A Hierarchy of Normalizing Flows for Modelling the Galaxy–Halo Relationship

Christopher C. Lovell [1 2] **Sultan Hassan** [3] **Daniel Anglés-Alcázar** [4 5] **Greg Bryan** [6] **Giulio Fabbian** [5 7] **Shy Genel** [5 8] **ChangHoon Hahn** [9] **Kartheik Iyer** [6] **James Kwon** [10] **Natalí de Santi** [5 11] **Francisco Villaescusa-Navarro** [5 9]

Abstract

Using a large sample of galaxies taken from the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project, a suite of hydrodynamic simulations varying both cosmological and astrophysical parameters, we train a normalizing flow (NF) to map the probability of various galaxy and halo properties conditioned on astrophysical and cosmological parameters. By leveraging the learnt conditional relationships we can explore a wide range of interesting questions, whilst enabling simple marginalisation over nuisance parameters. We demonstrate how the model can be used as a generative model for arbitrary values of our conditional parameters; we generate halo masses and matched galaxy properties, and produce realisations of the halo mass function as well as a number of galaxy scaling relations and distribution functions. The model represents a unique and flexible approach to modelling the galaxy–halo relationship.

[1]Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth, PO1 3FX, UK [2]Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK [3]School of Computation, University of Edenborrow, Edenborrow, United Kingdom [4]Department of Physics, University of Connecticut, 196 Auditorium Road, U-3046, Storrs, CT 06269-3046, USA [5]Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY, 10010, USA [6]Department of Astronomy, Columbia University, 550 West 120th Street, New York, NY, 10027, USA [7]School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, Wales CF24 3AA, United Kingdom [8]Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY, 10027, US [9]Department of Astrophysical Sciences, Princeton University, Princeton NJ 08544, USA [10]Department of Physics, University of California, Santa Barbara, CA 93106, US [11]Instituto de Física, Universidade de São Paulo, R. do Matão 1371, 05508-090, São Paulo, Brasil. Correspondence to: Christopher C. Lovell <christopher.lovell@port.ac.uk>.

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

1. Introduction

Galaxies form within dark matter haloes, and their evolution is closely tied to the evolutionary history of their host halo – an understanding of the galaxy–halo relationship is key to a cosmological interpretation of galaxy populations (Wechsler & Tinker, 2018). Many computational modelling methods take explicit advantage of the galaxy–halo connection, populating haloes in less computationally expensive Dark-Matter only N-body simulations with galaxies in order to achieve larger volumes, or explore a larger range of parameters (Benson, 2010; Somerville & Davé, 2015). In the past decade a growing number of supervised machine learning (ML) methods for modelling the galaxy–halo relationship have emerged, using properties of the halo as features from which to predict the host galaxy properties (*e.g.* Kamdar et al., 2016; Agarwal et al., 2018; Jo & Kim, 2019; Lovell et al., 2022; de Santi et al., 2022; Jespersen et al., 2022; Icaza-Lizaola et al., 2023; Chittenden & Tojeiro, 2023). Almost all of these methods are deterministic; a given set of halo properties leads to a single predicted galaxy property.[1] However, galaxy evolution is not *entirely* determined by the host halo; other factors contribute to the properties of a galaxy at a given time that are not encoded in the halo properties and assembly history, *e.g.* the stochastic nature of stellar and AGN feedback. Deterministic methods are therefore susceptible to underpredicting the scatter in galaxy properties for a fixed set of input halo properties; there is insufficient information to model the true scatter. Finally, many studies have demonstrated the intrinsic stochasticity in results from numerical galaxy formation simulations, due to both explicit randomness (Genel et al., 2019) and the computational architecture (Borrow et al., 2022).

What we require is a non-deterministic method for populating haloes with galaxies, that can model the multi-dimensional joint distribution of galaxy properties, accounting for the scatter introduced by all latent variables. *Generative models*, particularly those for density estimation,

[1]Rodrigues et al. (2023) demonstrate a non-deterministic approach, however this relies on binning combined with a classification procedure.

are an approach with promise in this domain (Kingma & Welling, 2013; Goodfellow et al., 2014; Jimenez Rezende et al., 2014). Normalizing flows (NF; Dinh et al., 2015; Jimenez Rezende & Mohamed, 2015) are one such technique, offering exact density estimation (equivalent to the multi-dimensional likelihood) and efficient sampling. Hassan et al. (2022) demonstrate the use of NFs on the CAMELS simulation suite (Villaescusa-Navarro et al., 2021; 2023) by training a model on maps of atomic hydrogen density. They build a generative model that can produce HI maps for arbitrary cosmological and astrophysical parameters. Friedman & Hassan (2022) present an update to this model, fully utilising the spatial information from the map using the Glow NF model (Kingma & Dhariwal, 2018) to produce better constraints on cosmological parameters.

In this paper we build a generative model for discrete halo (M_h) and galaxy properties ($M_\star$, $M_{\rm gas}$, $M_\bullet$, SFR), using a hierarchy of NF's trained on haloes and galaxies taken from the CAMELS simulation suite.

2. Methods

A normalizing flow (NF) models some data $\mathbf{x}$ as a bijective transformation of some base distribution, typically a gaussian noise variable $\mathbf{u}$,

$$\mathbf{x} = f_\theta(\mathbf{u}) \tag{1}$$

$$\mathbf{u} \sim \pi(\mathbf{u}) \ , \tag{2}$$

where f_θ is invertible and differentiable, with parameters θ. This allows the target density $p_\phi(\mathbf{x})$ to be written as

$$p_\phi(\mathbf{x}) = \pi(f_\theta^{-1}(\mathbf{x})) \left| \det \left(\frac{\delta f_\theta^{-1}}{\delta \mathbf{x}} \right) \right| \ . \tag{3}$$

For maximum flexibility f_θ and f_θ^{-1} are modelled using invertible neural networks (NN). f_θ can be represented by multiple stacked layers, in order to produce highly complex mappings from the noise to the target density.

In order to build a conditional model we require a dataset with pairs of variables, $\mathcal{D} = \{(\mathbf{z}, \mathbf{x})\}$. Here, the $\mathbf{z}$ parameters are responsible for the generation of $\mathbf{x}$, and we wish to model $p_x(\mathbf{x}|\mathbf{z})$. To include this conditional dependence in our model we incorporate these parameters in our transformation, $\mathbf{x} = f_\theta(\mathbf{u}, \mathbf{z})$ (Winkler et al., 2019). We implement a version of a Neural spline flow (Durkan et al., 2019; Dolatabadi et al., 2020).

The Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS; Villaescusa-Navarro et al., 2021) are a large ensemble of N-body and hydrodynamic simulations exploring the effect of cosmological and astrophysical parameter choices on galaxy evolution and structure formation. In this study we focus on the SIMBA simulation suite only (Davé et al., 2019). For full details please refer to Villaescusa-Navarro et al. (2021; 2023); Ni et al. (2023). Each simulation is defined by the initial random phases, as well as 4 astrophysical parameters ($A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$) and 2 cosmological parameters ($\Omega_{\rm m}$, σ_8). The following cosmological parameters are kept fixed in all simulations: $\Omega_{\rm b} = 0.049$, $h = 0.6711$, $n_s = 0.9624$, $M_\nu = 0.0$ eV, $w = -1$, $\Omega_K = 0$. The fiducial astrophysical parameters are defined at $A = 1.0$ and varied around this value to control the relative strength of the various feedback implementations in each simulation. There are a number of different simulation sets within the CAMELS suite; the Latin Hypercube (LH) set contains 1000 simulations where the 6 parameters are varied using a latin hypercube; the cosmic variance CV set contains 27 simulations that only differ in the value of the random seed in the initial conditions.

We train three complementary flows, each conditional on the cosmological and astrophysical parameters (an illustration of the different flows is shown in Figure 1). The *abundance flow* models the absolute abundance of subhaloes with mass $> 10^{10}\mathrm{M}_\odot$, $p_\phi(\mathbf{n} \,|\, \mathbf{z})$. We add gaussian noise to the data in the LH set equal to the scatter in the abundance in the CV set 50 times, and train on this augmented data set, to mimic the effect of cosmic variance. The *halo flow* models the density distribution of halo masses, $p_\phi(\mathbf{y} \,|\, \mathbf{z})$. By coupling the *abundance* and *halo flows*, we can generate the volume normalised halo mass function for arbitrary parameters; an example is shown in the top left corner of Figure 2. Finally, the *galaxy flow* models the distribution of galaxy properties within dark matter haloes by further conditioning on the subhalo mass, $p_\phi(\mathbf{x} \,|\, \mathbf{z}, \mathbf{y})$. We predict the stellar mass, gas mass, black hole mass and star formation rate.

We reserve a random subset of entire LH set simulations for testing (15%), and use the rest for training and validation; this ensures there is no overlap between the train and test sets of galaxies with the same astrophysical and cosmological parameters. We use the $z = 0$ snapshot from each simulation, and reserve a study of the redshift dependence for future work. Each flow contains 16 layers, each consisting of a linear rational spline bijection (with 256 segments) coupled to an autoregressive NN layer consisting of two hidden layers with 256 and 128 nodes, respectively. We train using the ADAM optimizer (Kingma & Ba, 2014), with a multi-step learning rate starting at 5×10^{-3}, with $\gamma = 0.1$, using mini batches of size 2048 that are randomly shuffled after each epoch. At the end of each epoch we evaluate on the validation set, and save the model if the validation error has improved, to avoid overfitting.

3. Results

In this section we demonstrate an example use case for the model by predicting the galaxy and halo properties for a set of parameters not used in the training procedure. We take

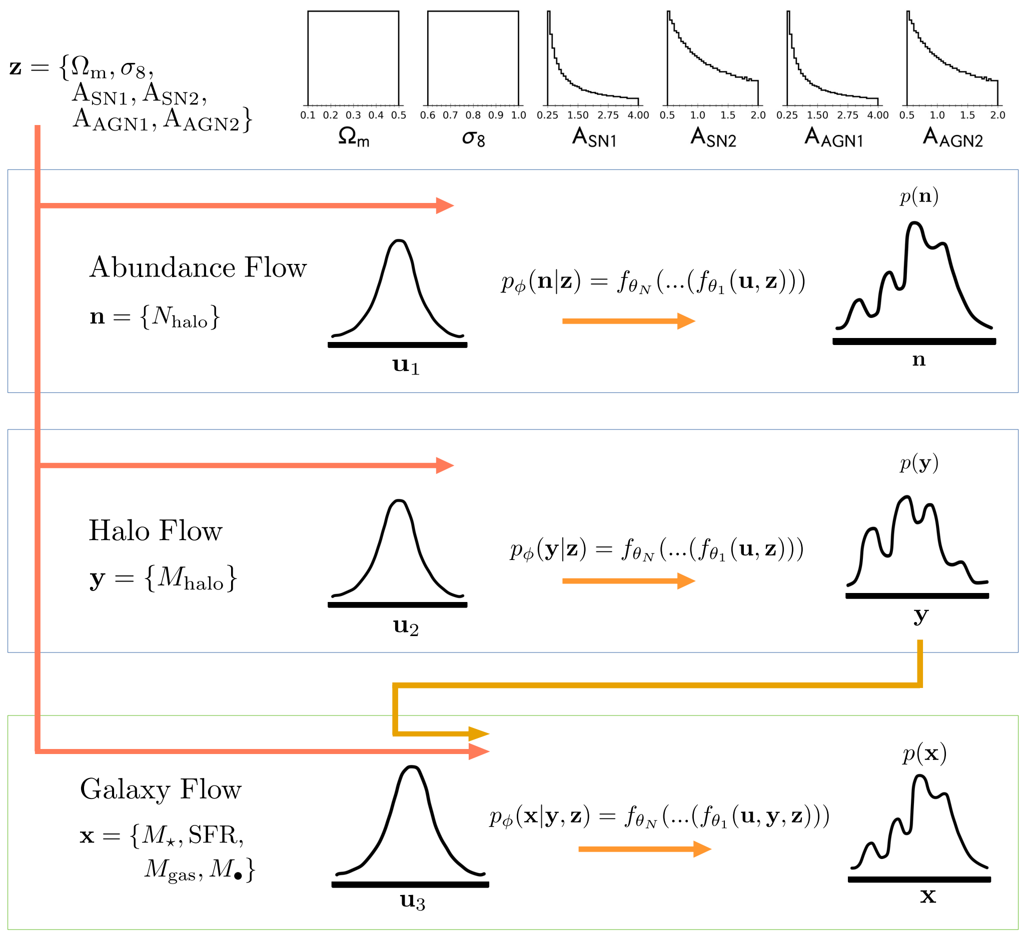

Figure 1. High level diagram of the model. The distribution of the conditional cosmological and astrophysical parameters is shown at the top. The *abundance*, *halo* and *galaxy flows* are shown below. The arrows highlight the direction of conditional dependence, as well as the mapping from each simple base distribution to the complex target density distribution.

these parameters from an LH set simulation from the test set, and first predict the halo mass function given the input parameters $\mathbf{z}$. We then use the *abundance flow* to predict the cumulative number of subhaloes with mass $M_\mathrm{halo} > 10^{10}\,M_\odot$, $\mathbf{n}$, and the *halo flow* to predict the distribution of their masses, $\mathbf{y}$. Combined we can produce the halo mass function (HMF), shown in the top left panel of Figure 2 for 50 realisations, and compared to the true HMF from the corresponding LH set simulation. The model successfully reproduces the distribution function within the scatter of the realisations. We can also change one of the conditional parameters and explore the impact on the HMF. This is shown in the top row of Figure 2; there is a strong positive correlation between Ω_m and the normalisation of the HMF.

We can also predict the properties of the galaxy within each host subhalo by providing the subhalo mass as well as the other conditional parameters to the *galaxy flow*. Whilst galaxy properties may be dependent on additional parameters as well as mass, the flow is able to model the full distribution of those properties at a given mass, marginalising over these unknown additional dependencies. The first panel in the second row of Figure 2 shows the galaxy stellar mass function (GSMF) produced when applied to haloes generated from the *abundance* and *halo flows*. The GSMF is reproduced within the scatter of the 50 realisations. We can, again, fix parameters and explore the impact on the GSMF; we show this for Ω_m, A_SN1 & A_SN2 in the second row of Figure 2.

The third, fourth, fifth and sixth rows in Figure 2 also show predictions for the star forming sequence, the stellar mass–gas mass relation, the stellar mass–black hole mass relation, and the stellar–halo mass relation, and the impact of chang-

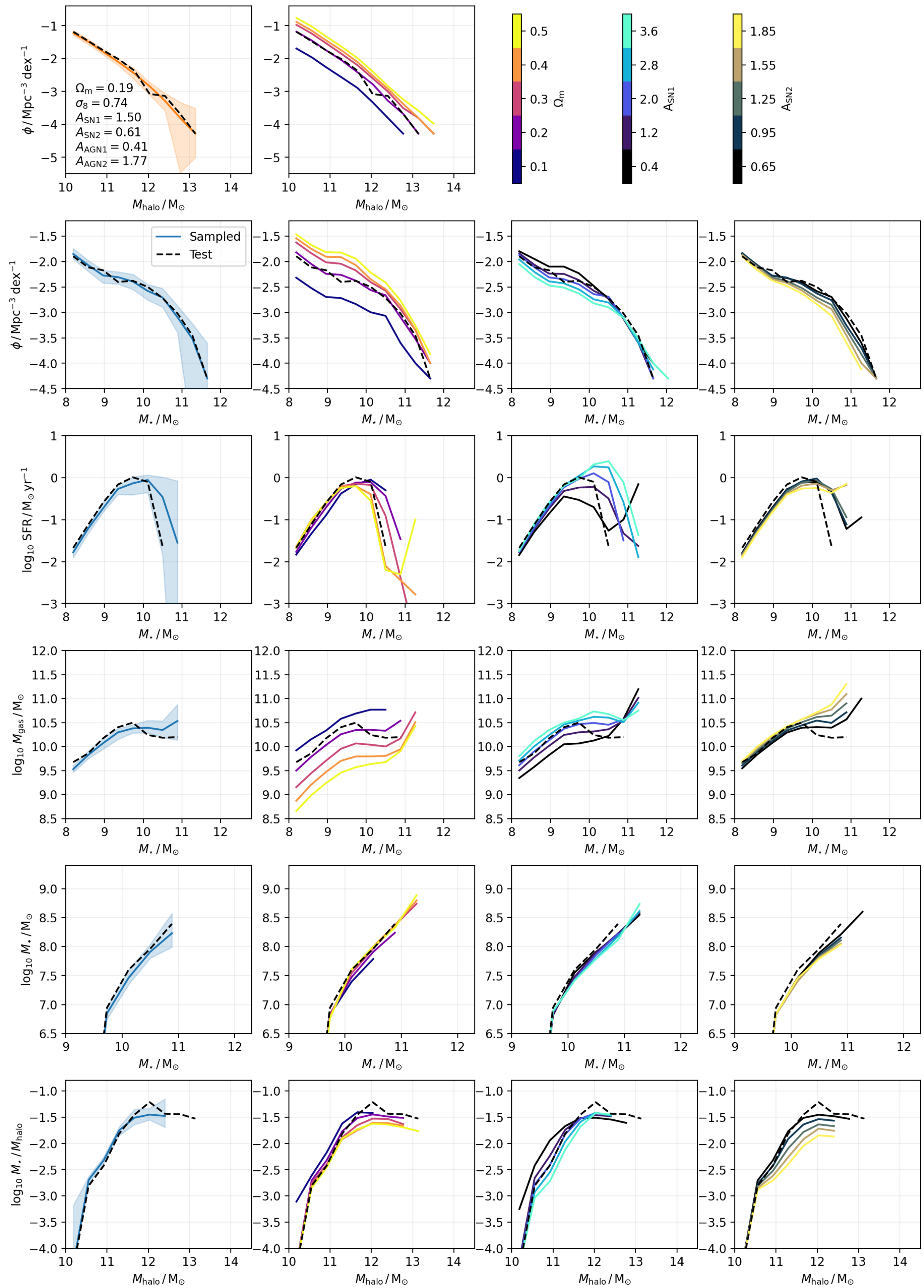

Figure 2. An example of the model predictions when used as a generative model for haloes and galaxies, for fixed and varying parameters. The top row shows the halo mass function (HMF), the second row the galaxy stellar mass function, the third row the star forming sequence, the fourth row the stellar mass–gas mass relation, the fifth row the stellar mass–black hole mass relation, and finally the stellar–halo mass relation. The first row shows predictions using haloes generated from the *abundance* and *halo flows*, as well as haloes taken directly from the LH set simulation.

ing conditional parameters ($\Omega_{\rm m}$, $A_{\rm SN1}$, $A_{\rm SN2}$) on each of these relations in turn. We emphasise that galaxy properties are predicted jointly, enabling us to predict these relations self consistently.

4. Conclusions

We present a novel approach to modelling the galaxy–halo relationship, using the density estimation capabilities of normalising flows to model the coupled halo and galaxy distribution conditioned on astrophysical and cosmological parameters. The model is able to self-consistently predict a number of halo and galaxy relations, and shows interesting correlations with different cosmological and astrophysical parameters, whilst marginalising over other nuisance parameters. There are a number of applications for such a model, from rapid generation of galaxy properties in dark matter only N-body simulations, to direct and indirect inference of astrophysical and cosmological parameters from individual galaxy properties or predicted scaling relations through simulation based inference (SBI) (Cranmer et al., 2019), an increasingly popular and flexible approach to inference (Papamakarios et al., 2017; Alsing et al., 2019; Hahn et al., 2019; Zhang et al., 2021; Dax et al., 2021; Hahn & Melchior, 2022; Huppenkothen & Bachetti, 2022; Wang et al., 2023).

5. Acknowledgements

CCL acknowledges support from a Dennis Sciama fellowship funded by the University of Portsmouth for the Institute of Cosmology and Gravitation. DAA acknowledges support by NSF grants AST-2009687 and AST-2108944, CXO grant TM2-23006X, Simons Foundation Award CCA-1018464, and Cottrell Scholar Award CS-CSA-2023-028 by the Research Corporation for Science Advancement. GF acknowledges the support of the European Research Council under the Marie Skłodowska Curie actions through the Individual Global Fellowship No. 892401 PiCOGAMBAS and of the Simons Foundation. NSMS acknowledges financial support from FAPESP, grants 2019/13108-0 and 2022/03589-4.

References

Agarwal, S., Davé, R., and Bassett, B. A. Painting galaxies into dark matter haloes using machine learning. *MNRAS*, 478:3410–3422, August 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty1169. URL `http://adsabs.harvard.edu/abs/2018MNRAS.478.3410A`.

Alsing, J., Charnock, T., Feeney, S., and Wandelt, B. Fast likelihood-free cosmology with neural density estimators and active learning. *MNRAS*, 488:4440–4458, September 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1960. URL `https://ui.adsabs.harvard.edu/abs/2019MNRAS.488.4440A`. ADS Bibcode: 2019MNRAS.488.4440A.

Benson, A. J. Galaxy Formation Theory. *Phys. Rep.*, 495(2-3):33–86, October 2010. ISSN 03701573. doi: 10.1016/j.physrep.2010.06.001. URL `http://arxiv.org/abs/1006.5394`. arXiv: 1006.5394.

Borrow, J., Schaller, M., Bahe, Y. M., Schaye, J., Ludlow, A. D., Ploeckinger, S., Nobels, F. S. J., and Altamura, E. The impact of stochastic modeling on the predictive power of galaxy formation simulations. *arXiv.2211.08442*, November 2022. doi: 10.48550/arXiv.2211.08442. URL `http://arxiv.org/abs/2211.08442`. arXiv:2211.08442 [astro-ph].

Chittenden, H. G. and Tojeiro, R. Modelling the galaxy-halo connection with semi-recurrent neural networks. *MNRAS*, 518:5670–5692, January 2023. ISSN 0035-8711. doi: 10.1093/mnras/stac3498. URL `https://ui.adsabs.harvard.edu/abs/2023MNRAS.518.5670C`. ADS Bibcode: 2023MNRAS.518.5670C.

Cranmer, M. D., Galvez, R., Anderson, L., Spergel, D. N., and Ho, S. Modeling the Gaia Color-Magnitude Diagram with Bayesian Neural Flows to Constrain Distance Estimates. *arXiv.1908.08045*, August 2019. doi: 10.48550/arXiv.1908.08045. URL `https://ui.adsabs.harvard.edu/abs/2019arXiv190808045C`. Publication Title: arXiv e-prints ADS Bibcode: 2019arXiv190808045C Type: article.

Davé, R., Anglés-Alcázar, D., Narayanan, D., Li, Q., Rafieferantsoa, M. H., and Appleby, S. SIMBA: Cosmological simulations with black hole growth and feedback. *MNRAS*, 486(2):2827, June 2019. doi: 10.1093/mnras/stz937. URL `https://ui.adsabs.harvard.edu/abs/2019MNRAS.486.2827D/abstract`.

Dax, M., Green, S. R., Gair, J., Macke, J. H., Buonanno, A., and Schölkopf, B. Real-Time Gravitational Wave Science with Neural Posterior Estimation. *Phys. Rev. L*, 127:241103, December 2021. ISSN 0031-9007. doi: 10.1103/PhysRevLett.127.241103. URL `https://ui.adsabs.harvard.edu/abs/2021PhRvL.127x1103D`. ADS Bibcode: 2021PhRvL.127x1103D.

de Santi, N. S. M., Rodrigues, N. V. N., Montero-Dorta, A. D., Abramo, L. R., Tucci, B., and Artale, M. C. Mimicking the halo-galaxy connection using machine learning. *MNRAS*, 514:2463–2478, August 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac1469. URL `https://ui.adsabs.harvard.edu/abs/`

2022MNRAS.514.2463D. ADS Bibcode: 2022MNRAS.514.2463D.

Dinh, L., Krueger, D., and Bengio, Y. NICE: Non-linear Independent Components Estimation. In *ICLR*, 2015. URL https://ui.adsabs.harvard.edu/abs/2014arXiv1410.8516D.

Dolatabadi, H. M., Erfani, S., and Leckie, C. Invertible Generative Modeling using Linear Rational Splines. In *23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, January 2020. URL https://ui.adsabs.harvard.edu/abs/2020arXiv200105168D.

Durkan, C., Bekasov, A., Murray, I., and Papamakarios, G. Neural Spline Flows. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, June 2019. URL https://ui.adsabs.harvard.edu/abs/2019arXiv190604032D.

Friedman, R. and Hassan, S. HIGlow: Conditional Normalizing Flows for High-Fidelity HI Map Modeling. In *Machine Learning and the Physical Sciences workshop, NeurIPS 2022*, November 2022. URL https://ui.adsabs.harvard.edu/abs/2022arXiv221112724F.

Genel, S., Bryan, G. L., Springel, V., Hernquist, L., Nelson, D., Pillepich, A., Weinberger, R., Pakmor, R., Marinacci, F., and Vogelsberger, M. A Quantification of the Butterfly Effect in Cosmological Simulations and Implications for Galaxy Scaling Relations. *The Astrophysical Journal*, 871:21, January 2019. ISSN 0004-637X. doi: 10.3847/1538-4357/aaf4bb. URL https://ui.adsabs.harvard.edu/abs/2019ApJ...871...21G. ADS Bibcode: 2019ApJ...871...21G.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014. URL https://papers.nips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

Hahn, C. and Melchior, P. Accelerated Bayesian SED Modeling Using Amortized Neural Posterior Estimation. *ApJ*, 938:11, October 2022. ISSN 0004-637X. doi: 10.3847/1538-4357/ac7b84. URL https://ui.adsabs.harvard.edu/abs/2022ApJ...938...11H. ADS Bibcode: 2022ApJ...938...11H.

Hahn, C., Beutler, F., Sinha, M., Berlind, A., Ho, S., and Hogg, D. W. Likelihood non-Gaussianity in large-scale structure analyses. *MNRAS*, 485:2956–2969, May 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz558. URL https://ui.adsabs.harvard.edu/abs/2019MNRAS.485.2956H. ADS Bibcode: 2019MNRAS.485.2956H.

Hassan, S., Villaescusa-Navarro, F., Wandelt, B., Spergel, D. N., Anglés-Alcázar, D., Genel, S., Cranmer, M., Bryan, G. L., Davé, R., Somerville, R. S., Eickenberg, M., Narayanan, D., Ho, S., and Andrianomena, S. HIFLOW: Generating Diverse HI Maps and Inferring Cosmology while Marginalizing over Astrophysics Using Normalizing Flows. *ApJ*, 937:83, October 2022. ISSN 0004-637X. doi: 10.3847/1538-4357/ac8b09. URL https://ui.adsabs.harvard.edu/abs/2022ApJ...937...83H. ADS Bibcode: 2022ApJ...937...83H.

Huppenkothen, D. and Bachetti, M. Accurate X-ray timing in the presence of systematic biases with simulation-based inference. *MNRAS*, 511:5689–5708, April 2022. ISSN 0035-8711. doi: 10.1093/mnras/stab3437. URL https://ui.adsabs.harvard.edu/abs/2022MNRAS.511.5689H. ADS Bibcode: 2022MNRAS.511.5689H.

Icaza-Lizaola, M., Bower, R. G., Norberg, P., Cole, S., and Schaller, M. A sparse regression approach for populating dark matter haloes and subhaloes with galaxies. *MNRAS*, 518:2903–2920, January 2023. ISSN 0035-8711. doi: 10.1093/mnras/stac3265. URL https://ui.adsabs.harvard.edu/abs/2023MNRAS.518.2903I. ADS Bibcode: 2023MNRAS.518.2903I.

Jespersen, C. K., Cranmer, M., Melchior, P., Ho, S., Somerville, R. S., and Gabrielpillai, A. Mangrove: Learning Galaxy Properties from Merger Trees. *ApJ*, 941:7, December 2022. ISSN 0004-637X. doi: 10.3847/1538-4357/ac9b18. URL https://ui.adsabs.harvard.edu/abs/2022ApJ...941....7J. ADS Bibcode: 2022ApJ...941....7J.

Jimenez Rezende, D. and Mohamed, S. Variational Inference with Normalizing Flows. In *Proceedings of the 32nd International Conference on Machine Learning*, May 2015. URL https://ui.adsabs.harvard.edu/abs/2015arXiv150505770J.

Jimenez Rezende, D., Mohamed, S., and Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *Proceedings of the 31st International Conference on Machine Learning*

(ICML), January 2014. URL https://ui.adsabs.harvard.edu/abs/2014arXiv1401.4082J.

Jo, Y. and Kim, J.-h. Machine-assisted semi-simulation model (MSSM): estimating galactic baryonic properties from their dark matter using a machine trained on hydrodynamic simulations. *MNRAS*, 489:3565–3581, November 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz2304. URL http://adsabs.harvard.edu/abs/2019MNRAS.489.3565J.

Kamdar, H. M., Turk, M. J., and Brunner, R. J. Machine learning and cosmological simulations – I. Semi-analytical models. *MNRAS*, 455(1):642–658, January 2016. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/stv2310. URL http://mnras.oxfordjournals.org/content/455/1/642.

Kingma, D. P. and Ba, J. Adam: A Method for Stochastic Optimization. In *3rd International Conference for Learning Representations*, December 2014. URL http://adsabs.harvard.edu/abs/2014arXiv1412.6980K.

Kingma, D. P. and Dhariwal, P. Glow: Generative Flow with Invertible 1x1 Convolutions. In *Advances in Neural Information Processing Systems*, July 2018. URL http://arxiv.org/abs/1807.03039. arXiv:1807.03039 [cs, stat].

Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations (ICLR2014)*, December 2013. URL https://ui.adsabs.harvard.edu/abs/2013arXiv1312.6114K.

Lovell, C. C., Wilkins, S. M., Thomas, P. A., Schaller, M., Baugh, C. M., Fabbian, G., and Bahé, Y. A machine learning approach to mapping baryons on to dark matter haloes using the EAGLE and C-EAGLE simulations. *MNRAS*, 509:5046–5061, February 2022. ISSN 0035-8711. doi: 10.1093/mnras/stab3221. URL https://ui.adsabs.harvard.edu/abs/2022MNRAS.509.5046L. ADS Bibcode: 2022MNRAS.509.5046L.

Ni, Y., Genel, S., Anglés-Alcázar, D., Villaescusa-Navarro, F., Jo, Y., Bird, S., Di Matteo, T., Croft, R., Chen, N., de Santi, N. S. M., Gebhardt, M., Shao, H., Pandey, S., Hernquist, L., and Dave, R. The CAMELS project: Expanding the galaxy formation model space with new ASTRID and 28-parameter TNG and SIMBA suites. *arXiv.2304.02096*, April 2023. doi: 10.48550/arXiv.2304.02096. URL https://ui.adsabs.harvard.edu/abs/2023arXiv230402096N. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230402096N Type: article.

Papamakarios, G., Pavlakou, T., and Murray, I. Masked Autoregressive Flow for Density Estimation. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/6c1da886822c67822bcf3679d04369fa-Paper.pdf.

Rodrigues, N. V. N., de Santi, N. S. M., Montero-Dorta, A. D., and Abramo, L. R. High-fidelity reproduction of central galaxy joint distributions with Neural Networks. Technical report, January 2023. URL https://ui.adsabs.harvard.edu/abs/2023arXiv230106398R. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230106398R Type: article.

Somerville, R. S. and Davé, R. Physical Models of Galaxy Formation in a Cosmological Framework. *ARAA*, 53(1): 51–113, August 2015. ISSN 0066-4146, 1545-4282. doi: 10.1146/annurev-astro-082812-140951. URL http://arxiv.org/abs/1412.2712. arXiv: 1412.2712.

Villaescusa-Navarro, F., Anglés-Alcázar, D., Genel, S., Spergel, D. N., Somerville, R. S., Dave, R., Pillepich, A., Hernquist, L., Nelson, D., Torrey, P., Narayanan, D., Li, Y., Philcox, O., La Torre, V., Maria Delgado, A., Ho, S., Hassan, S., Burkhart, B., Wadekar, D., Battaglia, N., Contardo, G., and Bryan, G. L. The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations. *ApJ*, 915:71, July 2021. ISSN 0004-637X. doi: 10.3847/1538-4357/abf7ba. URL https://ui.adsabs.harvard.edu/abs/2021ApJ...915...71V. ADS Bibcode: 2021ApJ...915...71V.

Villaescusa-Navarro, F., Genel, S., Anglés-Alcázar, D., Perez, L. A., Villanueva-Domingo, P., Wadekar, D., Shao, H., Mohammad, F. G., Hassan, S., Moser, E., Lau, E. T., Machado Poletti Valle, L. F., Nicola, A., Thiele, L., Jo, Y., Philcox, O. H. E., Oppenheimer, B. D., Tillman, M., Hahn, C., Kaushal, N., Pisani, A., Gebhardt, M., Delgado, A. M., Caliendo, J., Kreisch, C., Wong, K. W. K., Coulton, W. R., Eickenberg, M., Parimbelli, G., Ni, Y., Steinwandel, U. P., La Torre, V., Dave, R., Battaglia, N., Nagai, D., Spergel, D. N., Hernquist, L., Burkhart, B., Narayanan, D., Wandelt, B., Somerville, R. S., Bryan, G. L., Viel, M., Li, Y., Irsic, V., Kraljic, K., Marinacci, F., and Vogelsberger, M. The CAMELS Project: Public Data Release. *ApJSS*, 265:54, April 2023. ISSN 0067-0049. doi: 10.3847/1538-4365/acbf47. URL https://ui.adsabs.harvard.edu/abs/2023ApJS..265...54V. ADS Bibcode: 2023ApJS..265...54V.

Wang, B., Leja, J., Villar, V. A., and Speagle, J. S. SBI++: Flexible, Ultra-fast Likelihood-free Inference Customized for Astronomical Application. *arXiv.2304.05281*, April 2023. doi: 10.48550/arXiv.2304.05281. URL https://ui.adsabs.harvard.edu/abs/2023arXiv230405281W. Publication Title: arXiv e-prints ADS Bibcode: 2023arXiv230405281W Type: article.

Wechsler, R. H. and Tinker, J. L. The Connection Between Galaxies and Their Dark Matter Halos. *ARAA*, 56(1):435–487, 2018. doi: 10.1146/annurev-astro-081817-051756. URL https://doi.org/10.1146/annurev-astro-081817-051756. _eprint: https://doi.org/10.1146/annurev-astro-081817-051756.

Winkler, C., Worrall, D., Hoogeboom, E., and Welling, M. Learning Likelihoods with Conditional Normalizing Flows. *arXiv.1912.00042*, November 2019. doi: 10.48550/arXiv.1912.00042. URL https://ui.adsabs.harvard.edu/abs/2019arXiv191200042W. Publication Title: arXiv e-prints ADS Bibcode: 2019arXiv191200042W Type: article.

Zhang, K., Bloom, J. S., Gaudi, B. S., Lanusse, F., Lam, C., and Lu, J. R. Real-time Likelihood-free Inference of Roman Binary Microlensing Events with Amortized Neural Posterior Estimation. *AJ*, 161(6):262, May 2021. ISSN 1538-3881. doi: 10.3847/1538-3881/abf42e. URL https://dx.doi.org/10.3847/1538-3881/abf42e. Publisher: The American Astronomical Society.