
Simulation-based inference with scattering representations: scattering is all you need

Kiyam Lin

Mullard Space Science Laboratory, University College London, Holmbury St Mary,
Dorking, Surrey RH5 6NT, UK

Department of Physics and Astronomy, University College London, Gower Street,
London, WC1E 6BT, UK
kiyam.lin@ucl.ac.uk

Benjamin Joachimi

Department of Physics and Astronomy, University College London, Gower Street,
London, WC1E 6BT, UK
b.joachimi@ucl.ac.uk

Jason D. McEwen

Mullard Space Science Laboratory, University College London, Holmbury St Mary,
Dorking, Surrey RH5 6NT, UK
jason.mcewen@ucl.ac.uk

Abstract

We demonstrate the first successful use of scattering representations without further compression for simulation-based inference (SBI) with images (i.e. field-level), illustrated with a cosmological case study. Scattering representations provide a highly effective representational space for subsequent learning tasks, although the higher dimensional compressed space introduces challenges. We overcome these through spatial averaging, coupled with more expressive density estimators. Compared to alternative methods, such an approach does not require additional simulations for either training or computing derivatives, is interpretable, and resilient to covariate shift. As expected, we show that a scattering only approach extracts more information than traditional second order summary statistics.

1 Introduction

Simulation-based inference (SBI) has recently emerged as a powerful approach for scientific inference that frees the task of parameter inference from defining a tractable and accurate likelihood [20]. The premise of neural density approaches to SBI is to make use of forward simulations to learn a density estimator representing the likelihood [44], posterior [42, 24] or likelihood-to-evidence ratio [21, 19]. The benefits of SBI mean that as long as we are capable of performing accurate stochastic forward simulations, it becomes possible to infer an unbiased posterior regardless of the complexity of the data being simulated. Nevertheless, compression is a critical component of SBI for many problems and requires further study, which is the focus of the current article. Throughout, we consider an illustrative case study from the field of cosmology involving non-Gaussian random fields.

In many scientific fields, traditional parameter inference approaches assume a Gaussian likelihood of second order statistics and simple parametric modelling of systematics to simplify the task of statistical inference [7, 1, 28]. Such an approach, however, is in many cases not sufficient to fully

exploit the richness of upcoming data [32, 52] and it becomes important to make use of summary statistics that contain beyond second order information.

For image data, one such approach is through the use of higher order statistics [e.g. 47, 13, 25, 38, 39, 40, 41]. These statistics, however, are difficult to compute and it is often even more difficult to accurately characterise their covariance structure for use with an analytical likelihood. Another direction involves the use of image-level (commonly called field-level) inference [e.g. 45, 9, 49, 6, 46]. Such approaches, however, must overcome the challenges of high dimensional Markov chain Monte Carlo (MCMC) sampling and approximate likelihoods. SBI has recently been demonstrated as an effective alternative for field-level inference, eliminating these challenges [27, 22, 31]. However, due to the high dimensionality of field-level data, there is a need for substantial data compression.

There are currently two main approaches to performing data compression for field-level SBI in cosmology. One route makes use of statistical compression methods, such as MOPED [26] or score compression [3]. These methods are designed such that the compression can be lossless with respect to the Fisher information. Statistical compression methods have implementation challenges, however, as they require simulations to be run at perturbed parameters so that derivatives can be computed by finite differences. Some simulation suites support this (allocating computational budget for this purpose) but most do not. Alternatively, automatic differentiation could be leveraged to compute derivatives [14], although that requires codes implemented within a framework that provides automatic differentiation, which are not always readily available. Furthermore, statistical compression methods are not generally lossless and only asymptotically lossless around a maximum likelihood estimation point. An alternative avenue is to make use of neural compression, of which convolutional neural networks (CNNs) have proven to be a popular architecture for field-level data [48, 27]. The drawback of neural compression is that it requires a substantial additional simulation budget for training. Furthermore, it is not interpretable and inference can suffer from covariate shift if the training data does not match the distribution of real data. Hybrid statistical and neural compression techniques have also been considered, such as Information Maximising Neural Networks (IMNN) [15]. While this realizes the advantages of both frameworks, it also realizes all of the drawbacks of both, requiring access to numerical derivatives, a significant simulation budget for generating training data, and can suffer covariate shift.

Wavelet scattering representations provide an alternative summary statistic [34, 12]. Whilst scattering representations are inspired by CNNs, they are designed and not learnt. This has the distinct advantage of not being a neural method. Consequently, they are interpretable [17, 2, 16] and do not require any training data. Scattering representations have gained traction in cosmology recently and appear to be highly effective at capturing informative non-Gaussian information [17, 2, 16, 50, 18]. In the asymptotic case of infinite depth they capture all information content of the field analysed, while also capturing the majority of information in a relatively small number of levels [34, 36, 16]. Furthermore, scattering transforms have recently been considered in cosmology as a form of compression for SBI [22, 8]. However, they have either been combined with neural compression [22] or demonstrated to provide unstable inferences [8]. The use of scattering representations without further neural compression has not been demonstrated successfully.

In this work we demonstrate the successful use of scattering representations without further compression as summary statistics of non-Gaussian cosmological random fields for accurate SBI. This approach does not suffer from the drawbacks of other statistical and neural compression techniques, i.e. it does not require additional simulations for computing derivatives, it does not require any simulations for training, is interpretable and not sensitive to covariate shift during data compression.

2 Wavelet Scattering Representation

Wavelet scattering representations [34, 12] leverage stable wavelet representations [33] to create powerful hierarchical representational spaces that satisfy isometric invariance up to a particular scale, are stable to diffeomorphisms and probe all content of the underlying field. Consequently, they provide an excellent representational space for subsequent learning [11, 37]. In this work we consider scattering representations as the compressed representational space on which to perform neural likelihood estimation [44].

The scattering propagator of field f with wavelet ψ_λ , for parameters $\lambda = (j, \chi)$ with scale j and orientation χ , is given by

$$U[\lambda]f = |f \star \psi_\lambda|, \quad (1)$$

where \star denotes the convolution operator and the modulus function is adopted to introduce non-linearity since it is non-expansive and thus preserves the stability of wavelet representations. Scattering representations for the path $p = (\lambda_1, \lambda_2, \dots, \lambda_d)$, with depth d , can then be constructed by applying a cascade of propagators:

$$U[p]f = U[\lambda_1, \lambda_2, \dots, \lambda_d]f = U[\lambda_d] \dots U[\lambda_2]U[\lambda_1]f = |||f \star \psi_{\lambda_1} | \star \psi_{\lambda_2} | \dots \star \psi_{\lambda_d} |. \quad (2)$$

By adopting carefully designed filters, i.e. wavelets, combined with a non-expansive activation function, scattering propagators inherit the stability properties from the underlying wavelets, yielding representations that are stable to diffeomorphisms. Scattering propagators themselves, however, do not yield representations that exhibit isometric invariances. The introduction of a form of averaging is necessary to ensure isometric invariance. Scale-dependent averaging is performed by convolution with a scaling function ϕ to yield the scattering representation [34, 12]

$$S[p]f = |||f \star \psi_{\lambda_1} | \star \psi_{\lambda_2} | \dots \star \psi_{\lambda_d} | \star \phi. \quad (3)$$

For compression, scattering representations are more challenging to work with than aforementioned neural and statistical compression due to their relatively higher dimensional representational space. We perform spatial averaging [cf. 17] to further compresses the representational space.

3 Methodology

Simulations: We adopt the high resolution Quijote Λ CDM N-body simulations [51], extracting 20 fields per simulation by slicing each simulation cube depth-wise and making use of all 2,000 dark matter only simulations run over a latin-hypercube. The non-linear and correlated cosmological parameters, including the cold dark matter density Ω_m , baryonic matter density Ω_b , scalar spectral index n_s , Hubble parameter h , and the root-mean-square matter fluctuation σ_8 , are varied over the ranges shown in Appendix A (Table 2). For this simulation suite numerical derivatives are not available, hence it is not possible to perform statistical compression (MOPED or score) or hybrid compression (IMNN). As we are not applying any neural based compression schema, we can use all simulations for training the neural likelihood density estimator without the need to reserve any to train a compressor.

Compression: To calculate the compressed scattering representation we adopt the `kymatio` package¹ [5] using a Morlet mother wavelet, introduced for astrophysics in [35], that is rotated and dilated to recover the set of wavelets considered. Since dark matter fields are expected to be isotropic on large scales (due to the cosmological principle), we set the number of spatial orientations probed to two and take the average. We take randomly selected fields from the simulation cube sliced in depth with resolution of 512^2 pixels and only probe scales, J , up to $2^{J+1} = 512$. This reduces the dimensionality of the summary statistic to a data vector of length $n = 1 + J + \frac{1}{2}J \times (J - 1) = 37$. For comparison, we consider an alternative simple compression, which also does not require derivatives or training, based on binning the power spectrum of the field. This bandpower approach, however, only captures second order information. We first calculate the power spectrum using the `pylians` package², with 362 bins, and further bin into bandpowers using log spaced bins with the binning scheme described in [10], resulting in a compressed dimension of 14. We also consider a summary statistic where scattering coefficients are concatenated with bandpowers.

SBI: We perform neural likelihood estimation (NLE) [44] using the `PyDELFI` package³ [4] with the compressed summary statistics described above. Whilst the performance of neural posterior estimation [24] may be considered in a future study, for cosmology, NLE invaluablely gives direct access to a learned likelihood that can be tested such as with a goodness of fit test. We follow the approach presented in [32, 52] where an ensemble of Masked Autoregressive Flows (MAFs; [43]) are considered, each with differing numbers of Masked Autoencoder for Distribution Estimation (MADE) components [23], and then model averaged, weighted by the relative validation likelihood. We found that standard `PyDELFI` settings for the MAFs are not sufficiently expressive when learning

¹<https://github.com/kymatio/kymatio>

²<https://github.com/franciscovillaescusa/Pylians3>

³<https://github.com/justinalsing/pydelfi>

Parameter	Truth	Scattering	Bandpowers	Combined
σ_8	0.869	0.876 ± 0.018	0.894 ± 0.046	0.883 ± 0.015
Ω_m	0.336	$0.298^{+0.052}_{-0.083}$	$0.238^{+0.045}_{-0.077}$	$0.283^{+0.044}_{-0.064}$

Table 1: Posterior constraints on σ_8 and Ω_m with one standard deviation errors for different compressed representations.

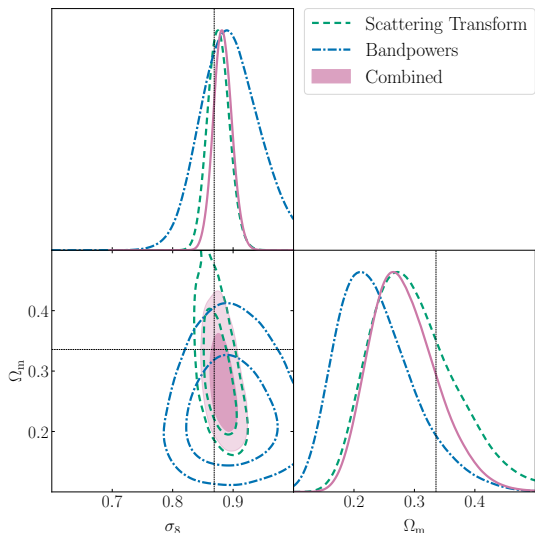


Figure 1: Posterior constraints for different compressed representations. Scattering representations, without further neural compression, are highly effective when using more expressive density estimators.

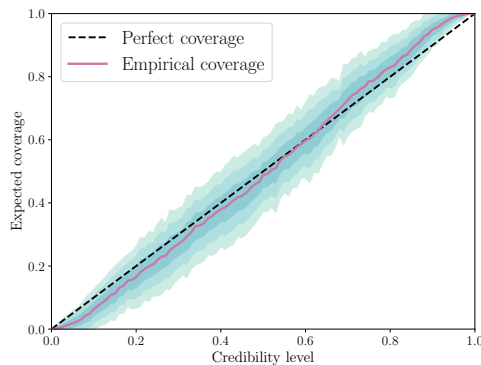


Figure 2: Coverage tests using TARP [30] for the combined scattering and bandpower summary statistic (similar results are found when considering each summary statistic separately). The blue regions decreasing in saturation indicate the 1, 2 and 3 σ errors obtained from bootstrapping over 100 realisations. The dashed black line denotes perfect coverage. Empirical coverage is generally good.

our relatively high dimensional data vector. Instead, we consider MAFs with four hidden layers of 50 neurons each. Finally, we sample the posterior defined in terms of the learned likelihood using the *nautilus* package⁴ [29].

4 Results

Performing SBI on a mock data vector (parameters in Table 1), we find that SBI with scattering performs well without further compression as indicated by the results of Figure 1 and Table 1 (on a single CPU core with 8GB RAM). We only show Ω_m and σ_8 results as the other three cosmological parameters are not well constrained by dark matter only N-body simulations, as expected. Inspection of the ensemble of trained MAFs demonstrates the density estimators are converged, in contrast with cases considered in [8]. For comparison, bandpowers are less effective as they only capture second order statistical information. Since the scattering and bandpower representations capture different statistical information, constraints can be improved by combining both summary statistics. To quantify the improvement, we find that scattering alone can yield $\sim 60\%$ tighter constraints in σ_8 than bandpowers alone, boosting the improvement to $\sim 70\%$ in the combined summary statistics case.

To ensure that the posterior distribution we infer is representative and unbiased, we conduct coverage tests based on the Tests of Accuracy with Random Points (TARP) algorithm [30]. TARP measures the expected coverage probability for a given credibility level empirically from samples. Results are shown in Figure 2, where the empirical coverage is shown to approximate the ideal perfect coverage scenario reasonably well. Here we consider the combined summary statistic, although we perform the same test on all considered summary statistics and find good coverage, with some minor overdispersion only around the 90% credibility level.

⁴<https://github.com/johannesulf/nautilus>

5 Conclusions

We demonstrate for the first time that field-level SBI can be performed effectively using wavelet scattering representations alone to provide a compressed representation without any further neural compression. The challenge of a higher dimensional compressed representation is overcome by incorporating spatial averaging and a judicious choice of parameters, coupled with more expressive density estimators. This approach has considerable advantages compared to the alternative statistical or neural compression approaches typically considered in that it does not require additional simulations to train a neural compressor, it does not require simulation suites with perturbed parameters needed to compute numerical derivatives for statistical compression approaches, it is more interpretable than the use of black-box neural networks, and it is not sensitive to covariate shift. Furthermore, we demonstrate that SBI with scattering representations extracts beyond two-point statistical information, as expected, since recovered parameter constraints are distinct and tighter than those recovered by bandpowers (thus, scattering representations and bandpowers can be combined to improve constraining power further still). In a sense, for field-level SBI, scattering is all you need for compression. That said, we do not dogmatically advocate that scattering alone should always be used. While the scattering approach presented has numerous advantages, our findings motivate further studies directly comparing the effectiveness of different compression approaches for field-level SBI.

Acknowledgments and Disclosure of Funding

The authors would like to thank Matt Price and Alessio Spurio Mancini for invaluable discussions. KL thanks the UCL Cosmoparticle Initiative and STFC (grant number ST/W001136/1). BJ acknowledges support by the ERC-selected UKRI Frontier Research Grant EP/Y03015X/1 and by STFC Consolidated Grant ST/V000780/1. JDM is supported by EPSRC (grant number EP/W007673/1) and STFC (grant number ST/W001136/1). This work used facilities provided by the UCL Cosmoparticle Initiative. We thank Dr Edd Edmonson for technical support at the UCL HPC facilities.

References

- [1] T. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila, D. Bacon, et al. Dark energy survey year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Physical Review D*, 105(2):023520, 2022.
- [2] E. Allys, T. Marchand, J.-F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. p. Mallat. New interpretable statistics for large-scale structure analysis and generation. *Physical Review D*, 102(10):103506, 2020.
- [3] J. Alsing and B. Wandelt. Generalized massive optimal data compression. *Monthly Notices of the Royal Astronomical Society: Letters*, 476(1):L60–L64, 2018.
- [4] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. *Monthly Notices of the Royal Astronomical Society*, 488(3):4440–4458, 2019.
- [5] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andron, E. Belilovsky, J. Bruna, V. Lostanlen, M. Chaudhary, M. J. Hirn, E. Oyallon, S. Zhang, C. Cella, and M. Eickenberg. Kymatio: Scattering transforms in python. *Journal of Machine Learning Research*, 21(60):1–6, 2020. URL <http://jmlr.org/papers/v21/19-047.html>.
- [6] A. Andrews, J. Jasche, G. Lavaux, and F. Schmidt. Bayesian field-level inference of primordial non-gaussianity using next-generation galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 520(4):5746–5763, 2023.
- [7] M. Asgari, C.-A. Lin, B. Joachimi, B. Giblin, C. Heymans, H. Hildebrandt, A. Kannawadi, B. Stölzner, T. Tröster, J. L. van den Busch, et al. KiDS-1000 Cosmology: Cosmic shear constraints and comparison between two point statistics. *Astronomy & Astrophysics*, 645:A104, 2021.

- [8] B. R.-S. Blancard, C. Hahn, S. Ho, J. Hou, P. Lemos, E. Massara, C. Modi, A. M. Dizgah, L. Parker, Y. Yao, et al. Galaxy clustering analysis with SIMBIG and the wavelet scattering transform. *arXiv preprint arXiv:2310.15250*, 2023.
- [9] S. S. Boruah, E. Rozo, and P. Fiedorowicz. Map-based cosmology inference with lognormal cosmic shear maps. *Monthly Notices of the Royal Astronomical Society*, 516(3):4111–4122, 2022.
- [10] M. L. Brown, P. G. Castro, and A. N. Taylor. Cosmic microwave background temperature and polarization pseudo- C_l estimators and covariances. *Monthly Notices of the Royal Astronomical Society*, 360(4):1262–1280, 7 2005. doi: 10.1111/j.1365-2966.2005.09111.x.
- [11] J. Bruna and S. Mallat. Classification with scattering operators. In *CVPR 2011*, pages 1561–1566. IEEE, 2011.
- [12] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.
- [13] P. A. Burger, O. Friedrich, J. Harnois-Déraps, P. Schneider, M. Asgari, M. Bilicki, H. Hildebrandt, A. H. Wright, T. Castro, K. Dolag, et al. KiDS-1000 cosmology: Constraints from density split statistics. *Astronomy & Astrophysics*, 669:A69, 2023.
- [14] J.-E. Campagne, F. Lanusse, J. Zuntz, A. Boucaud, S. Casas, M. Karamanis, D. Kirkby, D. Lanzieri, Y. Li, and A. Peel. JAX-COSMO: An end-to-end differentiable and gpu accelerated cosmology library. *arXiv preprint arXiv:2302.05163*, 2023.
- [15] T. Charnock, G. Lavaux, and B. D. Wandelt. IMNN: Information maximizing neural networks. *Astrophysics Source Code Library*, pages ascl-1804, 2018.
- [16] S. Cheng and B. Ménard. How to quantify fields or textures? A guide to the scattering transform. *arXiv preprint arXiv:2112.01288*, 2021.
- [17] S. Cheng, Y.-S. Ting, B. Ménard, and J. Bruna. A new approach to observational cosmology using the scattering transform. *Monthly Notices of the Royal Astronomical Society*, 499(4): 5902–5914, 2020.
- [18] S. Cheng, G. A. Marques, D. Grandón, L. Thiele, M. Shirasaki, B. Ménard, and J. Liu. Cosmological constraints from weak lensing scattering transform using HSC Y1 data. *arXiv preprint arXiv:2404.16085*, 2024.
- [19] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger. Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation. *Journal of Cosmology and Astroparticle Physics*, 2022(09):004, 2022.
- [20] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.
- [21] C. Durkan, I. Murray, and G. Papamakarios. On contrastive learning for likelihood-free inference. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 2771–2781. PMLR, 7 2020. URL <https://proceedings.mlr.press/v119/durkan20a.html>.
- [22] M. Gatti, N. Jeffrey, L. Whiteway, J. Williamson, B. Jain, V. Ajani, D. Anbajagane, G. Giannini, C. Zhou, A. Porredon, et al. Dark energy survey year 3 results: Simulation-based cosmological inference with wavelet harmonics, scattering transforms, and moments of weak lensing mass maps. validation on simulations. *Physical Review D*, 109(6):063534, 2024.
- [23] M. Germain, K. Gregor, I. Murray, and H. Larochelle. Made: Masked autoencoder for distribution estimation. In *International conference on machine learning*, pages 881–889. PMLR, 2015.
- [24] D. Greenberg, M. Nonnenmacher, and J. Macke. Automatic posterior transformation for likelihood-free inference. In *International Conference on Machine Learning*, pages 2404–2414. PMLR, 2019.

- [25] J. Harnois-Déraps, N. Martinet, T. Castro, K. Dolag, B. Giblin, C. Heymans, H. Hildebrandt, and Q. Xia. Cosmic shear cosmology beyond two-point statistics: a combined peak count and correlation function analysis of DES-Y1. *Monthly Notices of the Royal Astronomical Society*, 506(2):1623–1650, 2021.
- [26] A. F. Heavens, R. Jimenez, and O. Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, 317(4):965–972, 2000.
- [27] N. Jeffrey, L. Whiteway, M. Gatti, J. Williamson, J. Alsing, A. Porredon, J. Prat, C. Doux, B. Jain, C. Chang, et al. Dark energy survey year 3 results: likelihood-free, simulation-based w cdm inference with neural compression of weak-lensing map statistics. *arXiv preprint arXiv:2403.02314*, 2024.
- [28] B. Joachimi, C.-A. Lin, M. Asgari, T. Tröster, C. Heymans, H. Hildebrandt, F. Köhlinger, A. Sánchez, A. Wright, M. Bilicki, et al. KiDS-1000 methodology: Modelling and inference for joint weak gravitational lensing and spectroscopic galaxy clustering analysis. *Astronomy & Astrophysics*, 646:A129, 2021.
- [29] J. U. Lange. nautilus: boosting Bayesian importance nested sampling with deep learning. *Monthly Notices of the Royal Astronomical Society*, 525(2):3181–3194, 2023.
- [30] P. Lemos, A. Coogan, Y. Hezaveh, and L. Perreault-Levasseur. Sampling-Based Accuracy Testing of Posterior Estimators for General Inference. *40th International Conference on Machine Learning*, 202:19256–19273, 1 2023. doi: 10.48550/arXiv.2302.03026.
- [31] P. Lemos, L. Parker, C. Hahn, S. Ho, M. Eickenberg, J. Hou, E. Massara, C. Modi, A. Moradinezhad Dizgah, B. Régalo-Saint Blancard, et al. SimBIG: Field-level simulation-based inference of large-scale structure. *Machine Learning for Astrophysics*, page 18, 2023.
- [32] K. Lin, M. von Wietersheim-Kramsta, B. Joachimi, and S. Feeney. A simulation-based inference pipeline for cosmic shear with the Kilo-Degree Survey. *Monthly Notices of the Royal Astronomical Society*, 524(4):6167–6180, 2023.
- [33] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1999.
- [34] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.
- [35] J. D. McEwen, M. Hobson, A. Lasenby, and D. Mortlock. A high-significance detection of non-gaussianity in the WMAP 3-yr data using directional spherical wavelets. *Monthly Notices of the Royal Astronomical Society: Letters*, 371(1):L50–L54, 2006.
- [36] J. D. McEwen, C. G. Wallis, and A. N. Mavor-Parker. Scattering networks on the sphere for scalable and rotationally equivariant spherical cnns. *arXiv preprint arXiv:2102.02828*, 2021.
- [37] J. D. McEwen, C. G. R. Wallis, and A. N. Mavor-Parker. Scattering networks on the sphere for scalable and rotationally equivariant spherical CNNs. In *International Conference on Learning Representations (ICLR)*, 1 2022.
- [38] D. Munshi, J. McEwen, T. Kitching, P. Fosalba, R. Teyssier, and J. Stadel. Estimating the integrated bispectrum from weak lensing maps. *Journal of Cosmology and Astroparticle Physics*, 2020(05):043, 2020.
- [39] D. Munshi, T. Namikawa, J. McEwen, T. Kitching, and F. Bouchet. Morphology of weak lensing convergence maps. *Monthly Notices of the Royal Astronomical Society*, 507(1):1421–1433, 2021.
- [40] D. Munshi, H. Lee, C. Dvorkin, and J. D. McEwen. Weak lensing trispectrum and Kurt-spectra. *Journal of Cosmology and Astroparticle Physics*, 2022(11):020, 2022.
- [41] D. Munshi, G. Jung, T. Kitching, J. McEwen, M. Liguori, T. Namikawa, and A. Heavens. Position-dependent correlation function of weak-lensing convergence. *Physical Review D*, 107(4):043516, 2023.

- [42] G. Papamakarios and I. Murray. Fast ε -free inference of simulation models with bayesian conditional density estimation. *Advances in neural information processing systems*, 29, 2016.
- [43] G. Papamakarios, T. Pavlakou, and I. Murray. Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, 30, 2017.
- [44] G. Papamakarios, D. Sterratt, and I. Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In *The 22nd international conference on artificial intelligence and statistics*, pages 837–848. PMLR, 2019.
- [45] N. Porqueres, A. Heavens, D. Mortlock, and G. Lavaux. Lifting weak lensing degeneracies with a field-based likelihood. *Monthly Notices of the Royal Astronomical Society*, 509(3):3194–3202, 2022.
- [46] N. Porqueres, A. Heavens, D. Mortlock, G. Lavaux, and T. L. Makinen. Field-level inference of cosmic shear with intrinsic alignments and baryons. *arXiv preprint arXiv:2304.04785*, 2023.
- [47] S. Pyne, A. Tenneti, and B. Joachimi. Three-point intrinsic alignments of dark matter haloes in the IllustrisTNG simulation. *Monthly Notices of the Royal Astronomical Society*, 516(2): 1829–1845, 2022.
- [48] D. Ribli, B. Á. Pataki, J. M. Zorrilla Matilla, D. Hsu, Z. Haiman, and I. Csabai. Weak lensing cosmology with convolutional neural networks on noisy data. *Monthly Notices of the Royal Astronomical Society*, 490(2):1843–1860, 2019.
- [49] E. Tsaprazi, N.-M. Nguyen, J. Jasche, F. Schmidt, and G. Lavaux. Field-level inference of galaxy intrinsic alignment from the SDSS-III BOSS survey. *Journal of Cosmology and Astroparticle Physics*, 2022(08):003, 2022.
- [50] G. Valogiannis, S. Yuan, and C. Dvorkin. Precise cosmological constraints from boss galaxy clustering with a simulation-based emulator of the wavelet scattering transform. *Physical Review D*, 109(10):103503, 2024.
- [51] F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys, et al. The quijote simulations. *The Astrophysical Journal Supplement Series*, 250(1):2, 2020.
- [52] M. von Wietersheim-Kramsta, K. Lin, N. Tessore, B. Joachimi, A. Loureiro, R. Reischke, and A. H. Wright. KiDS-SBI: Simulation-Based Inference Analysis of KiDS-1000 Cosmic Shear. *ArXiv*, 2024.

A Cosmological parameter ranges

Table 2 shows the priors ranges used both for the Quijote N-body simulation latin-hypercube and for inference. For inference, uniform priors were adopted for all parameters that matched the prior width of the simulations such that our learnt likelihood had full training data support.

Parameter	Prior range
Ω_m	[0.1, 0.5]
Ω_b	[0.03, 0.07]
h	[0.5, 0.9]
n_s	[0.8, 1.2]
σ_8	[0.6, 1.0]

Table 2: Cosmological parameter ranges.