SimBIG: Field-level Simulation-based Inference of Large-scale Structure

Pablo Lemos [*1,2,3,4] **Liam Parker** [*5] **ChangHoon Hahn** [5] **Shirley Ho** [4,5,6,7] **Michael Eickenberg** [8] **Jiamin Hou** [9,10] **Elena Massara** [11,12] **Chirag Modi** [4,8] **Azadeh Moradinezhad Dizgah** [13] **Bruno Régaldo-Saint Blancard** [4] **David Spergel** [4,5]

Abstract

We present the first simulation-based inference (SBI) of cosmological parameters from field-level analysis of galaxy clustering. Standard galaxy clustering analyses rely on analyzing summary statistics, such as the power spectrum, P_ℓ, with analytic models based on perturbation theory. Consequently, they do not fully exploit the non-linear and non-Gaussian features of the galaxy distribution. To address these limitations, we use the SIMBIG forward modelling framework to perform SBI using normalizing flows. We apply SIMBIG to a subset of the BOSS CMASS galaxy sample using a convolutional neural network with stochastic weight averaging to perform massive data compression of the galaxy field. We infer constraints on $\Omega_m = 0.295^{+0.036}_{-0.033}$ and $\sigma_8 = 0.753^{+0.040}_{-0.036}$ that are $1.03\times$ and $2.65\times$ tighter than those from standard P_ℓ analyses. Our analysis also provides constraints on the Hubble constant $H_0 = 63.1 \pm 4.1$ km/s/Mpc from galaxy clustering alone. This higher constraining power comes from additional non-Gaussian cosmological information, inaccessible with P_ℓ. We demonstrate the robustness of our analysis by showcasing our ability to infer unbiased cosmological constraints from a series of test simulations that are constructed using different forward models than the one used in our training dataset. This work not only presents competitive cosmological constraints but also introduces novel methods for leveraging additional cosmological information in upcoming galaxy surveys like DESI, PFS, and *Euclid*.

[1]Department of Physics, Université de Montréal, Montréal, 1375 Avenue Thérèse-Lavoie-Roux, QC H2V 0B3, Canada [2]Mila - Quebec Artificial Intelligence Institute, Montréal, 6666 Rue Saint-Urbain, QC H2S 3H1, Canada [3]Ciela - Montreal Institute for Astrophysical Data Analysis and Machine Learning, Montréal, Canada [4]Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA [5]Department of Physics, Princeton University, Princeton NJ 08544, USA [6]Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA [7]Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA [8]Center for Computational Mathematics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA [9]Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 32611, USA [10]Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstrasse 1, 85748 Garching bei München, Germany [11]Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada [12]Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada [13]Département de Physique Théorique, Université de Genève, 24 quai Ernest Ansermet, 1211 Genève 4, Switzerland. Correspondence to: Pablo Lemos <pablo.lemos@umontreal.ca>, Liam Parker <lhparker@princeton.edu>.

ICML 2023 Workshop on Machine Learning for Astrophysics, Honolulu, Hawaii, USA. PMLR 202, 2023.

1. Introduction:

Cosmological parameters play a crucial role in our understanding of the evolution and structure of the cosmos. These parameters can be inferred from a variety of observational data, including measurements of the statistical properties of the large-scale structure (LSS) of the universe.

Traditionally, cosmological parameter inference has relied on summary statistics — most often the power spectrum, $P_\ell(k)$ (Beutler et al., 2017; Ivanov et al., 2020; Kobayashi et al., 2022). Analyses also model galaxy clustering analytically using perturbation theory (PT; Bernardeau et al., 2002; Desjacques et al., 2018). Consequently, these analyses have been limited to large scales where the deviation from linear PT is small. Moreover, by relying on P_ℓ, these methods discard non-Gaussian information. However, a recent line of inquiry (*e.g.* D'Amico et al., 2022; Philcox & Ivanov, 2022) has established that there is significant non-Gaussian cosmological information on non-linear scales. Another major challenge of galaxy clustering analyses is their inability to fully account for observational systematics, such as fiber collisions (Hahn et al., 2017; Bianchi et al., 2018). Finally, these analyses are limited by the assumed Gaussian functional form of the likelihood function used.

To overcome these limitations, we instead use Simulation-Based Inference (SBI), which consists in using forward models of the observables, instead of theoretical predictions, and then predicting a posterior distribution over the parameters. This method allows for the incorporation of complex physical processes and modeling uncertainties, and can provide more robust inferences than methods based on analytical models. There have already been multiple applications of SBI in astronomy. In the specific context of analyzing galaxy clustering, Hahn et al.; 2022) introduced the SIMulation-Based Inference of Galaxies (SIMBIG) forward models, which produce realistic mock observations of the Baryon Oscillation Spectroscopic Survey (BOSS) (Eisenstein et al., 2011; Dawson et al., 2012) Southern Galactic Cap (SGC) at different cosmologies, and includes systematic effects such as survey geometry and fiber collisions. Using these models, the authors were able to robustly infer ΛCDM parameters from the BOSS CMASS-SGC. However, their analysis still relied on compressing to the power spectrum, which was unable to capture the non-Gaussian information present.

In this work, we extend the SIMBIG analysis by applying SBI to LSS directly at field-level. Specifically, we use convolutional neural networks (CNNs) to perform a step of massive data compression of the galaxy distribution, as has previously been done for SBI on weak lensing maps (Jeffrey et al., 2021). With this approach, we analyze $109,636$ galaxies from the Baryon Oscillation Spectroscopic Survey (BOSS) (Eisenstein et al., 2011; Dawson et al., 2012) Southern Galactic Cap (SGC) CMASS sample in the redshift range $(0.45 < z < 0.6)$. To apply CNNs to the data, we mesh the data into a box with voxel size $64 \times 128 \times 128$, convert to a density field, and impose a scale cut $k < k_{\rm max} = 0.28\,h/{\rm Mpc}$.

2. Methods

2.1. Forward Models

We use the SIMBIG forward modelling pipeline (Hahn et al., 2022) to generate field-level synthetic observations that aim to be statistically indistinguishable from the real BOSS observations. This pipeline consists of four distinct steps: (1) N-body simulations, (2) a dark matter halo finder, (3) a dark matter halo occupation distribution framework (HOD), and (4) application of survey realism. A schematic illustrating the pipeline is provided in Figure 1. For more details on the forward models, we refer the reader to **??**. Ultimately, the forward models are determined by 5 ΛCDM cosmological parameters, $\Omega_m, \Omega_b, h, n_s, \sigma_8$, and 9 HOD parameters.

To construct our training set, we use 2,518 high-resolution QUIJOTE N-body simulations arranged in a Latin hypercube configuration (LHC), which conservatively encompass the *Planck* cosmological constraints. For each simulation, we forward-model 10 CMASS-like galaxy catalogs using unique HOD parameters randomly sampled from a conservative prior. We split the resulting $25,180$ simulations into a $20,000$ and $5,180$ training and validation set.

Additionally, in order to demonstrate that we can infer unbiased cosmological constraints, we test our analysis on a suite of "mock challenge" data sets specifically designed for cross-validation with the BOSS CMASS sample (Hahn et al., 2022; Hahn et al.). These include three different suites of test simulations: (1) `TEST0`, 500 samples constructed using the same forward model as the training data, (2) `TEST1`, 500 samples constructed using a different halo finder and HOD model, and (3) `TEST2`, 1,000 samples constructed using a different N-body simulation and halo finder.

2.2. Compression

CNNs are flexible functions that can be optimized to extract maximally relevant features from their inputs. In this study, we train a three-dimensional CNN to compress the density fields produced by the SIMBIG forward models to the cosmological parameters of those models. Specifically, the CNN takes as input the three-dimensional tensor representing the discretized forward model, $x \in \mathbb{R}^{64\times128\times128}$, and outputs a prediction of the ΛCDM cosmological parameters, $\boldsymbol{\theta} = \{\Omega_m, \Omega_b, h, n_s, \sigma_8\}$, used to generate that forward model.

The CNN architecture consists of 5 convolutional blocks and 3 fully-connected blocks. Additionally, we include significant levels of dropout and ℓ_2 regularization to ensure robustness and generalizability, and thus protect against the fact that the SIMBIG forward models, and in general any forward model, are approximate.

In order to further prevent the CNN from overfitting on the training set, we perform a weight marginalization step, converting our CNN into a Bayesian Neural Network. Specifically, we use Stochastic Weight Averaging (SWA) (Maddox et al., 2019; Wilson & Izmailov, 2020). SWA approximates the posterior distribution of the weights of the CNN as a Normal distribution, whose mean and covariance are calculated from the weights of the network during a number of previous optimization steps. By adopting this scheme, SWA solutions tend to converge to the center of flat loss regions, thereby leading to more stable and generalizeable solutions (Wilson & Izmailov, 2020). Moreover, SWA has already been applied to both astrophysics (Cranmer et al., 2021) and cosmology (Lemos et al., 2023).

The compressed data that we feed as input to SBI is the output of the SWA network: a set of 10 samples of the posterior distribution weights of the CNN — a 50-dimensional

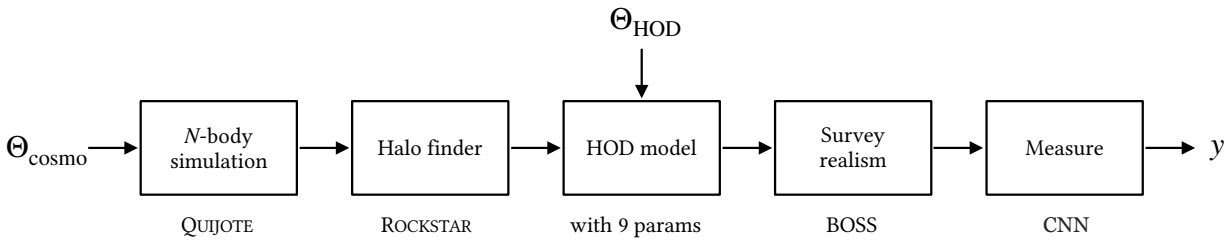


Figure 1. Schematic illustrating the various elements of the SIMBIG forward modelling pipeline, which generates synthetic galaxy catalogs that mimic the real BOSS observations. Given cosmological parameters $\Theta_{\rm cosmo}$, we first run an N-body simulation. We then find halos using ROCKSTAR. We build a galaxy catalog with our HOD presciption, which adds 9 more parameters sampled from a prior distribution. We add survey realism to mimic the BOSS observational data. Finally, our CNN performs a step of data compression.

data vector. For more details on the specifics of the CNN architecture, see Appendix B.

2.3. Simulation-Based Inference

After training the CNN, we use the SIMBIG SBI framework to estimate posterior distributions of the cosmological parameters, $\boldsymbol{\theta}$, from the compressed representation of the observables obtained from the CNN, $\hat{\boldsymbol{\theta}}$. We represent this posterior as $p(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$. To that end, we use Neural Posterior Estimation (NPE) (Rezende & Mohamed, 2015; Papamakarios & Murray, 2016; Lueckmann et al., 2018; Lueckmann et al., 2017; Greenberg et al., 2019). NPE uses a density estimator to learn an approximation to the posterior distribution. In this case, the training set consists of the ground-truth/CNN-compressed $\{\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\}$ parameter pairs of the SIMBIG forward models. We use the publicly available SBI implementation from Tejero-Cantero et al. (2020). In previous SIMBIG analyses, the authors employed a Masked Autoregressive Flow (MAF; Papamakarios et al., 2017) as their density estimator. However, for our density estimator, we instead use Neural Spline Flows (NSF; Durkan et al., 2019), a more expressive alternative which have been shown to outperform other methods such as Mixture Density Networks (MDN) or MAFs. Denoting our NSF as $q_\phi(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, where ϕ represents the hyperparameters, we train q_ϕ by minimizing the KL divergence between $p(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$ and $q_\phi(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})p(\hat{\boldsymbol{\theta}})$. This is equivalent to maximizing the log-likelihood over the training set of SIMBIG forward models. For a more robust inference, we use an ensemble of five NSFs (Lakshminarayanan et al., 2017), which have been shown to produce more reliable approximations (Hermans et al., 2022).

2.4. Validation

Before applying our analysis to observations, we validate our NDE on the suite of test simulations described in Section 2.1. Specifically, we follow Hahn & Melchior (2022) and assess robustness by comparing the likelihoods over the three test sets. We do this by computing the posterior mean μ and standard deviation σ for each parameter and each test simulation, and then studying the difference between μ and the true parameter value $\theta^{\rm fid}$ in units of σ. For a robust pipeline, we expect to find consistency of these estimates across all three datasets. On the other hand, variations between the distributions across the data sets would be indicative of likelihood variations as we change the forward model.

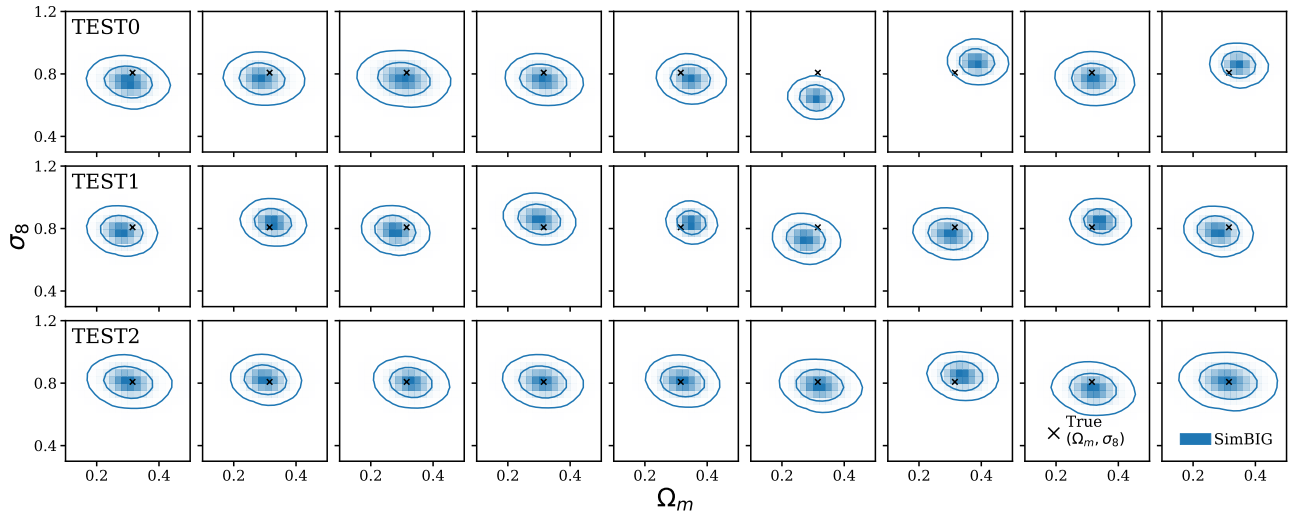
3. Results

3.1. Validation Tests

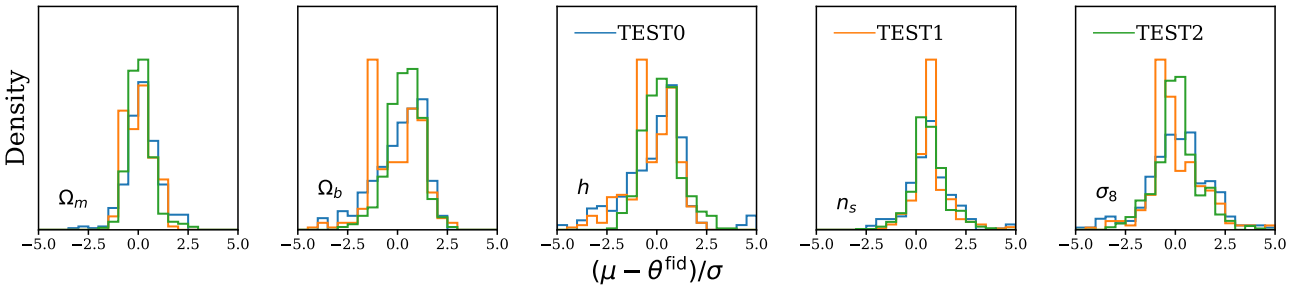
Before applying our NDE to the observational data, we validate our pipeline on the SIMBIG mock challenge data. Figure 2 (a) shows the marginalized two-dimensional posterior distribution in $\{\Omega_m, \sigma_8\}$ for 10 randomly selected simulations from each of the test sets. We see how in all three test sets, the posteriors appear to be well calibrated and unbiased. Figure 2 (b) shows a more quantitative metric, by plotting an estimated difference between the likelihoods of the three forward models, as described in Section 2.

The fact that Figure 2 (b) shows similar distributions for all parameters across test sets indicates that our posterior inference is robust to these variations in the forward model. This is likely helped by the use of weight marginalization, which leads to better generalization properties.

These validation tests form a crucial part of our analysis. We note that it is possible to obtain significantly tighter constraints that pass only validation tests on only `TEST0` or the training data. However, in doing so, we would need to assume that our forward model accurately models every aspect of the observations. Given the complexities of galaxy formation, *any* forward model of galaxy clustering is an approximate model. Hence, validating that we can successfully infer unbias cosmological constraints from simulated test galaxy catalogsgenerated with different forward models



(a) Posterior distributions for (Ω_m, σ_8). Each row shows 10 randomly selected examples taken from each test set, as labeled. True parameters are marked in orange, and the contours, from the outer to the inner ones, cover an estimated probability mass of 90% and 50%, respectively.



(b) Distributions of the differences between the posterior mean μ and the true parameter $\boldsymbol{\theta}^{\rm fid}$ normalized by the posterior standard deviation σ, for each cosmological parameter and test set.

Figure 2. Validation of our neural density estimator on the SIMBIG mock challenge data.

(TEST1 and TEST2) serves a powerful test against model misspecification, even if it come at the expense of significant constraining power.

3.2. Parameter Constraints

In Figure 3, we present the marginalized posterior distributions of the ΛCDM cosmological parameters, Ω_m and σ_8. We present the CNN constraints (orange) and compare them to those obtained from the SIMBIG P_ℓ analysis (grey; Hahn et al.). We also include constraints from the PT based P_ℓ analysis of the CMASS SGC sample (dashed; Ivanov et al., 2020).

Our field-level analysis using the CNN provides tighter, yet consistent, constraints on σ_8 and Ω_m to the previous BOSS analyses. Specifically, our constraints on Ω_m and σ_8 are $1.76\times$ and $1.92\times$ tighter than the SIMBIG P_ℓ analysis and $1.03\times$ and $2.65\times$ tighter than the P_ℓ. This higher constraining power is expected. Indeed, in using the CNN to compress the data, our posterior estimator is able to exploit non-Gaussian cosmological information on non-linear scales that is inaccessible to P_ℓ analyses. Moreover, in using the SIMBIG SBI approach, we are able to more robustly account for observational systematics compared to the standard clustering analyses.

With our field-level analysis, we also can place significant constraints on $H_0 = 63.1 \pm 4.1$ km/s/Mpc, albeit weaker than those on Ω_m and σ_8. This is in contrast to standard P_ℓ analyses, which cannot independently constrain H_0 and typically rely on priors from Big Bang Nucleosynthesis or CMB experiments. Our constraints support a low value of H_0, although we do not have not have enough constraining power to make strong statements. We also achieve good agreement with *Planck*. We will further investigate the consequences of this result, and how they compare with

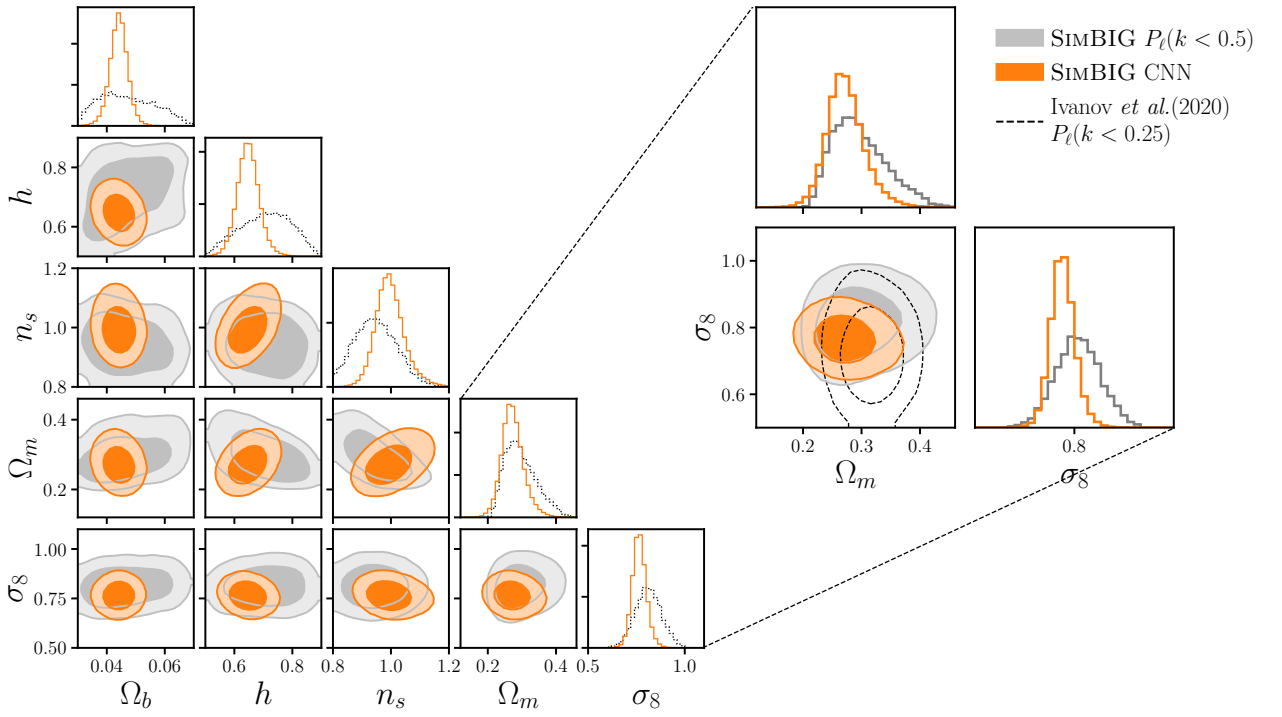


Figure 3. Left: Posterior distributions for all ΛCDM cosmological parameters from our CNN-based field level inference of BOSS observations (orange). For comparison, we include the SIMBIG P_ℓ analysis (gray). The contours represent the 68% and 95% confidence intervals. *Right*: Posterior distributions for Ω_m and σ_8. For comparison, we include posteriors from the SIMBIG P_ℓ analysis (gray) and the standard PT-based P_ℓ analysis (black dashed; Ivanov et al., 2020)

other surveys and cosmological probes in (Collaboration, 2023).

4. Conclusions

In this paper, we have presented constraints from field-level CMASS galaxy catalogs using a SBI approach. We have shown that our inference pipeline passes a number of stringent validation tests, including using a different halo finding algorithm and simulation code. These tests provide key validation against the main concern of this type of analysis, model misspecification. We have also shown that our method provides tighter constraints on the cosmological parameters than methods based on compressing the data to the power spectrum, both in likelihood-based and simulation-bases analysis. However, our constraints remain consistent.

As simulations become more realistic and efficient in the future, we will be able to extend our analyses to smaller scales an the larger volumes covered by upcoming surveys such as the Dark Energy Spectroscopic Instrument (DESI; Collaboration et al., 2016a;b; Abareshi et al., 2022), Subaru Prime Focus Spectrograph (PFS; Takada et al., 2014; Tamura et al., 2016), the ESA *Euclid* satellite mission (Laureijs et al., 2011), and the Nancy Grace Roman Space Telescope (Roman; Spergel et al., 2015; Wang et al., 2022). Our results demonstrate that these analyses will be able to produce leading cosmological constraints from galaxy clustering. The methodology and tests presented in this paper lays the groundwork for such analyses.

Acknowledgements

It is a pleasure to thank Mikhail M. Ivanov for providing us with the posteriors used for comparison, and Ben Wandelt for discussions that greatly helped the papers. We thanks the Learning the Universe Collaboration for helpful feedback and stimulating discussions. PL acknowledges support from the Simons Foundation. JH has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101025187. AMD acknowledges funding from Tomalla Foundation for Research in Gravity.

References

Abareshi, B., Aguilar, J., Ahlen, S., Alam, S., Alexander, D. M., Alfarsy, R., Allen, L., Prieto, C. A., Alves, O., Ameel, J., Armengaud, E., Asorey, J., Aviles, A., Bailey, S., Balaguera-Antolínez, A., Ballester, O., Baltay, C., Bault, A., Beltran, S. F., Benavides, B., BenZvi, S., Berti, A., Besuner, R., Beutler, F., Bianchi, D., Blake, C., Blanc, P., Blum, R., Bolton, A., Bose, S., Bramall, D., Brieden, S., Brodzeller, A., Brooks, D., Brownewell, C., Buckley-Geer, E., Cahn, R. N., Cai, Z., Canning, R., Rosell, A. C., Carton, P., Casas, R., Castander, F. J., Cervantes-Cota, J. L., Chabanier, S., Chaussidon, E., Chuang, C., Circosta, C., Cole, S., Cooper, A. P., da Costa, L., Cousinou, M.-C., Cuceu, A., Davis, T. M., Dawson, K., de la Cruz-Noriega, R., de la Macorra, A., de Mattia, A., Della Costa, J., Demmer, P., Derwent, M., Dey, A., Dey, B., Dhungana, G., Ding, Z., Dobson, C., Doel, P., Donald-McCann, J., Donaldson, J., Douglass, K., Duan, Y., Dunlop, P., Edelstein, J., Eftekharzadeh, S., Eisenstein, D. J., Enriquez-Vargas, M., Escoffier, S., Evatt, M., Fagrelius, P., Fan, X., Fanning, K., Fawcett, V. A., Ferraro, S., Ereza, J., Flaugher, B., Font-Ribera, A., Forero-Romero, J. E., Frenk, C. S., Fromenteau, S., Gänsicke, B. T., Garcia-Quintero, C., Garrison, L., Gaztañaga, E., Gerardi, F., Gil-Marín, H., Gontcho, S. G. A., Gonzalez-Morales, A. X., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gordon, C., Graur, O., Green, D., Grove, C., Gruen, D., Gutierrez, G., Guy, J., Hahn, C., Harris, S., Herrera, D., Herrera-Alcantar, H. K., Honscheid, K., Howlett, C., Huterer, D., Iršič, V., Ishak, M., Jelinsky, P., Jiang, L., Jimenez, J., Jing, Y. P., Joyce, R., Jullo, E., Juneau, S., Karaçaylı, N. G., Karamanis, M., Karcher, A., Karim, T., Kehoe, R., Kent, S., Kirkby, D., Kisner, T., Kitaura, F., Koposov, S. E., Kovács, A., Kremin, A., Krolewski, A., L'Huillier, B., Lahav, O., Lambert, A., Lamman, C., Lan, T.-W., Landriau, M., Lane, S., Lang, D., Lange, J. U., Lasker, J., Guillou, L. L., Leauthaud, A., Van Suu, A. L., Levi, M. E., Li, T. S., Magneville, C., Manera, M., Manser, C. J., Marshall, B., McCollam, W., McDonald, P., Meisner, A. M., Mezcua, J. M.-F. M., Miller, T., Miquel, R., Montero-Camacho, P., Moon, J., Martini, J. P., Meneses-Rizo, J., Moustakas, J., Mueller, E., Muñoz-Gutiérrez, A., Myers, A. D., Nadathur, S., Najita, J., Napolitano, L., Neilsen, E., Newman, J. A., Nie, J. D., Ning, Y., Niz, G., Norberg, P., Noriega, H. E., O'Brien, T., Obuljen, A., Palanque-Delabrouille, N., Palmese, A., Zhiwei, P., Pappalardo, D., Peng, X., Percival, W. J., Perruchot, S., Pogge, R., Poppett, C., Porredon, A., Prada, F., Prochaska, J., Pucha, R., Pérez-Fernández, A., Pérez-Ráfols, I., Rabinowitz, D., Raichoor, A., Ramirez-Solano, S., Ramírez-Pérez, C., Ravoux, C., Reil, K., Rezaie, M., Rocher, A., Rockosi, C., Roe, N. A., Roodman, A., Ross, A. J., Rossi, G., Ruggeri, R., Ruhlmann-Kleider, V., Sabiu, C. G., Safonova, S., Said, K., Saintonge, A., Catonga, J. S., Samushia, L., Sanchez, E., Saulder, C., Schaan, E., Schlafly, E., Schlegel, D., Schmoll, J., Scholte, D., Schubnell, M., Secroun, A., Seo, H., Serrano, S., Sharples, R. M., Sholl, M. J., Silber, J. H., Silva, D. R., Sirk, M., Siudek, M., Smith, A., Sprayberry, D., Staten, R., Stupak, B., Tan, T., Tarlé, G., Tie, S. S., Tojeiro, R., Ureña-López, L. A., Valdes, F., Valenzuela, O., Valluri, M., Vargas-Magaña, M., Verde, L., Walther, M., Wang, B., Wang, M. S., Weaver, B. A., Weaverdyck, C., Wechsler, R., Wilson, M. J., Yang, J., Yu, Y., Yuan, S., Yèche, C., Zhang, H., Zhang, K., Zhao, C., Zhou, R., Zhou, Z., Zou, H., Zou, J., Zou, S., and Zu, Y. Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument, May 2022.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.

Anderson, L., Aubourg, E., Bailey, S., Bizyaev, D., Blanton, M., Bolton, A. S., Brinkmann, J., Brownstein, J. R., Burden, A., Cuesta, A. J., et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 427(4):3435–3467, 2012.

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(2):109, 2012a.

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., Busha, M. T., Klypin, A. A., and Primack, J. R. Gravitationally consistent halo catalogs and merger trees for precision cosmology. *The Astrophysical Journal*, 763(1):18, 2012b.

Bernardeau, F., Colombi, S., Gaztañaga, E., and Scoccimarro, R. Large-scale structure of the universe and cosmological perturbation theory. *Physics Reports*, 367(1-3):1–248, sep 2002. doi: 10.1016/s0370-1573(02)00135-7. URL https://doi.org/10.1016%2Fs0370-1573%2802%2900135-7.

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.

Birdsall, C. K. and Fuss, D. Clouds-in-clouds, clouds-in-cells physics for many-body plasma simulation. *Journal of Computational Physics*, 3(4):494–511, 1969.

Carlson, J. and White, M. Embedding realistic surveys in simulations through volume remapping. *The Astrophysical Journal Supplement Series*, 190(2):311, 2010.

Collaboration, D., Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Prieto, C. A., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., Blake, C., Bleuler, H., Blomqvist, M., Blum, R., Bolton, A. S., Briceno, C., Brooks, D., Brownstein, J. R., Buckley-Geer, E., Burden, A., Burtin, E., Busca, N. G., Cahn, R. N., Cai, Y.-C., Cardiel-Sas, L., Carlberg, R. G., Carton, P.-H., Casas, R., Castander, F. J., Cervantes-Cota, J. L., Claybaugh, T. M., Close, M., Coker, C. T., Cole, S., Comparat, J., Cooper, A. P., Cousinou, M.-C., Crocce, M., Cuby, J.-G., Cunningham, D. P., Davis, T. M., Dawson, K. S., de la Macorra, A., De Vicente, J., Delubac, T., Derwent, M., Dey, A., Dhungana, G., Ding, Z., Doel, P., Duan, Y. T., Ealet, A., Edelstein, J., Eftekharzadeh, S., Eisenstein, D. J., Elliott, A., Escoffier, S., Evatt, M., Fagrelius, P., Fan, X., Fanning, K., Farahi, A., Farihi, J., Favole, G., Feng, Y., Fernandez, E., Findlay, J. R., Finkbeiner, D. P., Fitzpatrick, M. J., Flaugher, B., Flender, S., Font-Ribera, A., Forero-Romero, J. E., Fosalba, P., Frenk, C. S., Fumagalli, M., Gaensicke, B. T., Gallo, G., Garcia-Bellido, J., Gaztanaga, E., Fusillo, N. P. G., Gerard, T., Gershkovich, I., Giannantonio, T., Gillet, D., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gott, S., Graur, O., Gutierrez, G., Guy, J., Habib, S., Heetderks, H., Heetderks, I., Heitmann, K., Hellwing, W. A., Herrera, D. A., Ho, S., Holland, S., Honscheid, K., Huff, E., Hutchinson, T. A., Huterer, D., Hwang, H. S., Laguna, J. M. I., Ishikawa, Y., Jacobs, D., Jeffrey, N., Jelinsky, P., Jennings, E., Jiang, L., Jimenez, J., Johnson, J., Joyce, R., Jullo, E., Juneau, S., Kama, S., Karcher, A., Karkar, S., Kehoe, R., Kennamer, N., Kent, S., Kilbinger, M., Kim, A. G., Kirkby, D., Kisner, T., Kitanidis, E., Kneib, J.-P., Koposov, S., Kovacs, E., Koyama, K., Kremin, A., Kron, R., Kronig, L., Kueter-Young, A., Lacey, C. G., Lafever, R., Lahav, O., Lambert, A., Lampton, M., Landriau, M., Lang, D., Lauer, T. R., Goff, J.-M. L., Guillou, L. L., Van Suu, A. L., Lee, J. H., Lee, S.-J., Leitner, D., Lesser, M., Levi, M. E., L'Huillier, B., Li, B., Liang, M., Lin, H., Linder, E., Loebman, S. R., Lukić, Z., Ma, J., MacCrann, N., Magneville, C., Makarem, L., Manera, M., Manser, C. J., Marshall, R., Martini, P., Massey, R., Matheson, T., McCauley, J., McDonald, P., McGreer, I. D., Meisner, A., Metcalfe, N., Miller, T. N., Miquel, R., Moustakas, J., Myers, A., Naik, M., Newman, J. A., Nichol, R. C., Nicola, A., da Costa, L. N., Nie, J., Niz, G., Norberg, P., Nord, B., Norman, D., Nugent, P., O'Brien, T., Oh, M., Olsen, K. A. G., Padilla, C., Padmanabhan, H., Padmanabhan, N., Palanque-Delabrouille, N., Palmese, A., Pappalardo, D., Pâris, I., Park, C., Patej, A., Peacock, J. A., Peiris, H. V., Peng, X., Percival, W. J., Perruchot, S., Pieri, M. M., Pogge, R., Pollack, J. E., Poppett, C., Prada, F., Prakash, A., Probst, R. G., Rabinowitz, D., Raichoor, A., Ree, C. H., Refregier, A., Regal, X., Reid, B., Reil, K., Rezaie, M., Rockosi, C. M., Roe, N., Ronayette, S., Roodman, A., Ross, A. J., Ross, N. P., Rossi, G., Rozo, E., Ruhlmann-Kleider, V., Rykoff, E. S., Sabiu, C., Samushia, L., Sanchez, E., Sanchez, J., Schlegel, D. J., Schneider, M., Schubnell, M., Secroun, A., Seljak, U., Seo, H.-J., Serrano, S., Shafieloo, A., Shan, H., Sharples, R., Sholl, M. J., Shourt, W. V., Silber, J. H., Silva, D. R., Sirk, M. M., Slosar, A., Smith, A., Smoot, G. F., Som, D., Song, Y.-S., Sprayberry, D., Staten, R., Stefanik, A., Tarle, G., Tie, S. S., Tinker, J. L., Tojeiro, R., Valdes, F., Valenzuela, O., Valluri, M., Vargas-Magana, M., Verde, L., Walker, A. R., Wang, J., Wang, Y., Weaver, B. A., Weaverdyck, C., Wechsler, R. H., Weinberg, D. H., White, M., Yang, Q., Yeche, C., Zhang, T., Zhao, G.-B., Zheng, Y., Zhou, X., Zhou, Z., Zhu, Y., Zou, H., and Zu, Y. The DESI Experiment Part I: Science,Targeting, and Survey Design. *arXiv:1611.00036 [astro-ph]*, October 2016a.

Collaboration, D., Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Prieto, C. A., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., Blake, C., Bleuler, H., Blomqvist, M., Blum, R., Bolton, A. S., Briceno, C., Brooks, D., Brownstein, J. R., Buckley-Geer, E., Burden, A., Burtin, E., Busca, N. G., Cahn, R. N., Cai, Y.-C., Cardiel-Sas, L., Carlberg, R. G., Carton, P.-H., Casas, R., Castander, F. J., Cervantes-Cota, J. L., Claybaugh, T. M., Close, M., Coker, C. T., Cole, S., Comparat, J., Cooper, A. P., Cousinou, M.-C., Crocce, M., Cuby, J.-G., Cunningham, D. P., Davis, T. M., Dawson, K. S., de la Macorra, A., De Vicente, J., Delubac, T., Derwent, M., Dey, A., Dhungana, G., Ding, Z., Doel, P., Duan, Y. T., Ealet, A., Edelstein, J., Eftekharzadeh, S., Eisenstein, D. J., Elliott, A., Escoffier, S., Evatt, M., Fagrelius, P., Fan, X., Fanning, K., Farahi, A., Farihi, J., Favole, G., Feng, Y., Fernandez, E., Findlay, J. R., Finkbeiner, D. P., Fitzpatrick, M. J., Flaugher, B., Flender, S., Font-Ribera, A., Forero-Romero, J. E., Fosalba, P., Frenk, C. S., Fumagalli, M., Gaensicke, B. T.,

Gallo, G., Garcia-Bellido, J., Gaztanaga, E., Fusillo, N. P. G., Gerard, T., Gershkovich, I., Giannantonio, T., Gillet, D., Gonzalez-de-Rivera, G., Gonzalez-Perez, V., Gott, S., Graur, O., Gutierrez, G., Guy, J., Habib, S., Heetderks, H., Heetderks, I., Heitmann, K., Hellwing, W. A., Herrera, D. A., Ho, S., Holland, S., Honscheid, K., Huff, E., Hutchinson, T. A., Huterer, D., Hwang, H. S., Laguna, J. M. I., Ishikawa, Y., Jacobs, D., Jeffrey, N., Jelinsky, P., Jennings, E., Jiang, L., Jimenez, J., Johnson, J., Joyce, R., Jullo, E., Juneau, S., Kama, S., Karcher, A., Karkar, S., Kehoe, R., Kennamer, N., Kent, S., Kilbinger, M., Kim, A. G., Kirkby, D., Kisner, T., Kitanidis, E., Kneib, J.-P., Koposov, S., Kovacs, E., Koyama, K., Kremin, A., Kron, R., Kronig, L., Kueter-Young, A., Lacey, C. G., Lafever, R., Lahav, O., Lambert, A., Lampton, M., Landriau, M., Lang, D., Lauer, T. R., Goff, J.-M. L., Guillou, L. L., Van Suu, A. L., Lee, J. H., Lee, S.-J., Leitner, D., Lesser, M., Levi, M. E., L'Huillier, B., Li, B., Liang, M., Lin, H., Linder, E., Loebman, S. R., Lukić, Z., Ma, J., MacCrann, N., Magneville, C., Makarem, L., Manera, M., Manser, C. J., Marshall, R., Martini, P., Massey, R., Matheson, T., McCauley, J., McDonald, P., McGreer, I. D., Meisner, A., Metcalfe, N., Miller, T. N., Miquel, R., Moustakas, J., Myers, A., Naik, M., Newman, J. A., Nichol, R. C., Nicola, A., da Costa, L. N., Nie, J., Niz, G., Norberg, P., Nord, B., Norman, D., Nugent, P., O'Brien, T., Oh, M., Olsen, K. A. G., Padilla, C., Padmanabhan, H., Padmanabhan, N., Palanque-Delabrouille, N., Palmese, A., Pappalardo, D., Pâris, I., Park, C., Patej, A., Peacock, J. A., Peiris, H. V., Peng, X., Percival, W. J., Perruchot, S., Pieri, M. M., Pogge, R., Pollack, J. E., Poppett, C., Prada, F., Prakash, A., Probst, R. G., Rabinowitz, D., Raichoor, A., Ree, C. H., Refregier, A., Regal, X., Reid, B., Reil, K., Rezaie, M., Rockosi, C. M., Roe, N., Ronayette, S., Roodman, A., Ross, A. J., Ross, N. P., Rossi, G., Rozo, E., Ruhlmann-Kleider, V., Rykoff, E. S., Sabiu, C., Samushia, L., Sanchez, E., Sanchez, J., Schlegel, D. J., Schneider, M., Schubnell, M., Secroun, A., Seljak, U., Seo, H.-J., Serrano, S., Shafieloo, A., Shan, H., Sharples, R., Sholl, M. J., Shourt, W. V., Silber, J. H., Silva, D. R., Sirk, M. M., Slosar, A., Smith, A., Smoot, G. F., Som, D., Song, Y.-S., Sprayberry, D., Staten, R., Stefanik, A., Tarle, G., Tie, S. S., Tinker, J. L., Tojeiro, R., Valdes, F., Valenzuela, O., Valluri, M., Vargas-Magana, M., Verde, L., Walker, A. R., Wang, J., Wang, Y., Weaver, B. A., Weaverdyck, C., Wechsler, R. H., Weinberg, D. H., White, M., Yang, Q., Yeche, C., Zhang, T., Zhao, G.-B., Zheng, Y., Zhou, X., Zhou, Z., Zhu, Y., Zou, H., and Zu, Y. The DESI Experiment Part II: Instrument Design. *arXiv:1611.00037 [astro-ph]*, October 2016b.

Collaboration, S. Simbig: Comparison of different observables with external surveys. *In preparation*, 2023.

Cranmer, M., Tamayo, D., Rein, H., Battaglia, P., Hadden, S., Armitage, P. J., Ho, S., and Spergel, D. N. A bayesian neural network predicts the dissolution of compact planetary systems. *Proceedings of the National Academy of Sciences*, 118(40):e2026053118, 2021.

D'Amico, G., Donath, Y., Lewandowski, M., Senatore, L., and Zhang, P. The BOSS bispectrum analysis at one loop from the Effective Field Theory of Large-Scale Structure, June 2022.

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal*, 292: 371–394, 1985.

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., et al. The baryon oscillation spectroscopic survey of sdss-iii. *The Astronomical Journal*, 145(1):10, 2012.

Desjacques, V., Jeong, D., and Schmidt, F. Large-scale galaxy bias. *Physics Reports*, 733:1–193, feb 2018. doi: 10.1016/j.physrep.2017.12.002. URL https://doi.org/10.1016%2Fj.physrep.2017.12.002.

Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. Neural spline flows. *Advances in neural information processing systems*, 32, 2019.

Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Prieto, C. A., Anderson, S. F., Arns, J. A., Aubourg, É., Bailey, S., Balbinot, E., et al. Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *The Astronomical Journal*, 142(3):72, 2011.

Gal, Y. and Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016.

Graves, A. Practical variational inference for neural networks. *Advances in neural information processing systems*, 24, 2011.

Greenberg, D., Nonnenmacher, M., and Macke, J. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pp. 2404–2414. PMLR, 2019.

Hadzhiyska, B., Liu, S., Somerville, R. S., Gabrielpillai, A., Bose, S., Eisenstein, D., and Hernquist, L. Galaxy assembly bias and large-scale distribution: a comparison between illustristng and a semi-analytic model. *Monthly Notices of the Royal Astronomical Society*, 508(1):698–718, 2021.

Hadzhiyska, B., Eisenstein, D., Bose, S., Garrison, L. H., and Maksimova, N. compaso: A new halo finder for competitive assignment to spherical overdensities. *Monthly Notices of the Royal Astronomical Society*, 509(1):501–521, 2022a.

Hadzhiyska, B., Eisenstein, D., Hernquist, L., Pakmor, R., Bose, S., Delgado, A. M., Contreras, S., Kannan, R., White, S. D. M., Springel, V., Frenk, C., Hernández-Aguayo, C., Ferlito, F., and Barrera, M. The millenniumtng project: An improved two-halo model for the galaxy-halo connection of red and blue galaxies, 2022b. URL https://arxiv.org/abs/2210.10072.

Hahn, C. and Melchior, P. Accelerated Bayesian SED Modeling using Amortized Neural Posterior Estimation, March 2022.

Hahn, C., Abidi, M., Eickenberg, M., Ho, S., Lemos, P., Massara, E., Dizgah, A. M., and Régaldo-Saint Blancard, B. Simbig: Likelihood-free inference of galaxy clustering.

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 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx185.

Hahn, C., Eickenberg, M., Ho, S., Hou, J., Lemos, P., Massara, E., Modi, C., Dizgah, A. M., Blancard, B. R.-S., and Abidi, M. M. Simbig: A forward modeling approach to analyzing galaxy clustering. *arXiv preprint arXiv:2211.00660*, 2022.

Hahn, C., Eickenberg, M., Ho, S., Hou, J., Lemos, P., Massara, E., Modi, C., Moradinezhad Dizgah, A., Régaldo-Saint Blancard, B., and Abidi, M. M. SIMBIG: mock challenge for a forward modeling approach to galaxy clustering. , 2023(4):010, April 2023. doi: 10.1088/1475-7516/2023/04/010.

Hermans, J., Delaunoy, A., Rozet, F., Wehenkel, A., Begy, V., and Louppe, G. A trust crisis in simulation-based inference? your posterior approximations can be unfaithful. *stat*, 1050:4, 2022.

Ivanov, M. M., Simonović, M., and Zaldarriaga, M. Cosmological parameters and neutrino masses from the final p l a n c k and full-shape boss data. *Physical Review D*, 101 (8):083504, 2020.

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(1):954–969, 2021.

Kobayashi, Y., Nishimichi, T., Takada, M., and Miyatake, H. Full-shape cosmology analysis of the SDSS-III BOSS galaxy power spectrum using an emulator-based halo model: A 5% determination of σ_8. , 105(8):083517, April 2022. doi: 10.1103/PhysRevD.105.083517.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J.-L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., Pasian, F., Percival, W., Rhodes, J., Saavedra Criado, G., Sauvage, M., Scaramella, R., Valenziano, L., Warren, S., Bender, R., Castander, F., Cimatti, A., Le Fèvre, O., Kurki-Suonio, H., Levi, M., Lilje, P., Meylan, G., Nichol, R., Pedersen, K., Popa, V., Rebolo Lopez, R., Rix, H.-W., Rottgering, H., Zeilinger, W., Grupp, F., Hudelot, P., Massey, R., Meneghetti, M., Miller, L., Paltani, S., Paulin-Henriksson, S., Pires, S., Saxton, C., Schrabback, T., Seidel, G., Walsh, J., Aghanim, N., Amendola, L., Bartlett, J., Baccigalupi, C., Beaulieu, J.-P., Benabed, K., Cuby, J.-G., Elbaz, D., Fosalba, P., Gavazzi, G., Helmi, A., Hook, I., Irwin, M., Kneib, J.-P., Kunz, M., Mannucci, F., Moscardini, L., Tao, C., Teyssier, R., Weller, J., Zamorani, G., Zapatero Osorio, M. R., Boulade, O., Foumond, J. J., Di Giorgio, A., Guttridge, P., James, A., Kemp, M., Martignac, J., Spencer, A., Walton, D., Blümchen, T., Bonoli, C., Bortoletto, F., Cerna, C., Corcione, L., Fabron, C., Jahnke, K., Ligori, S., Madrid, F., Martin, L., Morgante, G., Pamplona, T., Prieto, E., Riva, M., Toledo, R., Trifoglio, M., Zerbi, F., Abdalla, F., Douspis, M., Grenet, C., Borgani, S., Bouwens, R., Courbin, F., Delouis, J.-M., Dubath, P., Fontana, A., Frailis, M., Grazian, A., Koppenhöfer, J., Mansutti, O., Melchior, M., Mignoli, M., Mohr, J., Neissner, C., Noddle, K., Poncet, M., Scodeggio, M., Serrano, S., Shane, N., Starck, J.-L., Surace, C., Taylor, A., Verdoes-Kleijn, G., Vuerli, C., Williams, O. R., Zacchei, A., Altieri, B., Escudero Sanz, I., Kohley, R., Oosterbroek, T., Astier, P., Bacon, D., Bardelli, S., Baugh, C., Bellagamba, F., Benoist, C., Bianchi, D., Biviano, A., Branchini, E., Carbone, C., Cardone, V., Clements, D., Colombi, S., Conselice, C., Cresci, G., Deacon, N., Dunlop, J., Fedeli, C., Fontanot, F., Franzetti, P., Giocoli, C., Garcia-Bellido, J., Gow, J., Heavens, A., Hewett, P., Heymans, C., Holland, A., Huang, Z., Ilbert, O., Joachimi, B., Jennins, E., Kerins, E., Kiessling, A., Kirk, D., Kotak, R., Krause,

O., Lahav, O., van Leeuwen, F., Lesgourgues, J., Lombardi, M., Magliocchetti, M., Maguire, K., Majerotto, E., Maoli, R., Marulli, F., Maurogordato, S., McCracken, H., McLure, R., Melchiorri, A., Merson, A., Moresco, M., Nonino, M., Norberg, P., Peacock, J., Pello, R., Penny, M., Pettorino, V., Di Porto, C., Pozzetti, L., Quercellini, C., Radovich, M., Rassat, A., Roche, N., Ronayette, S., Rossetti, E., Sartoris, B., Schneider, P., Semboloni, E., Serjeant, S., Simpson, F., Skordis, C., Smadja, G., Smartt, S., Spano, P., Spiro, S., Sullivan, M., Tilquin, A., Trotta, R., Verde, L., Wang, Y., Williger, G., Zhao, G., Zoubian, J., and Zucca, E. Euclid Definition Study Report. *arXiv e-prints*, pp. arXiv:1110.3193, October 2011.

Lemos, P., Cranmer, M., Abidi, M., Hahn, C., Eickenberg, M., Massara, E., Yallup, D., and Ho, S. Robust simulation-based inference in cosmology with bayesian neural networks. *Machine Learning: Science and Technology*, 4(1):01LT01, 2023.

Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., and Macke, J. H. Flexible statistical inference for mechanistic models of neural dynamics. *Advances in neural information processing systems*, 30, 2017.

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

Maddox, W. J., Izmailov, P., Garipov, T., Vetrov, D. P., and Wilson, A. G. A simple baseline for bayesian uncertainty in deep learning. *Advances in neural information processing systems*, 32, 2019.

Maksimova, N. A., Garrison, L. H., Eisenstein, D. J., Hadzhiyska, B., Bose, S., and Satterthwaite, T. P. Abacussummit: a massive set of high-accuracy, high-resolution n-body simulations. *Monthly Notices of the Royal Astronomical Society*, 508(3):4017–4037, 2021.

More, S., Miyatake, H., Takada, M., Diemer, B., Kravtsov, A. V., Dalal, N. K., More, A., Murata, R., Mandelbaum, R., Rozo, E., et al. Detection of the splashback radius and halo assembly bias of massive galaxy clusters. *The Astrophysical Journal*, 825(1):39, 2016.

Papamakarios, G. and Murray, I. Fast ε-free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016.

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

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Philcox, O. H. and Ivanov, M. M. Boss dr12 full-shape cosmology: λ cdm constraints from the large-scale galaxy power spectrum and bispectrum monopole. *Physical Review D*, 105(4):043517, 2022.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Ross, A. J., Percival, W. J., Sánchez, A. G., Samushia, L., Ho, S., Kazin, E., Manera, M., Reid, B., White, M., Tojeiro, R., et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: analysis of potential systematics. *Monthly Notices of the Royal Astronomical Society*, 424(1):564–590, 2012.

Santurkar, S., Tsipras, D., Ilyas, A., and Madry, A. How does batch normalization help optimization? *Advances in neural information processing systems*, 31, 2018.

Smith, L. N. and Topin, N. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, volume 11006, pp. 369–386. SPIE, 2019.

Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Macintosh, B., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., Wang, Y., Weinberg, D., Benford, D., Hudson, M., Jeong, W. S., Mellier, Y., Traub, W., Yamada, T., Capak, P., Colbert, J., Masters, D., Penny, M., Savransky, D., Stern, D., Zimmerman, N., Barry, R., Bartusek, L., Carpenter, K., Cheng, E., Content, D., Dekens, F., Demers, R., Grady, K., Jackson, C., Kuan, G., Kruk, J., Melton, M., Nemati, B., Parvin, B., Poberezhskiy, I., Peddie, C., Ruffa, J., Wallace, J. K., Whipple, A., Wollack, E., and Zhao, F. Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, March 2015.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Takada, M., Ellis, R. S., Chiba, M., Greene, J. E., Aihara, H., Arimoto, N., Bundy, K., Cohen, J., Doré, O., Graves, G., Gunn, J. E., Heckman, T., Hirata, C. M., Ho, P., Kneib,

J.-P., Le Fèvre, O., Lin, L., More, S., Murayama, H., Nagao, T., Ouchi, M., Seiffert, M., Silverman, J. D., Sodré, L., Spergel, D. N., Strauss, M. A., Sugai, H., Suto, Y., Takami, H., and Wyse, R. Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph. *Publications of the Astronomical Society of Japan*, 66:R1, February 2014. ISSN 0004-6264. doi: 10.1093/pasj/pst019.

Tamura, N., Takato, N., Shimono, A., Moritani, Y., Yabe, K., Ishizuka, Y., Ueda, A., Kamata, Y., Aghazarian, H., Arnouts, S., Barban, G., Barkhouser, R. H., Borges, R. C., Braun, D. F., Carr, M. A., Chabaud, P.-Y., Chang, Y.-C., Chen, H.-Y., Chiba, M., Chou, R. C. Y., Chu, Y.-H., Cohen, J., de Almeida, R. P., de Oliveira, A. C., de Oliveira, L. S., Dekany, R. G., Dohlen, K., dos Santos, J. B., dos Santos, L. H., Ellis, R., Fabricius, M., Ferrand, D., Ferreira, D., Golebiowski, M., Greene, J. E., Gross, J., Gunn, J. E., Hammond, R., Harding, A., Hart, M., Heckman, T. M., Hirata, C. M., Ho, P., Hope, S. C., Hovland, L., Hsu, S.-F., Hu, Y.-S., Huang, P.-J., Jaquet, M., Jing, Y., Karr, J., Kimura, M., King, M. E., Komatsu, E., Le Brun, V., Le Fèvre, O., Le Fur, A., Le Mignant, D., Ling, H.-H., Loomis, C. P., Lupton, R. H., Madec, F., Mao, P., Marrara, L. S., Mendes de Oliveira, C., Minowa, Y., Morantz, C., Murayama, H., Murray, G. J., Ohyama, Y., Orndorff, J., Pascal, S., Pereira, J. M., Reiley, D., Reinecke, M., Ritter, A., Roberts, M., Schwochert, M. A., Seiffert, M. D., Smee, S. A., Sodre, L., Spergel, D. N., Steinkraus, A. J., Strauss, M. A., Surace, C., Suto, Y., Suzuki, N., Swinbank, J., Tait, P. J., Takada, M., Tamura, T., Tanaka, Y., Tresse, L., Verducci, O., Vibert, D., Vidal, C., Wang, S.-Y., Wen, C.-Y., Yan, C.-H., and Yasuda, N. Prime Focus Spectrograph (PFS) for the Subaru telescope: Overview, recent progress, and future perspectives. In *Ground-Based and Airborne Instrumentation for Astronomy VI*, volume 9908, pp. 99081M, eprint: arXiv:1608.01075, August 2016. doi: 10.1117/12.2232103.

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, 2020. doi: 10.21105/joss.02505. URL https://doi.org/10.21105/joss.02505.

Vakili, M. and Hahn, C. How are galaxies assigned to halos? searching for assembly bias in the sdss galaxy clustering. *The Astrophysical Journal*, 872(1):115, 2019.

Villaescusa-Navarro, F., Hahn, C., Massara, E., Banerjee, A., Delgado, A. M., Ramanah, D. K., Charnock, T., Giusarma, E., Li, Y., Allys, E., et al. The quijote simulations. *The Astrophysical Journal Supplement Series*, 250 (1):2, 2020.

Wang, Y., Zhai, Z., Alavi, A., Massara, E., Pisani, A., Benson, A., Hirata, C. M., Samushia, L., Weinberg, D. H., Colbert, J., Doré, O., Eifler, T., Heinrich, C., Ho, S., Krause, E., Padmanabhan, N., Spergel, D., and Teplitz, H. I. The High Latitude Spectroscopic Survey on the Nancy Grace Roman Space Telescope. *The Astrophysical Journal*, 928(1):1, March 2022. ISSN 0004-637X. doi: 10.3847/1538-4357/ac4973.

Wilson, A. G. and Izmailov, P. Bayesian deep learning and a probabilistic perspective of generalization. *Advances in neural information processing systems*, 33:4697–4708, 2020.

Zentner, A. R., Hearin, A., van den Bosch, F. C., Lange, J. U., and Villarreal, A. S. Constraints on assembly bias from galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, 485(1):1196–1209, 2019.

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(2):760, 2007.

A. Data

In this section, we describe in greater detail the observational data on which our cosmological parameters are inferred, as well as the synthetic training and test data generated.

A.1. Observational Data

We use a sample of CMASS Luminous Red Galaxies from the BOSS Data Release 12 as our observational data[1]. We limit our analysis to the subsample of CMASS sample at the Southern Galactic Cap (SGC) within the angular footprint, $\mathrm{DEC} > -6$ deg. and $-25 < \mathrm{RA} < 28$ deg., and redshift range, $0.45 < z < 0.6$. The reason for this is that the QUIJOTE simulation boxes are not big enough to include the rest of the CMASS sample. In total, our sample consists of 109,636 galaxies. Visual illustrations of the sample can be found in (Hahn & Melchior, 2022; Hahn et al., 2022).

A.2. SIMBIG Forward Model

We use the SIMBIG forward modelling pipeline (Hahn et al., 2022) to generate field-level synthetic observations that aim to be statistically indistinguishable from BOSS observations. This pipeline consists of four distinct steps: (1) N-body simulations, (2) a dark matter halo finder, (3) a halo occupation distribution framework (HOD), and (4) application of survey realism. A schematic illustrating the pipeline is provided in Figure 1.

The N-body simulations are taken from the QUIJOTE suite (Villaescusa-Navarro et al., 2020). The simulations evolve 1024^3 cold dark matter particles from $z = 127$ to $z = 0.5$ in a cosmological volume $1(h^{-1}Gpc)^3$ using the TreePM GADGET-III. These simulations accurately model matter clustering down to non-linear scales beyond $k = 0.5h/Mpc$.

From these N-body simulations, dark matter halos are identified using the ROCKSTAR halo finder (Behroozi et al., 2012a). The ROCKSTAR approach has been shown to robustly and accurately track dark matter halo location and substructure using phase-space information (Behroozi et al., 2012b). Then, halos are populated with a state-of-the-art HOD framework. Specifically, the standard Zheng, et al. (Zheng et al., 2007) HOD model, which populates halos using M_h and five free HOD parameters, is expanded by including assembly, concentration, and velocity biases. These biases add the necessary flexibility to account for recent evidence suggesting that galaxies occupy halos in ways that depend on halo properties beyond M_h(*e.g.* assembly history; Hadzhiyska et al., 2021; Vakili & Hahn, 2019; Zentner et al., 2019; Hadzhiyska et al., 2022b; More et al., 2016).

Finally, survey realism is applied to the HOD galaxy catalog to produce a CMASS-like galaxy catalog. First, the $1(h^{-1}Gpc)^3$ box is remapped to a cuboid (Carlson & White, 2010) and then cut to the BOSS survey geometry. Then, the galaxy catalog is trimmed to $z \in (0.45, 0.6)$, and fiber collisions are applied. Ultimately, the forward models are determined by 5 ΛCDM cosmological parameters, $\Omega_m, \Omega_b, h, n_s, \sigma_8$, and 9 HOD parameters. We refer readers to Hahn et al. (2022); Hahn et al. for further details.

To construct our training set, we use 2,518 high-resolution QUIJOTE N-body simulations arranged in a Latin hypercube configuration (LHC). The LHC imposes uniform priors on the cosmological parameters, which conservatively encompass the *Planck* cosmological constraints. For each simulation, we forward-model 10 CMASS-like galaxy catalogs using unique HOD parameters randomly sampled from a conservative prior. We split the resulting $25,180$ simulations into a $20,000$ and $5,180$ training and validation set.

A.3. Test simulations

In order to demonstrate that we can infer accurate and unbiased cosmological constraints, we test our analysis on three different sets of realistic test simulations that differ from the training dataset and have been developed within SIMBIG and introduced in (Hahn et al., 2023): TEST0, TEST1, and TEST2.

TEST0 uses QUIJOTE N-body simulations that have the same specifications as those arranged in the LHC, but were run at a fiducial cosmology with $\Omega_m = 0.3175, \Omega_b = 0.049, h = 0.6711, n_s = 0.9624, \sigma_8 = 0.834$. The halo finder, HOD framework and survey realism are the same as those used in the training set, but the HOD parameters span a narrower prior. This test dataset contains 500 synthetic galaxy catalogs.

[1] https://data.sdss.org/sas/dr12/boss/lss/

TEST1 involves the same N-body simulations as TEST0, but a different halo finder: the Friend-of-Friend algorithm (Davis et al., 1985). Moreover, assembly, concentration, and satellite velocity biases are not considered in the HOD model. Central velocity bias is implemented, as the halo velocities in FoF halo catalogs correspond to the bulk velocity of the dark matter particles in the halo rather than the velocity of the central density peak of the halo. This test dataset contains 500 synthetic galaxy catalogs.

TEST2 uses 25 ABACUSSUMMIT N-body simulations (Maksimova et al., 2021) in the "base" configuration of the suite. The simulations contain 6912^3 particles in a $(2h^{-1}Gpc)^3$ volume box. Halo catalogs are constructed from these simulations using the COMPASO halo finder (Hadzhiyska et al., 2022a) and each of them is divided into 8 boxes of volume $1(h^{-1}Gpc)^3$. Halos are populated with galaxies using the same HOD model implemented in the training set, with HOD parameters that sample the same narrower priors used in TEST0. This test dataset contains 1,000 synthetic galaxy catalogs.

All three test datasets incorporate the same survey realism as the training dataset to produce CMASS-like galaxy catalogs.

A.4. Galaxy Density Field

To apply CNNs to our observational and simulated galaxy samples, we mesh the galaxy distribution into a box, with voxel size $64 \times 128 \times 128$. We choose this size because divisibility by two allows for easier downsampling in the CNN. First, we place the distribution into a $[707, 1414, 1414]$ Mpc$/h$ box and convert it into a 3D density field using a cloud-in-cell mass assignment (Birdsall & Fuss, 1969). For our observational sample, we include systematics weights for multiple effects, redshift failures, stellar density, and seeing conditions (Ross et al., 2012; Anderson et al., 2012), in the mass assignment.

Since our data occupies a $[577.3, 1414, 1224]$ Mpc$/h$ box, we fill some of the box with zero-valued voxels. Our voxels have size $\sim [11, 11, 11]$ Mpc$/h$, thus we impose an effective scale cut of $k < k_{\rm max} = 0.28\ h/{\rm Mpc}$. While this is larger than the scale cut imposed in the SIMBIG P_ℓ analysis (Hahn et al.), we find that it is sufficient to place significant cosmological constraints. Moreover, pushing to even smaller scale cuts presents significant computational challenges in terms of the required memory.

B. Convolutional Neural Network

We provide additional details about the architecture and hyperparameters of our 3D CNN in this section.

The CNN architecture consists of 5 convolutional blocks. Each convolutional block begins with a convolutional layer that convolves its input with a number of $3 \times 3 \times 3$ kernels. This convolution is performed with 1-voxel zero-padding. This is followed by a rectified linear unit (ReLU). The output of the ReLU unit is then downsampled using max-pooling, which enables the network to learn features at increasing scales by reducing the size of its internal representations. Finally, batch normalization is applied, which typically speeds up training and enables generalization (Santurkar et al., 2018). Following the convolutional blocks, the activation maps are flattened and fed into three fully-connected layers that output $\hat{\boldsymbol{\theta}}$. These layers also use ReLU activation functions, but do not perform batch normalization.

In order to prevent overfitting on the training simulations, we include in the CNN's final architecture significant levels of dropout. This technique randomly sets to zero a percentage of neuron activations during training. Specifically, we use dropout percentages of $p = 0.15$ for each convolutional block and $p = 0.4$ for each fully connected blocks. Additionally, we introduce a large ℓ_2 penalty term with normalization value $\lambda = 0.0275$ on the network weights. In applying dropout in both the convolutional and fully connected layers, the network is forced to train on a smaller subset of active neurons, leading to underutilization of the network's capacity. Moreover, with the ℓ_2 penalty term in the loss function, the network's flexibility, and subsequently its ability to learn specific features, is limited. While these measures ultimately limit the constraining power of the CNN, they ensure robustness and generalizability, and thus protect against the fact that the SIMBIG forward models, and in general any forward model, are approximate.

CNN training is performed using a supervised learning approach. Specifically, we optimize the weights of the network to minimize the mean-squared-error (MSE) loss between $\hat{\boldsymbol{\theta}}_{\rm normed}$ and $\boldsymbol{\theta}^{\rm true}_{\rm normed}$, where we normalize both $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{\rm true}$ to $(0, 1)$, to prevent their varying ranges from affecting the loss differently. The optimization is performed using Stochastic Gradient Descent (SGD) with momentum $\beta = 0.9$. The neural network is trained in mini-batches of 32 maps. We use the OneCycleLR learning rate scheduler, which involves gradually increasing and then decreasing the learning rate during a single training cycle, and has been shown to lead to faster convergence and improved generalization (Smith & Topin, 2019). We use a maximal learning rate of $r = 0.01$. During training, the input maps are also randomly flipped horizontally and

vertically with $p = 0.5$ to further improve network generalization. We train the CNN on a single A100 GPU core until the MSE computed on the validation set has not improved for 20 consecutive epochs. Training the CNN in this context takes roughly 8 hours. Our CNN is implemented using PyTORCH (Paszke et al., 2019).

The CNN's architecture and hyperparameters are determined through experimentation and are roughly modelled off of previous successful image classifiers (Szegedy et al., 2016; Krizhevsky et al., 2017). To determine the specifics of our network, we train 60 networks using the Optuna hyperparameter framework (Akiba et al., 2019). Specifically, we vary the number of convolutional blocks between 3 and 6, the number of fully-connected layers between 1 and 6, the base number of channels of the convolutional blocks between 2 and 14, the width of the fully-connected layers between 128 and 1024, the dropout in both convolutional and fully-connected layers between $p = 0$ and $p = 0.5$, the ℓ_2 penalty between $\lambda = 10^{-4}$ and $\lambda = 10^{-1}$, and the max learning rate between $r = 10^{-5}$ and $r = 10^{-2}$. Ultimately, we aim to maximize the network's ability to extract relevant features from the galaxy density field while maintaining its ability to generalize beyond the SIMBIG training simulations. To that end, we select the network configuration that maximizes the network's MSE on the held-out validation models while minimizing the ratio between training MSE and validation MSE. However, in order to pass the validation tests on the out-of-distribution TEST1 and TEST2, we found that it was necessary to impose slightly stricter regularization on the network. To that end, the dropout and ℓ_2 terms were increased through trial-and-error from the optuna output to their reported values.

B.1. Weight Marginalization

In order to further prevent the CNN from overfitting on the training set, we perform a weight marginalization step, converting our CNN into a Bayesian Neural Network (BNN). In contrast to other neural networks, BNNs train the model weights as a distribution rather than searching for an optimal value. This allows them to capture the uncertainty in the weights and outputs of the model. The ultimate goal of BNNs is to quantify the uncertainty introduced by the models in terms of outputs and weights so as to explain the trustworthiness of the prediction. There are multiple methods for weight marginalization, such as Monte Carlo Dropout (Gal & Ghahramani, 2016) or Variational Inference (Graves, 2011).

In this work, we use Stochastic Weight Averaging (SWA; Maddox et al., 2019; Wilson & Izmailov, 2020). SWA is predicated on the observation that the parameters of deep neural networks often converge to the edges of low-loss regions. This edge-type convergence is sub-optimal, as these solutions are more susceptible to the shift between train and test error surfaces. SWA approximates the posterior distribution of the weights of the CNN as a Normal distribution, whose mean and covariance are given by

$$\bar{w} = \frac{1}{N_{\rm swa}} \sum_{n=1}^{N_{\rm swa}} w_n, \quad \Sigma = \frac{1}{N_{\rm swa}} \sum_{n=1}^{N_{\rm swa}} (w_n - \bar{w})(w_n - \bar{w})^T, \tag{1}$$

respectively, where w are the weights of the network, n is the time step during network optimization/training, and $N_{\rm swa}$ are the total steps over which SWA is performed.

By adopting this scheme, SWA solutions tend to converge to the center of flat loss regions, thereby leading to more stable and generalizeable solutions. Indeed, SWA has already been shown to lead to better generalization to out-of-distribution data (Wilson & Izmailov, 2020), which is expected to improve the robustness of our analysis. Moreover, SWA has been shown to outperform competing methods in multiple tasks (Maddox et al., 2019), and has been previously applied to astrophysics (Cranmer et al., 2021) and cosmology (Lemos et al., 2023). We use the publicly available COSMOSWAG implementation[2]. The compressed galaxy field that we feed as input to SBI is the output of the SWA network: a set of 10 samples of the posterior distribution weights of the CNN — a 50-dimensional data vector.

[2] https://github.com/Pablo-Lemos/cosmoSWAG