SIMBIG: Likelihood-Free Inference of Galaxy Clustering

ChangHoon Hahn [1] **Muntazir Abidi** [2] **Michael Eickenberg** [3] **Shirley Ho** [4] **Pablo Lemos** [5 6] **Elena Massara** [7 8] **Azadeh Moradinezhad Dizgah** [9] **Bruno Régaldo-Saint Blancard** [2]

Abstract

We present SIMBIG, a likelihood-free inference (LFI) framework for analyzing galaxy clustering using a fully simulation-based approach. We apply SIMBIG to the BOSS CMASS galaxy sample using an N-body simulation based forward model that includes a flexible galaxy-halo model, detailed survey geometry, and realistic observational systematics. As demonstration and validation, we use SIMBIG to analyze the galaxy power spectrum out to $k_{\max} = 0.5\,h/\mathrm{Mpc}$. We derive constraints on Ω_m and σ_8 that are a factor of 1.1 and 3 tighter, respectively, than previous results. This improvement comes from the extra cosmological information available on non-linear scales that we can extract with our simulation-based approach. Furthermore, we use a suite of test simulations to confirm that our LFI approach produces conservative estimates of the true posterior. In subsequent work, we will apply SIMBIG to analyze higher-order statistics and non-standard observables such as the bispectrum, marked power spectrum, and wavelet scattering-like statistics.

1. Introduction

The statistical clustering of galaxies provide key cosmological information that can be used to constrain the nature of dark energy and the contents of the Universe. The next generation galaxy surveys, such as DESI, PFS, Euclid, and Roman, will probe galaxies over unprecedented cosmic volumes. Combined with constraints from the cosmic microwave background experiments, they will provide the most stringent tests of the ΛCDM model and potentially lead to new discoveries.

Current analyses utilize the galaxy power spectrum as the primary measurement of galaxy clustering (*e.g.* Beutler et al., 2017; Ivanov et al., 2020; Kobayashi et al., 2021). However, extra information is available in higher-order statistics. Recent studies using large suites of simulations have accurately quantified the information content in higher-order statistics (Hahn et al., 2020; Hahn & Villaescusa-Navarro, 2021). They also find significant constraining power on small scales in the non-linear regime. (Hahn & Villaescusa-Navarro, 2021) find that constraints on Ω_m, Ω_b, h, n_s, σ_8 improve by more than a factor of 2 when the bispectrum is included in the analysis. They also find that including smaller scales ($0.2 < k < 0.5\,h/\mathrm{Mpc}$) tightens constraints by $> 2\times$.

Despite the constraining power available in higher-order statistics and on non-linear scales, they cannot be exploited by standard methods. Standard analyses use analytic models based on perturbation theory, which cannot accurately model scales below the quasi-linear scales. Analytic models are even more limited for higher-order statistics. For instance, in (Philcox & Ivanov, 2021), they limit the power spectrum to $k < 0.2\,h/\mathrm{Mpc}$ and the bispectrum to $k < 0.08\,h/\mathrm{Mpc}$. Perturbation theory also cannot be used to model various newly proposed observables (Eickenberg et al., 2022; Naidoo et al., 2022).

Observational systematics are also a major challenge for current analyses. Fiber collisions, for instance, significantly bias the power spectrum on scales smaller than $k > 0.1\,h/\mathrm{Mpc}$ (Guo et al., 2012; Hahn et al., 2017a; Bianchi et al., 2018). Accurate correction schemes are limited and only available for the power spectrum. Other systematics such as the survey footprint, masks, and completeness also significantly impact clustering measurements.

An alternative *simulation-based* approach can be used to analyze galaxy clustering. With N-body simulations, we can accurately model nonlinear structure formation. We can also more easily forward model observational systematics onto the simulations (*e.g.* Rodríguez-Torres et al., 2016; Rossi et al., 2021). More importantly, recent advances in likelihood-free inference (LFI) enable accurate Bayesian inference with a fully simulation-based approach (*e.g.* Papamakarios et al., 2017; Alsing et al., 2019; Jeffrey et al., 2021; Tortorelli et al., 2021).

In this work, we present SIMulation-Based Inference of Galaxies (SIMBIG), a Likelihood-free inference (LFI; also known as "simulation-based" inference) framework for an-

ICML 2022 Workshop on Machine Learning for Astrophysics, Baltimore, Maryland, USA, 2022.

alyzing galaxy clustering with a full forward model of the observed galaxy survey. We apply SIMBIG to the BOSS CMASS galaxy sample and analyze the galaxy power spectrum multipoles. We demonstrate that we can accurately infer posteriors on cosmological parameters that exploit the constraining power available in the non-linear regime.

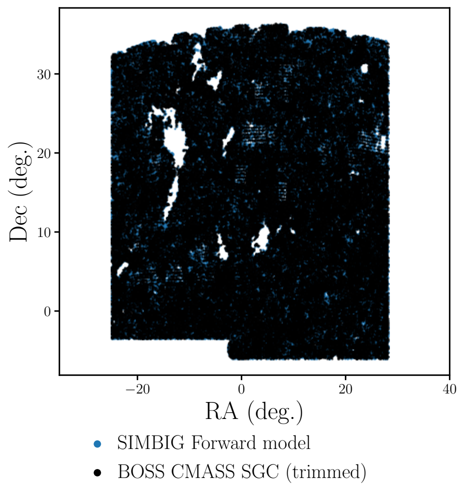

Figure 1. Our forward model of the BOSS CMASS SGC galaxy sample that includes a flexible galaxy-halo connection model, survey geometry, observational systematics, and nuisance parameters. We present the angular distribution of galaxies in a single forward model realization (blue) compared to the BOSS sample (black).

2. Observations: SDSS-III BOSS

In this work, we analyze observations from the Sloan Digital Sky Survey SDSS-III (Eisenstein et al., 2011; Dawson et al., 2013) Baryon Oscillation Spectroscopic Survey (BOSS) Data Release 12. In particular, we focus on the CMASS galaxy sample, which selects high stellar mass Luminous Red Galaxies (LRGs) over the redshift $0.43 < z < 0.7$, in the Southern Galactic Cap (SGC). We refer readers to (Reid et al., 2016) for further details on the galaxy sample.

SIMBIG requires a realistic forward model of the observations; however, our forward model is based on N-body simulations with a cosmological volume of $1(h^{-1}\mathrm{Gpc})^3$ (§ 3), which cannot fit the entire CMASS sample. We therefore restrict the sample to $0.45 < z < 0.6$ and impose the following angular cuts: Dec > -6 and $-25 <$ RA < 28. In Fig. 1, we present the angular distribution of the trimmed CMASS SGC galaxy sample (black).

3. Likelihood-Free Inference

Modern cosmological analyses use Bayesian inference to constrain the posterior probability distribution $p(\theta \,|\, x)$ of cosmological parameters, θ, given observation x. In standard galaxy clustering analyses, the posterior is typically evaluated using Bayes' rule, where the likelihood is assumed to have a Gaussian functional form and evaluated using an analytic perturbation theory model.

LFI offers an alternative that requires no assumptions on the form of the likelihood. Instead, LFI uses a generative forward model, *i.e.* a simulation, to generate mock observations x' given parameters θ'. It uses simulated pairs (θ', x') to directly estimate either the posterior, the likelihood, or $p(\theta, x)$. Hence, it provides an inference framework that can exploit high-fidelity simulations. LFI has already been successfully applied to a wide range of inference problems in astronomy and cosmology (Cameron & Pettitt, 2012; Weyant et al., 2013; Hahn et al., 2017b; Huppenkothen & Bachetti, 2021; Zhang et al., 2021).

In this work, we utilize neural density estimation based LFI, where a neural network with parameters ϕ is trained using the simulated pairs (θ', x') to estimate the density $p(\theta \,|\, x')$. In particular, we use normalizing flows as density estimators (Tabak & Vanden-Eijnden, 2010; Tabak & Turner, 2013).

3.1. Forward Model

A key ingredient in LFI is a forward model of the observations, *i.e.* the BOSS SGC galaxy sample. LFI assumes that the forward model is capable of generating mock observationss that is practically indistinguishable from the observations. In our case, this means that our forward model must generate a mock galaxy sample that includes the exact survey geometry and observational systematics.

We start with 2,000 N-body simulations from the QUIJOTE suite (Villaescusa-Navarro et al., 2020). The simulations are run using the TreePM GADGET-III code. They have a cosmological volume of $1(h^{-1}\mathrm{Gpc})^3$ and follow the evolution of 1024^3 cold dark matter (CDM) particles that were initalized at $z = 127$ using 2LPT. We use the $z = 0.5$ snapshot. Each simulation is constructed for a different cosmology that was selected based on a Latin-hypercube (LH) configuration. We use N-body simulations, despite their significant computational cost, due to their accuracy in modeling small-scale, non-linear, clustering.

The N-body simulations provide the dark matter distribution. Next, we identify dark matter halos using ROCKSTAR (Behroozi et al., 2013) and populate them with galaxies using the Halo Occupation Distribution (HOD) framework. HOD models provide a statistical prescription for populating halos with galaxies by sampling $p(\theta_g \,|\, \theta_h)$. θ_g

represents the properties of galaxies in a halo such as occupation number, halo-centric position, and velocity; θ_h represents halo properties such as mass and concentration. In this work, we use an HOD model that includes assembly, concentration, and velocity biases for central and satellite galaxies on top of the standard (Zheng et al., 2007) model. For each N-body simulation, we construct 5 different galaxy realizations each with a different set of HOD parameters sampled from a prior based on (Reid et al., 2014).

Next, we impose the survey geometry and observational systematics. We first perform volume remapping on the simulated galaxy catalog to a cuboid volume using (Carlson & White, 2010). This step is necessary to maximize the survey volume that can be fit into the simulation. Then, we use MANGLE polygons (Swanson et al., 2008) to cut out the survey geometry as well as the veto mask, which includes masking for bright star, centerpost, bad field, collision priority (Dawson et al., 2013). We trim the forward modeled catalog to match the $0.45 < z < 0.6$ range of the observations. Lastly, we apply fiber collisions. We identify all pairs of galaxies within an angular scale of $62''$; then, for a randomly selected 60% of the pairs, we remove one of the galaxies from the sample.

Galaxy clustering analyses typically include a number of nuisance parameters. For instance, a correction for shot noise, $A_{\rm shot}$, is typically included to account for residual shot noise contribution (*e.g.* Beutler et al., 2017; Kobayashi et al., 2021). For consistency, we also include $A_{\rm shot}$ in our analysis. In Fig. 1, we compare the angular distribution of a single realization of our forward model (blue) to the observed BOSS galaxy sample (black). The forward model has the same angular footprint as the observations.

3.2. Summary Statistic: the Galaxy Power spectrum

Since the primary goal of this work is to demonstrate SIMBIG for analyzing galaxy clustering, we use a summary statistic that has been extensively used in the past: the galaxy power spectrum multipole, $P_\ell(k)$. By using P_ℓ, we can compare the constraints we derive from SIMBIG with previous constraints (*e.g.* Ivanov et al., 2020) and validate the accuracy of our approach. In principle, SIMBIG can be directly applied to the full galaxy catalog if the forward model is capable of accurately modeling observations at all scales. Even with N-body simulations, however, this is not the case due to limitatons on mass and time resolution. Instead, we can use summary statistics of the galaxy sample, where we can impose cuts, *e.g.* based on scale, to which our forward model is able to reproduce observations. We measure the power spectrum monopole, quadrupole, and hexadecapole ($\ell = 0, 2,$ and 4) using the (Hand et al., 2017) algorithm of out to $k_{\max} = 0.5\,h/\mathrm{Mpc}$. In Fig. 2, we present P_ℓ for all 10,000 forward modeled galaxy simulations: $\ell = 0$ (blue), 2 (orange), and 4 (green).

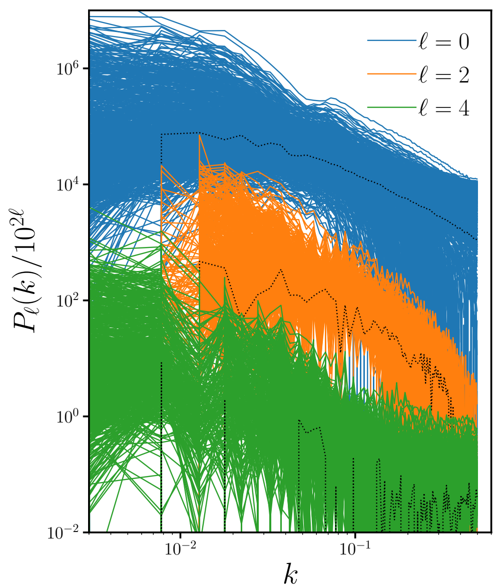

Figure 2. Power spectrum multipoles, $P_\ell(k)$, measured for each of the 10,000 forward modeled galaxy simulations used in this work. We represent the monopole, quadrpuole, and hexadecapole ($\ell = 0, 2, 4$) in blue, orange, and green, respectively, and offset them for clarity. We include $P_\ell(k)$ measured for our observational BOSS CMASS SGC galaxy sample (black dotted) for comparison.

3.3. Training the Neural Density Estimator

Among the various normalizing flow density estimators now available in the literature, we use a Masked Autoregressive Flow (MAF; Papamakarios et al., 2017). MAF combines normalizing flows with an autoregressive design (Uria et al., 2016), which is well-suited for estimating conditional probability distributions such as a posterior. A MAF model is built by stacking multiple Masked Autoencoder for Distribution Estimation (MADE; Germain et al., 2015) models so that it has the autoregressive structure of MADE but with more flexibility to describe complex probability distributions. We use the MAF implementation in the `sbi` Python package (Greenberg et al., 2019; Tejero-Cantero et al., 2020).

In training, our goal is to determine the parameters, ϕ, of our MAF model, $p_\phi(\theta\,|\,x)$, so that it accurately estimates the posterior, $p(\theta\,|\,x)$. We do this by minimizing the KL divergence between $p_\phi(\theta\,|\,x)$ and $p(\theta\,|\,x)$. We split the training data into a training and validation set with a 90/10 split, then maximize the total log likelihood $\sum_i \log p_\phi(\theta_i|x_i)$ over training set, which is equivalent to minimizing the KL divergence. We use the ADAM optimizer (Kingma & Ba, 2017) with a learning rate of 5×10^{-4}. We prevent overfitting by stopping the training when the validation log likelihood fails to increase after 20 epochs. We determine the architecture of our MAF model through experimentation. Our final trained MAF model has 6 MADE blocks, each with 2 hidden layers and 301 hidden units.

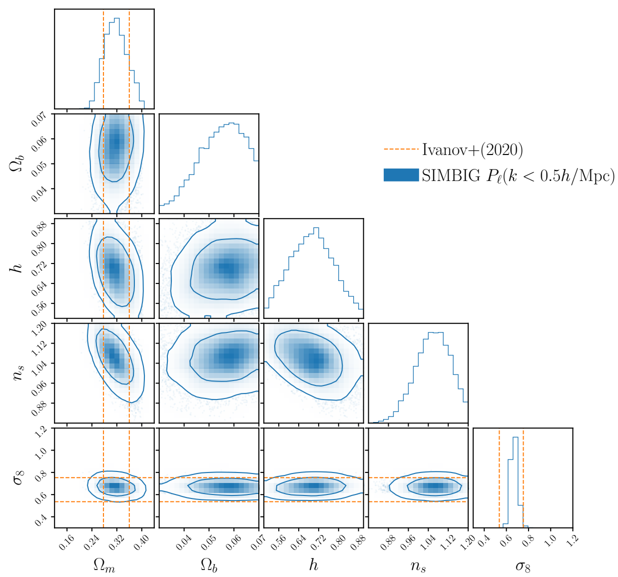

Figure 3. Posterior of cosmological parameters derived from $P_\ell(k < 0.5\,h/\mathrm{Mpc})$ using LFI with SIMBIG. We mark the 1 and 2σ confidence intervals on the contours. We include the 64% percentile range of the marginalized posteriors on Ω_m and σ_8 from the P_ℓ analysis of (Ivanov et al., 2020). They analyze P_ℓ to $k_{\mathrm{max}} < 0.25\,h/\mathrm{Mpc}$. With our simulation-based approach, we can accurately model and extract cosmological information from the non-linear regime. Hence, with SIMBIG we derive constraints on Ω_m and σ_8 ~1.1 and 3× tighter than previous work.

4. Results

We train our MAF model to accurately estimate the posterior for the galaxy power spectrum. In Fig. 3, we present the posterior of the cosmological parameters, evaluated using $p_\phi(\theta\,|\,P_\ell)$. For clarity, we do not include the posteriors on the HOD parameters or A_{shot}. We do not derive tight constraints on Ω_b, h, or n_s, which is consistent with previous galaxy power spectrum constraints. On the other hand, we derive significant constraints on Ω_m and σ_8. We infer $\Omega_m = 0.316^{+0.040}_{-0.036}$. This is in good agreement and a slight 10% improvement over the recent (Ivanov et al., 2020) constraint for the same CMASS SGC high-z sample. The orange dashed lines mark the 64% percentile range of their posterior. Meanwhile, our $\sigma_8 = 0.668^{+0.0324}_{-0.0284}$ constraint is significantly tighter than that of (Ivanov et al., 2020) and corresponds to an improvement of a factor of ~ 3.

The improvements we find over previous constraints are driven by extra cosmological information that we extract from the small scale clustering. (Ivanov et al., 2020) limit their analysis to $k_{\mathrm{max}} < 0.25\,h/\mathrm{Mpc}$ due to limitations in their analytical model. Meanwhile, we can accurately extend our P_ℓ analysis down to k=$0.5\,h/\mathrm{Mpc}$. When we repeat our analysis with k_{max}=$0.25\,h/\mathrm{Mpc}$, we find roughly the same constraints as (Ivanov et al., 2020). We also note that our improvements are consistent with the (Hahn & Villaescusa-Navarro, 2021) forecast for $P_\ell(k < 0.5\,h/\mathrm{Mpc})$ versus $P_\ell(k < 0.2\,h/\mathrm{Mpc})$.

To further validate our posteriors, we forward model a test set of 500 galaxy catalogs at a fiducial cosmology using a different set of QUIJOTE simulations. We derive posteriors for each of the test catalogs and since we know the true cosmological parameters that went into generating them, we use simulation-based calibration (Talts et al., 2020). For each test catalog, we measure the rank statistic of the true parameter values within the marginalized posterior distribution of the cosmological parameters. Then, we examine the distribution of the rank statistics for all of the test catalogs to confirm that $p_\phi(\theta\,|\,x)$ does not underestimate the width of our posteriors. More specificially, we find that our $p_\phi(\theta\,|\,x)$ accurately estimates the σ_8 constraints but slightly overestimates the constraints for all other cosmological parameters. Hence, the posterior in Fig 3 are conservative estimates of the true posterior. The conservative estimate of $p(\theta\,|\,P_\ell)$ is due to the fact that we have a limited sampling of the cosmological parameter space (2,000 different cosmologies) in our training data. We are currently constructing more N-body simulations to address this limitation.

5. Conclusions

In this work, we present SIMBIG, a simulation-based framework for analyzing galaxy clustering using LFI. SIMBIG utilizes a full forward model of observations based on N-body simulations that enables us to accurately model small scale clustering as well as observation systematics such as fiber collisions. We apply SIMBIG to the BOSS CMASS SGC $z \sim 0.5$ galaxy sample and analyze the power spectrum multipoles to demonstrate and validate our approach. Using neural density estimation based LFI, we infer $\Omega_m = 0.316^{+0.040}_{-0.036}$ and $\sigma_8 = 0.668^{+0.0324}_{-0.0284}$. These constraints are significantly tighter than the (Ivanov et al., 2020) constraints on the same galaxy sample. This improvement demonstrates that there is significant constraining power on scales beyond $k \gtrsim 0.2\,h/\mathrm{Mpc}$. In subsequent work, we will use SIMBIG to analyse the BOSS galaxies using higher-order statistics and non-standard observables that contain even more cosmological information, such as the bispectrum, marked powerspectrum, and wavelet scattering-like statistics. As upcoming galaxy surveys (DESI, PFS, Euclid, Roman) probe unprecedented cosmic volumes, SIMBIG provides a framework to maximize their statistical power to tightly constrain cosmological parameters.

References

Alsing, J., Charnock, T., Feeney, S., and Wandelt, B. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488:4440–4458, September 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1960.

Behroozi, P. S., Wechsler, R. H., and Wu, H.-Y. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *The Astrophysical Journal*, 762:109, January 2013. ISSN 0004-637X. doi: 10.1088/0004-637X/762/2/109.

Beutler, F., Seo, H.-J., Saito, S., Chuang, C.-H., Cuesta, A. J., Eisenstein, D. J., Gil-Marín, H., Grieb, J. N., Hand, N., Kitaura, F.-S., Modi, C., Nichol, R. C., Olmstead, M. D., Percival, W. J., Prada, F., Sánchez, A. G., Rodriguez-Torres, S., Ross, A. J., Ross, N. P., Schneider, D. P., Tinker, J., Tojeiro, R., and Vargas-Magaña, M. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Anisotropic galaxy clustering in Fourier space. *Monthly Notices of the Royal Astronomical Society*, 466:2242–2260, April 2017. ISSN 0035-8711. doi: 10.1093/mnras/stw3298.

Bianchi, D., Burden, A., Percival, W. J., Brooks, D., Cahn, R. N., Forero-Romero, J. E., Levi, M., Ross, A. J., and Tarle, G. Unbiased clustering estimates with the DESI fibre assignment. *Monthly Notices of the Royal Astronomical Society*, 481:2338–2348, December 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty2377.

Cameron, E. and Pettitt, A. N. Approximate Bayesian Computation for astronomical model analysis: A case study in galaxy demographics and morphological transformation at high redshift. *Monthly Notices of the Royal Astronomical Society*, 425:44–65, September 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.21371.x.

Carlson, J. and White, M. Embedding realistic surveys in simulations through volume remapping. *The Astrophysical Journal Supplement Series*, 190(2):311–314, October 2010. ISSN 0067-0049, 1538-4365. doi: 10.1088/0067-0049/190/2/311.

Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Filiz Ak, N., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene, M. A., Kneib, J.-P., Le Goff, J.-M., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D., Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West, A. A., White, M., Wood-Vasey, W. M., Yeche, C., Zehavi, I., Zhao, G.-B., and Zheng, Z. The Baryon Oscillation Spectroscopic Survey of SDSS-III. *The Astronomical Journal*, 145:10, January 2013. ISSN 0004-6256. doi: 10.1088/0004-6256/145/1/10.

Eickenberg, M., Allys, E., Moradinezhad Dizgah, A., Lemos, P., Massara, E., Abidi, M., Hahn, C., Hassan, S., Regaldo-Saint Blancard, B., Ho, S., Mallat, S., Andén, J., and Villaescusa-Navarro, F. Wavelet Moments for Cosmological Parameter Estimation, April 2022.

Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Allende Prieto, C., Anderson, S. F., Arns, J. A., Aubourg, É., Bailey, S., Balbinot, E., Barkhouser, R., Beers, T. C., Berlind, A. A., Bickerton, S. J., Bizyaev, D., Blanton, M. R., Bochanski, J. J., Bolton, A. S., Bosman, C. T., Bovy, J., Brandt, W. N., Breslauer, B., Brewington, H. J., Brinkmann, J., Brown, P. J., Brownstein, J. R., Burger, D., Busca, N. G., Campbell, H., Cargile, P. A., Carithers, W. C., Carlberg, J. K., Carr, M. A., Chang, L., Chen, Y., Chiappini, C., Comparat, J., Connolly, N., Cortes, M., Croft, R. A. C., Cunha, K., da Costa, L. N., Davenport, J. R. A., Dawson, K., De Lee, N., Porto de Mello, G. F., de Simoni, F., Dean, J., Dhital, S., Ealet, A., Ebelke, G. L., Edmondson, E. M., Eiting, J. M., Escoffier, S., Esposito, M., Evans, M. L., Fan, X., Femenía Castellá, B., Dutra Ferreira, L., Fitzgerald, G., Fleming, S. W., Font-Ribera, A., Ford, E. B., Frinchaboy, P. M., García Pérez, A. E., Gaudi, B. S., Ge, J., Ghezzi, L., Gillespie, B. A., Gilmore, G., Girardi, L., Gott, J. R., Gould, A., Grebel, E. K., Gunn, J. E., Hamilton, J.-C., Harding, P., Harris, D. W., Hawley, S. L., Hearty, F. R., Hennawi, J. F., González Hernández, J. I., Ho, S., Hogg, D. W., Holtzman, J. A., Honscheid, K., Inada, N., Ivans, I. I., Jiang, L., Jiang, P., Johnson, J. A., Jordan, C., Jordan, W. P., Kauffmann, G., Kazin, E., Kirkby, D., Klaene, M. A., Knapp, G. R., Kneib, J.-P., Kochanek, C. S., Koesterke, L., Kollmeier, J. A., Kron, R. G., Lampeitl, H., Lang, D., Lawler, J. E., Le Goff, J.-M., Lee, B. L., Lee, Y. S., Leisenring, J. M., Lin, Y.-T., Liu, J., Long, D. C., Loomis, C. P., Lucatello, S., Lundgren, B., Lupton, R. H., Ma, B., Ma, Z., MacDonald, N., Mack, C., Mahadevan, S., Maia, M. A. G., Majewski, S. R., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Maraston, C., Margala, D., Maseman, P., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena Requejo, O., Ménard, B., Miralda-Escudé, J., Morrison, H. L., Mullally, F., Muna, D., Murayama, H., Myers, A. D., Naugle, T., Neto, A. F., Nguyen, D. C., Nichol, R. C., Nidever, D. L., O'Connell, R. W., Ogando, R. L. C., Olmstead, M. D., Oravetz, D. J., Padmanabhan, N., Paegert, M., Palanque-Delabrouille, N., Pan, K., Pandey, P., Parejko, J. K., Pâris, I., Pellegrini, P., Pepper, J., Percival, W. J., Petitjean, P., Pfaffenberger, R., Pforr, J., Phleps, S., Pichon, C., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Ramos, B. H. F., Reid, I. N., Reyle, C., Rich, J., Richards, G. T., Rieke, G. H., Rieke, M. J., Rix, H.-W., Robin, A. C., Rocha-Pinto, H. J., Rockosi, C. M., Roe, N. A., Rollinde, E., Ross, A. J., Ross, N. P., Rossetto, B., Sánchez, A. G., Santiago, B., Sayres, C., Schiavon, R., Schlegel, D. J., Schlesinger, K. J., Schmidt, S. J., Schneider, D. P., Sellgren, K., Shelden, A., Sheldon, E., Shetrone, M., Shu, Y., Silverman, J. D., Simmerer, J., Simmons, A. E., Sivarani, T., Skrutskie, M. F., Slosar, A., Smee, S., Smith, V. V., Snedden, S. A., Stassun, K. G., Steele, O., Steinmetz, M., Stockett, M. H., Stollberg, T., Strauss, M. A., Szalay, A. S., Tanaka, M., Thakar, A. R., Thomas, D., Tinker, J. L., Tofflemire, B. M., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Vogt, N. P., Wake, D. A., Wan, X., Wang, J., Weaver, B. A., White, M., White, S. D. M., Wilson, J. C., Wisniewski, J. P., Wood-Vasey, W. M., Yanny, B., Yasuda, N., Yèche, C., York, D. G., Young, E., Zasowski, G., Zehavi, I., and Zhao, B. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *The Astronomical Journal*, 142:72, September 2011. ISSN 0004-6256. doi: 10.1088/0004-6256/142/3/72.

Germain, M., Gregor, K., Murray, I., and Larochelle, H. MADE: Masked Autoencoder for Distribution Estimation. *arXiv:1502.03509 [cs, stat]*, June 2015.

Greenberg, D. S., Nonnenmacher, M., and Macke, J. H. Automatic Posterior Transformation for Likelihood-Free Inference, May 2019.

Guo, H., Zehavi, I., and Zheng, Z. A New Method to Correct for Fiber Collisions in Galaxy Two-point Statistics. *The Astrophysical Journal*, 756:127, September 2012. ISSN 0004-637X. doi: 10.1088/0004-637X/756/2/127.

Hahn, C. and Villaescusa-Navarro, F. Constraining Mν with the bispectrum. Part II. The information content of the galaxy bispectrum monopole. *Journal of Cosmology and Astroparticle Physics*, 2021:029, April 2021. ISSN 1475-7516. doi: 10.1088/1475-7516/2021/04/029.

Hahn, C., Scoccimarro, R., Blanton, M. R., Tinker, J. L., and Rodríguez-Torres, S. A. The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles. *Monthly Notices of the Royal Astronomical Society*, 467:1940–1956, May 2017a. ISSN 0035-8711. doi: 10.1093/mnras/stx185.

Hahn, C., Vakili, M., Walsh, K., Hearin, A. P., Hogg, D. W., and Campbell, D. Approximate Bayesian Computation in Large Scale Structure: Constraining the galaxy-halo connection. *Monthly Notices of the Royal Astronomical Society*, 469(3):2791–2805, August 2017b. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stx894.

Hahn, C., Villaescusa-Navarro, F., Castorina, E., and Scoccimarro, R. Constraining Mν with the bispectrum. Part I. Breaking parameter degeneracies. *Journal of Cosmology and Astroparticle Physics*, 03:040, March 2020. ISSN 1475-7516. doi: 10.1088/1475-7516/2020/03/040.

Hand, N., Li, Y., Slepian, Z., and Seljak, U. An optimal FFT-based anisotropic power spectrum estimator. *Journal of Cosmology and Astro-Particle Physics*, 07:002, July 2017. ISSN 1475-7516. doi: 10.1088/1475-7516/2017/07/002.

Huppenkothen, D. and Bachetti, M. Accurate X-ray Timing in the Presence of Systematic Biases With Simulation-Based Inference, April 2021.

Ivanov, M. M., Simonović, M., and Zaldarriaga, M. Cosmological parameters from the BOSS galaxy power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2020:042, May 2020. ISSN 1475-7516. doi: 10.1088/1475-7516/2020/05/042.

Jeffrey, N., Alsing, J., and Lanusse, F. Likelihood-free inference with neural compression of DES SV weak lensing map statistics. *Monthly Notices of the Royal Astronomical Society*, 501:954–969, February 2021. ISSN 0035-8711. doi: 10.1093/mnras/staa3594.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017.

Kobayashi, Y., Nishimichi, T., Takada, M., and Miyatake, H. Full-shape cosmology analysis of SDSS-III BOSS galaxy power spectrum using emulator-based halo model: A \$5\\%\$ determination of \$\\sigma_8\$. *arXiv:2110.06969 [astro-ph]*, October 2021.

Naidoo, K., Massara, E., and Lahav, O. Cosmology and neutrino mass with the Minimum Spanning Tree. *Monthly Notices of the Royal Astronomical Society*, April 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac1138.

Papamakarios, G., Pavlakou, T., and Murray, I. Masked Autoregressive Flow for Density Estimation. *arXiv e-prints*, 1705:arXiv:1705.07057, May 2017.

Philcox, O. H. E. and Ivanov, M. M. The BOSS DR12 Full-Shape Cosmology: \$\\Lambda\$CDM Constraints from the Large-Scale Galaxy Power Spectrum and Bispectrum Monopole. *arXiv:2112.04515 [astro-ph, physics:hep-ex]*, December 2021.

Reid, B., Ho, S., Padmanabhan, N., Percival, W. J., Tinker, J., Tojeiro, R., White, M., Eisenstein, D. J., Maraston, C., Ross, A. J., Sánchez, A. G., Schlegel, D., Sheldon, E., Strauss, M. A., Thomas, D., Wake, D., Beutler, F., Bizyaev, D., Bolton, A. S., Brownstein, J. R., Chuang, C.-H., Dawson, K., Harding, P., Kitaura, F.-S., Leauthaud, A., Masters, K., McBride, C. K., More, S., Olmstead, M. D., Oravetz, D., Nuza, S. E., Pan, K., Parejko, J., Pforr, J., Prada, F., Rodríguez-Torres, S., Salazar-Albornoz, S., Samushia, L., Schneider, D. P., Scóccola, C. G., Simmons, A., and Vargas-Magana, M. SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: Galaxy target selection and large-scale structure catalogues. *Monthly Notices of the Royal Astronomical Society*, 455:1553–1573, January 2016. ISSN 0035-8711. doi: 10.1093/mnras/stv2382.

Reid, B. A., Seo, H.-J., Leauthaud, A., Tinker, J. L., and White, M. A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies. *Monthly Notices of the Royal Astronomical Society*, 444:476–502, October 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu1391.

Rodríguez-Torres, S. A., Chuang, C.-H., Prada, F., Guo, H., Klypin, A., Behroozi, P., Hahn, C. H., Comparat, J., Yepes, G., Montero-Dorta, A. D., Brownstein, J. R., Maraston, C., McBride, C. K., Tinker, J., Gottlöber, S., Favole, G., Shu, Y., Kitaura, F.-S., Bolton, A., Scoccimarro, R., Samushia, L., Schlegel, D., Schneider, D. P., and Thomas, D. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release. *Monthly Notices of the Royal Astronomical Society*, 460:1173–1187, August 2016. ISSN 0035-8711. doi: 10.1093/mnras/stw1014.

Rossi, G., Choi, P. D., Moon, J., Bautista, J. E., Gil-Marin, H., Paviot, R., Vargas-Magana, M., de la Torre, S., Fromenteau, S., Ross, A. J., Avila, S., Burtin, E., Dawson, K. S., Escoffier, S., Habib, S., Heitmann, K., Hou, J., Mueller, E.-M., Percival, W. J., Smith, A., Zhao, C., and Zhao, G.-B. The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: N-body Mock Challenge for Galaxy Clustering Measurements. *Monthly Notices of the Royal Astronomical Society*, 505(1):377–407, May 2021. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/staa3955.

Swanson, M. E. C., Tegmark, M., Hamilton, A. J. S., and Hill, J. C. Methods for rapidly processing angular masks of next-generation galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 387:1391–1402, July 2008. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2008.13296.x.

Tabak, E. G. and Turner, C. V. A Family of Nonparametric Density Estimation Algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013. ISSN 1097-0312. doi: 10.1002/cpa.21423.

Tabak, E. G. and Vanden-Eijnden, E. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, March 2010. ISSN 1945-0796. doi: 10.4310/CMS.2010.v8.n1.a11.

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. Validating Bayesian Inference Algorithms with Simulation-Based Calibration. *arXiv:1804.06788 [stat]*, October 2020.

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., Greenberg, D. S., and Macke, J. H. Sbi: A toolkit for simulation-based inference. *Journal of Open Source Software*, 5(52):2505, August 2020. ISSN 2475-9066. doi: 10.21105/joss.02505.

Tortorelli, L., Siudek, M., Moser, B., Kacprzak, T., Berner, P., Refregier, A., Amara, A., García-Bellido, J., Cabayol, L., Carretero, J., Castander, F. J., De Vicente, J., Eriksen, M., Fernandez, E., Gaztanaga, E., Hildebrandt, H., Joachimi, B., Miquel, R., Sevilla-Noarbe, I., Padilla, C., Renard, P., Sanchez, E., Serrano, S., Tallada-Crespí, P., and Wright, A. H. The PAU Survey: Measurement of Narrow-band galaxy properties with Approximate Bayesian Computation. *arXiv:2106.02651 [astro-ph]*, June 2021.

Uria, B., Côté, M.-A., Gregor, K., Murray, I., and Larochelle, H. Neural Autoregressive Distribution Estimation. *arXiv:1605.02226 [cs]*, May 2016.

Villaescusa-Navarro, F., Hahn, C., Massara, E., Banerjee, A., Delgado, A. M., Ramanah, D. K., Charnock, T., Giusarma, E., Li, Y., Allys, E., Brochard, A., Uhlemann, C., Chiang, C.-T., He, S., Pisani, A., Obuljen, A., Feng, Y., Castorina, E., Contardo, G., Kreisch, C. D., Nicola, A., Alsing, J., Scoccimarro, R., Verde, L., Viel, M., Ho, S., Mallat, S., Wandelt, B., and Spergel, D. N. The Quijote Simulations. *The Astrophysical Journal Supplement Series*, 250:2, September 2020. doi: 10.3847/1538-4365/ab9d82.

Weyant, A., Schafer, C., and Wood-Vasey, W. M. Likelihood-free Cosmological Inference with Type Ia Supernovae: Approximate Bayesian Computation for a Complete Treatment of Uncertainty. *The Astrophysical Journal*, 764:116, February 2013. ISSN 0004-637X. doi: 10.1088/0004-637X/764/2/116.

Zhang, K., Bloom, J. S., Gaudi, B. S., Lanusse, F., Lam, C., and Lu, J. R. Real-Time Likelihood-Free Inference of Roman Binary Microlensing Events with Amortized Neural Posterior Estimation. February 2021. doi: 10.3847/1538-3881/abf42e.

Zheng, Z., Coil, A. L., and Zehavi, I. Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering. *The Astrophysical Journal*, 667:760–779, October 2007. ISSN 0004-637X. doi: 10.1086/521074.