS Collaboration), Mon. Not. R. Astron. Soc. **466,** 2242 (2017).

[113] N. Aghanim *et al.* (Planck Collaboration), Astron. Astrophys. **641,** A6 (2020); **652,** C4(E) (2021).

[114] S.-F. Chen, P. Chakraborty, and C. Dvorkin, arXiv:2401.13036.

[115] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rep. **733,** 1 (2018).

[116] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00 (Association for Computing Machinery, New York, NY, USA, 2000), p. 93–104.

[117] F. T. Liu, K. M. Ting, and Z.-H. Zhou, *Isolation Forest. 2008 8th IEEE International Conference on Data Mining, Pisa, 2008*, (2008), pp. 413–422, 10.1109/ICDM.2008.17.

[118] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12,** 2825 (2011).

[119] S. Yuan, L. H. Garrison, D. J. Eisenstein, and R. H. Wechsler, Mon. Not. R. Astron. Soc. **515,** 871 (2022).

[120] E. Paillas *et al.*, arXiv:2309.16541.

[121] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, Adv. Neural Inf. Process. Syst. **32** (2019).