Cosmological Data Compression and Inference with Self-Supervised Machine Learning

Aizhan Akhmetzhanova [1] **Siddharth Mishra-Sharma** [2 3 1] **Cora Dvorkin** [1]

Abstract

The influx of massive amounts of new data from current and upcoming cosmological surveys necessitates compression schemes that can efficiently summarize the data with minimal loss of information. We investigate the potential of self-supervised machine learning to construct optimal summaries of cosmological datasets. Using a particular self-supervised machine learning method, VICReg (Variance-Invariance-Covariance Regularization) deployed on lognormal random fields as well as hydrodynamical cosmological simulations, we find that self-supervised learning can deliver highly informative summaries which can be used for downstream tasks, including providing precise and accurate constraints when used for parameter inference. Our results indicate that self-supervised machine learning techniques offer a promising new approach for cosmological data compression and analysis.

1. Introduction

Current and upcoming cosmological surveys such as DESI (Aghamousa et al., 2016), Euclid (Laureijs et al., 2011), the Vera C. Rubin Observatory (LSST) (Collaboration et al., 2012), and the Square Kilometer Array (SKA) (Weltman et al., 2020) will deliver massive amounts of data of various modalities; making full use of these complex datasets to probe cosmology is a challenging task. The raw datasets are typically first described in terms of a set of informative lower-dimensional data vectors or *summary statistics*, which are then used for parameter inference. These summary statistics are often motivated by inductive biases drawn from the physics of the problem at hand. Some widely used summary statistics include power spectra and higher n-point correlation functions, wavelet scattering transform coefficients, probability distribution functions, and many others.

While these statistics have been successful in placing tight constraints on the values of cosmological parameters, the sufficiency of manually-derived statistics (i.e., ability to compress all physically-relevant information) is the exception rather than the norm. Furthermore, in order to take advantage of recent advances in the field of simulation-based inference (SBI), the size of the summary statistic presents an important consideration due to the curse of dimensionality associated with the comparison of the simulated data to observations in high-dimensional space (Alsing & Wandelt, 2019).

A number of methods have been proposed to construct optimal statistics which are compact yet contain all of the relevant cosmological information. Some have focused on creating compression schemes which preserve the Fisher information content of the original dataset (Heavens et al., 2000; Zablocki & Dodelson, 2016; Alsing & Wandelt, 2018; Alsing & Wandelt, 2019; Charnock et al., 2018). Another line of research looked at compression schemes which optimize the mutual information between the summaries and the parameters of interest (Chen et al., 2020; Jeffrey et al., 2021).

In this work, we explore self-supervised machine learning as an alternative approach to obtaining compressed summary statistics. Instead of constructingsummaries by following a given prescription, self-supervised learning methods explore the data on their own and reduce dimensionality of the data based on its underlying structure and symmetries. Self-supervised learning has recently been applied to analyze astronomical images (Hayat et al., 2021) and particle collision events (Dillon et al., 2022; Dillon et al., 2022). We investigate the potential of the self-supervised machine learning techniques for compressing cosmological datasets into informative low-dimensional summaries. We compare the performance of this method to an equivalent supervised baseline model and, where applicable, theoretical constraints.

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA [2]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions [3]Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Correspondence to: Aizhan Akhmetzhanova <aakhmetzhanova@g.harvard.edu>.

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

2. Methodology

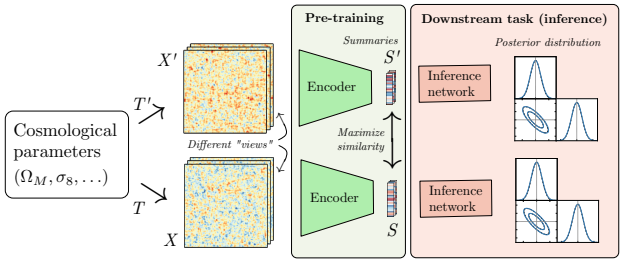

Figure 1. A schematic overview of the self-supervised learning pipeline implemented in this work. The T, T' are different transformations used to produce two views (e.g., lognormal density maps) X, X' of the same underlying cosmological parameters of interest (e.g., Ω_M and σ_8). The inference network is trained on the summaries S, S' obtained from the pre-training step.

In this work, we focus on a particular self-supervised method, VICReg (Variance-Invariance-Covariance Regularization) (Bardes et al., 2021). VICReg is designed to explicitly avoid a key challenge for self-supervised learning methods – the so-called *collapse* problem in which the neural network learns a trivial solution and produces the same constant summaries for different input vectors. VICReg addressed this problem through an easily interpretable *triple objective function* that maximizes the similarity of the summaries corresponding to the same image, while minimizing the redundancy between different features of the summary vectors and maintaining variance between summaries within a training batch.

Similarly to other self-supervised methods, VICReg can be divided into a pre-training step and a downstream task. During the pre-training step, the *encoder* network is first provided with two different *views* X and X' of an input I. In the image domain, so-called *views* are random transformations of the image I obtained by, for instance, cropping it at different locations, applying color jitters or blurring the image. In the context of cosmological data, different views might represent different realizations of an observable that corresponds to the same fundamental cosmological parameters, but, for instance, different initial conditions or evolution histories.

The encoder uses views X and X' to produce corresponding low-dimensional summaries S and S'. The summaries are then used as an input to a small expander network that maps them onto vectors Z and Z', called *embeddings*. Empirically, Bardes et al. (2021) and Zbontar et al. (2021) found that computing the VICReg loss on embeddings Z, Z' (which usually have more dimensions than summaries S, S') results in more informative summaries than computing the loss on directly on the summaries. The VICReg loss is computed on the level of embeddings, but the expander network is discarded after the pre-training step, and the summaries are used for the downstream tasks in the subsequent steps of the method. We show a schematic overview of the method in Fig. 1.

The three parts of the VICReg objective function are the invariance loss s, the variance loss v, and the covariance loss c. The invariance loss s measures the similarity between the outputs of the encoder by computing the mean-squared Euclidean distance between each pair of embeddings Z, Z'. For a batch of n pairs of views, the invariance loss is defined as: $s(Z, Z') = \frac{1}{n}\sum_i \|Z_i - Z'_i\|_2^2$.

The variance loss v is intended to prevent the norm collapse which occurs when the encoder maps every input to the same output. It measures the overall variance in a given batch across different dimensions in the embedding space and encourages the variance along each dimension j to be close to some constant γ: $v(Z) = \frac{1}{d}\sum_{j=1}^{d} \max\left(0, \gamma - \sqrt{\mathrm{Var}(Z_j) + \epsilon}\right)$. Here Z_j is a vector that consists of the values of the embeddings Z_i at dimension j, d is the dimensionality of embeddings Z, and γ and ϵ are fixed to 1 and 0.0001 respectively.

The covariance loss $c(Z)$ is used to address the informational collapse whereby different dimensions of the summaries encode the same information and are therefore redundant. It drives the covariance matrix $C(Z)$ to be close to a diagonal matrix by minimizing the sum of the squares of the off-diagonal entries of the covariance matrix: $c(Z) = \frac{1}{d}\sum_{i\neq j}[C(Z)]_{i,j}^2$.

The final loss function is a weighted sum of three loss terms:

$$\ell(Z, Z') = \lambda s(Z, Z') + \mu[v(Z) + v(Z')] + \nu[c(Z) + c(Z')], \tag{1}$$

where λ, μ, ν are hyperparameters controlling the weights assigned to each term in the loss function.

Once the pre-training step is complete, the summaries are used directly for downstream tasks by training a simple neural network, such as a multi-layer perceptron with a few layers. We use the summaries to infer cosmological parameters of interest and refer to the neural network used in this step as *inference network*. Assuming a Gaussian likelihood, we use the inference network to predict their means θ_n and covariances Σ_n by minimizing the negative log-likelihood function:

$$\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{1}{2}\ln|\Sigma_n| + \frac{1}{2}\left(\theta_n - \mu_n\right)^T \Sigma_n^{-1}\left(\theta_n - \mu_n\right)\right], \tag{2}$$

where μ_n are the true values of the parameters.

3. Experiments

3.1. Lognormal Overdensity Maps

We first test our methodology on mock cosmological data: lognormal random fields, which are commonly used as an approximation to matter density fields (Percival et al., 2004; Beutler et al., 2011; Cole et al., 2005).

Data and VICReg Set-up: We generate lognormal fields $\delta_{LN}(x)$ from 2D Gaussian overdensity fields $\delta_G(x)$ with a specified power spectrum $P_G(k)$. We then convert the Gaussian fields to obtain corresponding lognormal overdensity fields: $\delta_{LN}(x) = \exp\left(\delta_G(x) - \frac{1}{2}\sigma_G^2\right) - 1$, where σ_G^2 is the variance of the field $\delta_G(x)$. The Gaussian fields are produced with the `powerbox` package (Murray, 2018).

We take $P_G(k)$ to be linear matter power spectrum computed with the Eisenstein-Hu transfer function (Eisenstein & Hu, 1999) and generate the power spectra using the `pyccl` package (Chisari et al., 2019). For each $P_G(k)$, we vary two cosmological parameters: total matter density, Ω_M, and the r.m.s. of the present day (z = 0) density perturbations at scales of 8 h^{-1} Mpc, σ_8. We fix the remaining cosmological parameters to the following values: $\Omega_b = 0.05$, $h = 0.7$, $n_s = 0.96$, $N_{\rm eff} = 3.046$, $\sum m_\nu = 0$ eV. We use a grid of $N^2 = 100 \times 100$ points and set the volume of the box to be $V = L^2 = (1000\,{\rm Mpc})^2$.

We generate a set of 10,000 different combinations of cosmological parameters $\Omega_M \in [0.15, 0.45]$ and $\sigma_8 \in [0.65, 0.95]$. For each combination of Ω_M and σ_8, we simulate 10 different realizations of lognormal overdensity fields. These realizations, rotated and flipped at random, are used as different views to train the VICReg encoder network. We use 80% of the data for training, 10% for validation, and the remaining 10% for testing.

We compress the 100×100 maps down to summaries of length 16 using an encoder network with 9 convolutional layers and 2 fully-connected layers. The inference network used to infer parameters from the compressed summaries is a simple fully-connected neural network with 2 layers.

We train the encoder network for 200 epochs in the `PyTorch` (Paszke et al., 2019a) framework using `AdamW` (Kingma & Ba, 2014; Loshchilov & Hutter, 2019) optimizer with initial learning rate of 2×10^{-4} and cosine annealing. We set the λ, μ, and ν weights in the loss function to 5, 5, and 1 respectively.

We use the same training, validation, and test split when training the downstream inference network. The network is trained for 200 epochs with `AdamW`, with the initial learning rate 10^{-3}, reduced by a factor of 5 when the validation loss plateaus for 10 epochs. We evaluate the performance of both the encoder and inference networks on the validation set at the end of each epoch and save the model which yields the lowest validation loss.

Results: With both the encoder and the inference networks trained, we evaluate the performance of the VICReg method on the test dataset. We find the inference network trained on the VICReg summaries is able to recover the true values of cosmological parameters with both accuracy and precision, with relative errors on Ω_M and σ_8 equal to 5.2% and 1.3%, respectively. For comparison, a neural network with an equivalent architecture, trained on the maps directly in a fully-supervised manner, predicts the cosmological parameters with similar accuracy (relative errors on Ω_M and σ_8 are equal to 5.1% and 1.3%) which suggests that the encoder network has learnt an effective compression scheme which reduces the maps to summaries without substantial loss of information.

We also compare the Fisher information content of the lognormal fields and the summaries. In Fig. 2, we show the Fisher-forecasted constraints on Ω_M and σ_8. The Fisher contours from the lognormal fields and the VICReg summaries are in excellent agreement, demonstrating that the summaries preserve the Fisher information content of the maps almost entirely.

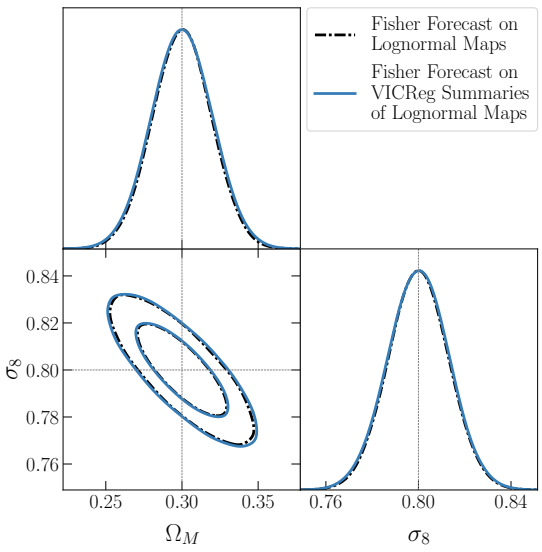

Figure 2. Constraints from Fisher forecast on the cosmological parameters Ω_M and σ_8 obtained from lognormal overdensity maps (black dashed line) and from summaries constructed with VICReg (blue solid line). The results shown on the plot were obtained for a fiducial cosmology with $\Omega_M = 0.3$ and $\sigma_8 = 0.8$.

3.2. CAMELS

We consider an application of the VICReg method to more complex data: total matter density maps from two publicly available hydrodynamic simulation suites, IllustrisTNG (Weinberger et al., 2017; Pillepich et al., 2018) and SIMBA

(Davé et al., 2019), which implement distinct galaxy formation models and are a part of the CAMELS project (Villaescusa-Navarro et al., 2021a;b). These maps represent spatial distribution of baryonic as well as dark matter within a slice of a given simulation. IllustrisTNG and SIMBA datasets contains 15,000 different maps each (1000 hydrodynamic simulations with 15 maps per simulation). We examine the efficiency of self-supervised summaries derived from these maps in conducting parameter inference.

VICReg Set-up: We modify the notion of two different views to represent total mass density maps from two different slices of the same simulation, rotated or flipped at random during training. This should enable the encoder network to learn relevant cosmological information from the maps and become insensitive to random spatial variations in the slices. We also find it helpful to modify the VICReg loss such that each batch includes 5 pairs of different 'views' from each simulation, as opposed to including a single pair per simulation (or per set of cosmological parameters). Since the CAMELS maps have more complexity than the lognormal maps, this allows the encoder network to learn from more variations.

Due to the high computational cost of running hydrodynamic simulations, IllustrisTNG and SIMBA have smaller size than the lognormal maps dataset used in Sec. 3.1, so we reserve more data for validation and testing purposes: 70% of the simulations for training, 20% for validation, and the remaining 10% for testing.

We use ResNet-18 (He et al., 2016; Paszke et al., 2019a) as the encoder which compresses the 256×256 maps to summaries of length 128. The inference network used for parameter inference is a simple fully-connected 2-layer neural network, with 512 units in each layer.

We train the encoder for 150 epochs with `AdamW` optimizer with initial learning rate 10^{-3}, which is multiplied by a factor of 0.3 when the validation loss plateaus for 10 epochs. The weights λ, μ, ν in the loss function are set to 25, 25, and 1, respectively. The inference network uses the same optimizer specifications with initial learning rate 7×10^{-4}. For both the encoder and the inference network, we save the models which perform best on the validation set.

Results: Figure 3 shows the predicted values of Ω_M (left panel) and σ_8 (right panel) against the true values for a subset of maps from the test set for the IllustrisTNG suite, with the error bars corresponding to predicted 1σ uncertainties (the corresponding plot for the SIMBA suite is shown in Appendix A). It can be seen that the inferred parameters provide a fairly accurate and unbiased estimate for the true parameters. Trained directly on the VICReg summaries, the inference model is able to infer the cosmological parameters with percent-level accuracy: the relative errors on Ω_M and σ_8 are 3.8% and 2.5% respectively for the SIMBA suite, and 3.7% and 1.9% for the IllustrisTNG suite. We find that performing field-level inference on the matter density maps with an equivalent (ResNet-18) supervised model results in similar constraints on the cosmological parameters: the relative errors on Ω_M and σ_8 are 3.3% and 2.3% respectively for the SIMBA suite and 3.3% and 1.8% for the IllustrisTNG suite. These results suggest that, despite massive reduction in the size and dimensionality of the data, the VICReg encoder network learns a near-optimal compression scheme.

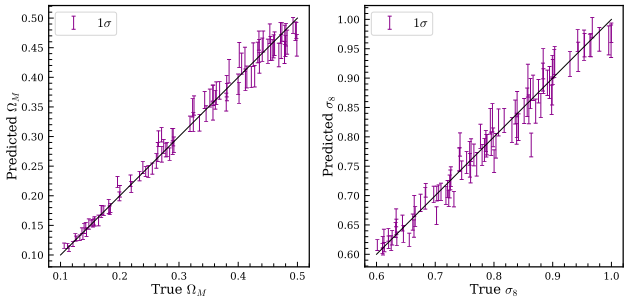

Figure 3. Predicted means and 1-σ uncertainties of cosmological parameters Ω_M and σ_8 compared to the true values of the parameters for total matter density maps from IllustrisTNG simulations. Predictions for the means and variances of the parameters were obtained by training simple inference neural network on VICReg summaries.

4. Conclusions

We have introduced the use of self-supervised machine learning for cosmological data compression and parameter inference. On applying our method to mock data with a tractable likelihood – lognormal random overdensity fields – we showed that parameter inference on the compressed data summaries saturates the theoretical Fisher information content. Deploying the method to more realistic data – total matter density maps based on hydrodynamic simulations from the CAMELS project – we found that, even for this more complex dataset of a smaller size (in terms of the number of simulations available for training), our method is able to construct informative summaries that achieve parameter inference performance on par with a fully-supervised baseline.

While follow-up studies are necessary before deploying our pipeline on real cosmological observations, with the influx of large amounts of complex, high-dimensional survey data and simulations products, as well rapid advances in machine learning, self-supervised learning methods such as VICReg offer a promising way to enable fast, efficient, and robust cosmological analyses.

Software and Data

Code used to reproduce the results of this paper is available at https://github.com/AizhanaAkhmet/SSL_for_Cosmology. This research made extensive use of the `Jupyter` (Kluyver et al., 2016), `Matplotlib` (Hunter, 2007), `Numpy` (Harris et al., 2020), `powerbox` (Murray, 2018), `pyccl`[1] (Chisari et al., 2019), `Pylians` (Villaescusa-Navarro, 2018), `PyTorch` (Paszke et al., 2019b), `PyTorch-Lightning` (Falcon et al., 2020), and `Scipy` (Virtanen et al., 2020) packages.

Acknowledgements

This work was partially supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, http://iaifi.org/). This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under grant Contract Number DE-SC0012567.

References

Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Prieto, C. A., Annis, J., Bailey, S., Balland, C., Ballester, O., et al. The desi experiment part i: science, targeting, and survey design. *arXiv preprint arXiv:1611.00036*, 2016.

Alsing, J. and Wandelt, B. Generalized massive optimal data compression. , 476(1):L60–L64, May 2018. doi: 10.1093/mnrasl/sly029.

Alsing, J. and Wandelt, B. Nuisance hardened data compression for fast likelihood-free inference. *Monthly Notices of the Royal Astronomical Society*, 488(4):5093–5103, 2019.

Bardes, A., Ponce, J., and LeCun, Y. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv preprint arXiv:2105.04906*, 2021.

Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Campbell, L., Parker, Q., Saunders, W., and Watson, F. The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *Monthly Notices of the Royal Astronomical Society*, 416(4):3017–3032, 09 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.19250.x. URL https://doi.org/10.1111/j.1365-2966.2011.19250.x.

Charnock, T., Lavaux, G., and Wandelt, B. D. Automatic physical inference with information maximizing neural networks. , 97(8):083004, April 2018. doi: 10.1103/PhysRevD.97.083004.

Chen, Y., Zhang, D., Gutmann, M., Courville, A., and Zhu, Z. Neural Approximate Sufficient Statistics for Implicit Models. *arXiv e-prints*, art. arXiv:2010.10079, October 2020. doi: 10.48550/arXiv.2010.10079.

Chisari, N. E., Alonso, D., Krause, E., Leonard, C. D., Bull, P., Neveu, J., Villarreal, A. S., Singh, S., McClintock, T., Ellison, J., Du, Z., Zuntz, J., Mead, A., Joudaki, S., Lorenz, C. S., Tröster, T., Sanchez, J., Lanusse, F., Ishak, M., Hlozek, R., Blazek, J., Campagne, J.-E., Almoubayyed, H., Eifler, T., Kirby, M., Kirkby, D., Plaszczynski, S., Slosar, A., Vrastil, M., Wagoner, E. L., and LSST Dark Energy Science Collaboration. Core Cosmology Library: Precision Cosmological Predictions for LSST. , 242(1):2, May 2019. doi: 10.3847/1538-4365/ab1658.

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., and Taylor, K. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. , 362(2):505–534, September 2005. doi: 10.1111/j.1365-2966.2005.09318.x.

Collaboration, L. D. E. S. et al. Large synoptic survey telescope: dark energy science collaboration. *arXiv preprint arXiv:1211.0310*, 2012.

Davé, R., Anglés-Alcázar, D., Narayanan, D., Li, Q., Rafieferantsoa, M. H., and Appleby, S. SIMBA: Cosmological simulations with black hole growth and feedback. , 486(2):2827–2849, June 2019. doi: 10.1093/mnras/stz937.

Dillon, B. M., Kasieczka, G., Olischlager, H., Plehn, T., Sorrenson, P., and Vogel, L. Symmetries, safety, and self-supervision. *SciPost Physics*, 12(6):188, 2022.

Dillon, B. M., Mastandrea, R., and Nachman, B. Self-supervised anomaly detection for new physics. , 106(5):056005, September 2022. doi: 10.1103/PhysRevD.106.056005.

Eisenstein, D. J. and Hu, W. Power Spectra for Cold Dark Matter and Its Variants. , 511(1):5–15, January 1999. doi: 10.1086/306640.

Falcon, W. et al. Pytorchlightning/pytorch-lightning: 0.7.6 release, May 2020. URL https://doi.org/10.5281/zenodo.3828935.

[1] https://github.com/LSSTDESC/CCL

Harris, C. R. et al. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Hayat, M. A., Stein, G., Harrington, P., Lukić, Z., and Mustafa, M. Self-supervised Representation Learning for Astronomical Images. , 911(2):L33, April 2021. doi: 10.3847/2041-8213/abf2c7.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Heavens, A. F., Jimenez, R., and Lahav, O. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, 317(4):965–972, 2000.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

Jeffrey, N., Alsing, J., and Lanusse, F. Likelihood-free inference with neural compression of DES SV weak lensing map statistics. , 501(1):954–969, February 2021. doi: 10.1093/mnras/staa3594.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kluyver, T. et al. Jupyter notebooks - a publishing format for reproducible computational workflows. In *ELPUB*, 2016.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J.-L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., et al. Euclid definition study report. *arXiv preprint arXiv:1110.3193*, 2011.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Murray, S. G. powerbox: A python package for creating structured fields with isotropic power spectra. *Journal of Open Source Software*, 3(28):850, 2018. doi: 10.21105/joss.00850. URL https://doi.org/10.21105/joss.00850.

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, 2019a.

Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019b. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance.pdf.

Percival, W. J., Verde, L., and Peacock, J. A. Fourier analysis of luminosity-dependent galaxy clustering. , 347(2):645–653, January 2004. doi: 10.1111/j.1365-2966.2004.07245.x.

Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., and Marinacci, F. Simulating galaxy formation with the IllustrisTNG model. , 473(3):4077–4106, January 2018. doi: 10.1093/mnras/stx2656.

Villaescusa-Navarro, F. Pylians: Python libraries for the analysis of numerical simulations. Astrophysics Source Code Library, record ascl:1811.008, November 2018.

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. , 915(1):71, July 2021a. doi: 10.3847/1538-4357/abf7ba.

Villaescusa-Navarro, F., Genel, S., Angles-Alcazar, D., Thiele, L., Dave, R., Narayanan, D., Nicola, A., Li, Y., Villanueva-Domingo, P., Wandelt, B., Spergel, D. N., Somerville, R. S., Zorrilla Matilla, J. M., Mohammad, F. G., Hassan, S., Shao, H., Wadekar, D., Eickenberg, M., Wong, K. W. K., Contardo, G., Jo, Y., Moser, E., Lau, E. T., Machado Poletti Valle, L. F., Perez, L. A., Nagai, D., Battaglia, N., and Vogelsberger, M. The CAMELS Multifield Dataset: Learning the Universe's Fundamental Parameters with Artificial Intelligence. *arXiv e-prints*, art. arXiv:2109.10915, September 2021b.

Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 2020. doi: https://doi.org/10.1038/s41592-019-0686-2.

Weinberger, R., Springel, V., Hernquist, L., Pillepich, A., Marinacci, F., Pakmor, R., Nelson, D., Genel, S., Vogelsberger, M., Naiman, J., and Torrey, P. Simulating galaxy formation with black hole driven thermal and kinetic feedback. , 465(3):3291–3308, March 2017. doi: 10.1093/mnras/stw2944.

Weltman, A., Bull, P., Camera, S., Kelley, K., Padmanabhan, H., Pritchard, J., Raccanelli, A., Riemer-Sørensen, S., Shao, L., Andrianomena, S., et al. Fundamental physics with the square kilometre array. *Publications of the Astronomical Society of Australia*, 37:e002, 2020.

Zablocki, A. and Dodelson, S. Extreme data compression for the cmb. *Physical Review D*, 93(8):083525, 2016.

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, pp. 12310–12320. PMLR, 2021.

A. CAMELS: SIMBA simulations suite

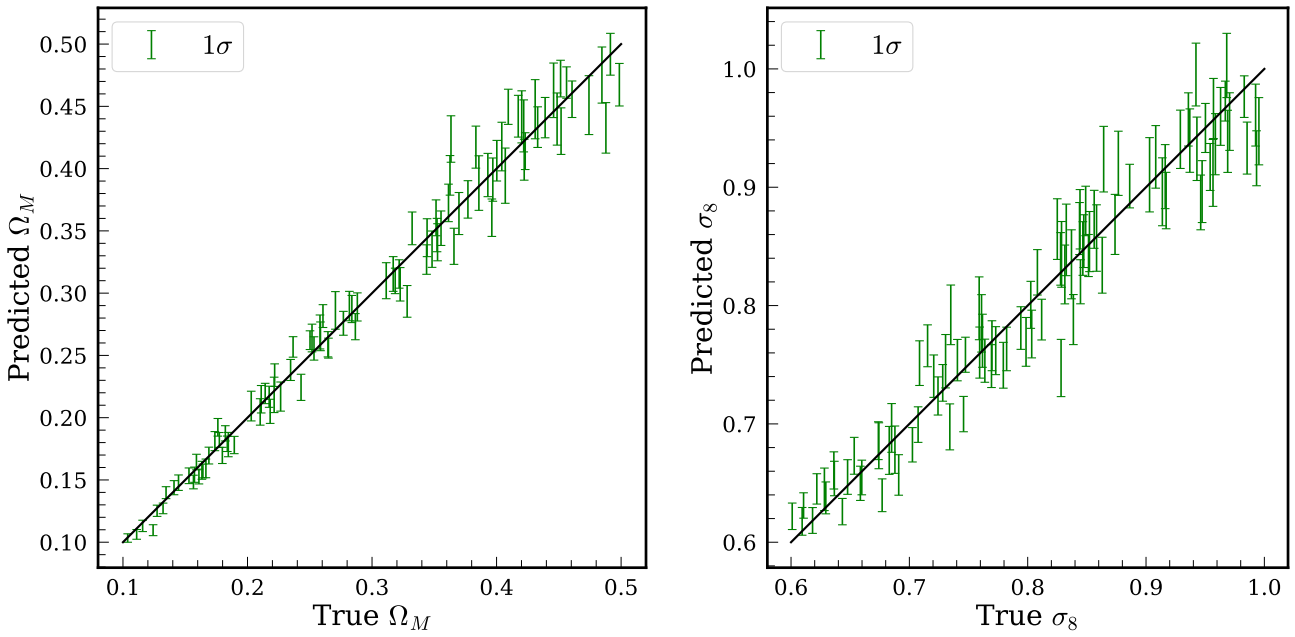

Figure 4. Predicted means and 1-σ uncertainties of cosmological parameters Ω_M and σ_8 compared to the true values of the parameters for total matter density maps from SIMBA simulations. Predictions for the means and variances of the parameters were obtained by training simple inference neural network on VICReg summaries.