Cosmological and astrophysical parameter inference with artificial intelligence

Olexandr Gugnin, Valerii Shchur, and Vladyslav Orlov
Taras Shevchenko National University of Kyiv

Anton Rudakovskyi and Musfer Adzhymambetov
Bogolybov Institute for Theoretical Physics of the NAS of Ukraine

Nazarii Drushchak and Roman Kovalchuk
Ukrainian Catholic University

Ivan Pyshnograiev
Igor Sikorsky Kyiv Polytechnic Institute

I. INTRODUCTION

The main goal of cosmological investigations is to advance the understanding of the universe's fundamental properties and its origin. Cosmology endeavours to address critical questions regarding the Universe's fate, its composition (including dark matter and dark energy), the formation and evolution of galaxies and clusters, and the precise measurement of cosmological parameters, taking into account the strong degeneracy between them. Through rigorous observational and theoretical investigations, cosmological research aims to refine existing models, validate theories, and uncover new insights into the nature of the cosmos. SBayesian inference is widely used to achieve this goal [1, 2]. However, applying popular Bayesian methods, such as Markov chain Monte Carlo or Nested Sampling, requires a lot of computational resources to compute the likelihood. A possible solution is provided by machine learning techniques (see, e.g., [3]), which allow the estimation of the posterior distributions of parameters through algorithms trained on a sample of simulated observables. The CAMELS [4] project aims to provide the opportunity to investigate the theoretical predictions for different observables, which will be obtained in upcoming observational missions. It is designed to sample the parameter space of cosmological and astrophysical models. In many cases, only statistical moments of the posterior are necessary. The momentum network [5] was proposed to estimate the mean and standard deviation for parameters from the observables. Momentum networks are used to estimate the cosmological parameters used in cosmological simulations [4, 6].

The aim of this project is improvement of the results of [4] via exploiting of the new parametrization and adding the prior knowledge. We use the modified version of the Jupyter Notebook in lectures by Francisco Villaescusa-Navarro[1]

[1] `https://github.com/franciscovillaescusa/ML_lectures`

II. METHODS

A. Data

In this work, we mostly focus on the SIMBA simulation [7] set with 1000 simulations performaed within CAMELS project [4], which were obtained via GRIZMO hydrodynamical code [8]. We also exploit the IllustriousTNG simulations (see, e.g., [9]) for testing of the robustness of our results. Each simulation was performed within different values of physical parameters $\{\Omega_m, \sigma_8, A_{SN1}, A_{SN2}, A_{AGN1}, A_{AGN1}\}$ (two cosmological and four astrophysical) and the initial random seed. The values of cosmological and astrophysical parameters are distributed as "latin" hypercube.

The σ_8 is the amplitude of the matter power spectrum, Ω_m is the matter contribution to total density of the Universe. A_{SN1}, A_{SN2} and A_{AGN1}, A_{AGN1} correspond to the supernova and active galactic nuclei feedbacks, that may expel the gas from the galaxies and supress the star formation. We further denote these 6 parameters $\vec{\theta}$[2].

Throughout this work, we use the star formation rate density (SFRD) as an observable value, which describes the mass density of the newly formed stars per time unit. We assume that the SFRD is a function of the astrophysical and cosmological parameters and redshift z:

$$\text{SFRD} = f\left(\vec{\theta}, z\right) , \tag{1}$$

and neglect the fact that SFRD in finite simulated volume also reflects the stochasticity of simulations, which may be significant for so small boxes.

The varying of different parameters has a diffrent impact on the SFRD. For example, at high ($z > 2$) redshifts SFRD strongly depends on the cosmological parameters σ_8 and Ω_m, because the star formation dominates at newly collapsed halos. The increasing σ_8 or Ω_m leads

[2] Note, that while SIMBA and IllustriousTNG use the same notations for the astrophysical parameters, they corresponds to quantities in the different physical models, see more in [4]

to the increasing of the SFRD. A_{SN1} impacts at all redshifts, while A_{SN2} has a significant effect on the SFRD at $z < 4$ On the other hand, the parameters A_{GN1} and A_{AGN2} don't affect significantly on star formation rate density at high redshifts, see e.g. Figure 9 in [4].

B. Neural Network architecture

We exploit a neural network for an estimation of the parameters of the simulations and standard deviation of the estimated parameters from the SFRD:

$$\vec{\theta} = g(\mathrm{SFRD}(z)),$$

where g denotes the reverse function to $\mathrm{SFRD}(\vec{\theta})$, approximated by neural network.

Our model adopts a multi-layer perceptron network architecture, processing 100 evenly spaced values of SFRD across the redshift range of $z = 0$ to 7. In the output layer, we have 12 values (2 for each of the 6 parameters θ_i).

The hidden layers are constructed using m perceptrons, each equipped with a Leaky ReLU activation function featuring a slope factor of 0.2 and a dropout rate p. We employ the Adam optimizer with β parameters set to $0.5, 0.999$, along with a learning rate denoted as lr and weight decay wd. The total number of hidden layers, represented by n, is another key hyperparameter of the neural network.

The dataset comprises 1000 simulations from both SIMBA and IllustrisTNG, randomly partitioned into training, validation, and test subsets according to a $0.7, 0.3, 0.3$ ratio. The hyperparameters n, m, p, lr, and wd are selected based on model performance on the validation subset. All the results presented in this paper were generated using the same optimized via grid-search set of hyperparameters: a single hidden layer with $m = 150$ perceptrons, a learning rate of $\mathrm{lr} = 10^{-4}$, weight decay $\mathrm{wd} = 0.1$, and a dropout rate of 0.2.

C. Loss function

The objective of our study is to infer the first two moments of the marginal distribution $p(\vec{\theta}, \mathbf{X})$ for each parameter θ_i. These two moments are the mean μ_i, and variance σ_i, which can be calculated as follows:

$$\mu_i = \int_\theta p(\vec{\theta}, \mathbf{X})\ \theta_i\ d\theta_1 \cdots d\theta_6, \tag{2}$$

$$\sigma_i^2 = \int_\theta p(\vec{\theta}, \mathbf{X})\,(\theta_i - \mu_i)^2\, d\theta_1 \cdots d\theta_6. \tag{3}$$

In the construction of our neural network we follow the original paper [10], utilizing a loss function, $\mathcal{L}$, inspired by the moment networks described in [5].

$$\mathcal{L}\left[\{\theta_{i,j}\}\right] = \sum_{i=1}^{6} \log\left(\sum_{j\in\mathrm{batch}} \mathcal{P}(\theta_{i,j})\,(\theta_{i,j} - \mu_{i,j})^2\right) \tag{4}$$
$$+ \lambda \sum_{i=1}^{6} \log\left(\sum_{j\in\mathrm{batch}} \mathcal{P}\left(\theta_{i,j}\right)\left((\theta_{i,j} - \mu_{i,j})^2 - \sigma_{i,j}^2\right)^2\right).$$

Here, the $\mu_{i,j}$ denotes the estimated mean values of the parameters, and $\sigma_{i,j}$ corresponds to the estimated standard deviations of the estimated parameters.

In this equation, the inner sum is taken over a batch, while the outer sum is over six parameters $\vec{\theta}$. The term $\mathcal{P}(\theta_{i,j})$ represents prior knowledge about the parameters in the Universe, which we will elaborate on later. In the absence of specific information, we assume $\mathcal{P}(\theta_{i,j}) \equiv 1$ to be uniformly equal to 1 in all simulations and similar to the [4, 6]. Finally, we introduce a hyperparameter λ to regulate the model's bias towards any of the moments.

We confirm the previous findings [4], that the estimations of σ_8 are highly inaccurate. Thereby, we use a new parameter, $S_8 = \sigma_8 \Omega_m^{1/2}$ [3], which is physically motivated by the fact that if the S_8 is the same, the galaxy clusters will have the roughly same velocity dispersions [11].

Now, for the modified set of parameters $\vec{\theta}^{'}$ with the substitution $\sigma_8 \to S_8$ we initially employ the same loss function $\mathcal{L}\left[\{\theta^{'}_{i,j}\}\right]$. Following that, we train our model with a prior distribution that depends solely on S_8

$$\mathcal{P}\left(\theta_{i,j}\right) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\left(S_{8j} - \hat{S_8}\right)^2 / 2\hat{\sigma}^2\right). \tag{5}$$

Here we manually chose $\hat{S_8} = 0.8(0.3)^{1/2} \approx 0.44$ based on our prior knowledge of approximate values for Ω_8 and σ_8 in the Universe, see e.g. [12].

III. RESULTS

A. Inference with new parametrization and priors

We train our model described in previous section on the SFRDs from SIMBA simulation set. The losses calculated on the train and test sets are shown in Figure 1. It is clearly visible that we don't overfit the data.

To evaluate our model, we investigate the standard deviation of the posterior for each of the six parameters. To successfully infer the cosmological and astrophysical parameters of the model from the SFRD using the neural network, the corresponding standard deviations, denoted as σ_i, must be significantly smaller than

[3] Note that our definition differs by a factor of $1/\sqrt{0.3}$ from the common one.

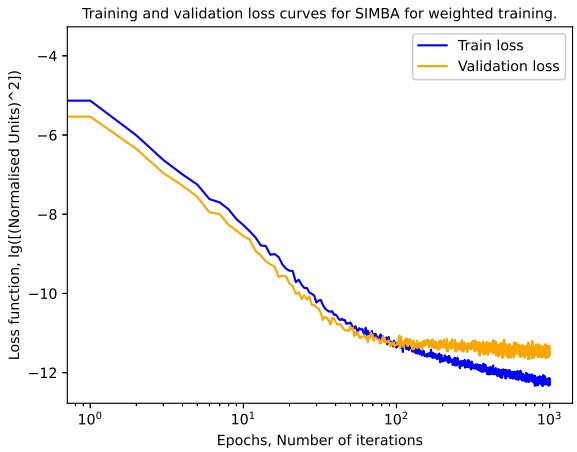

FIG. 1. The loss function calculated on the train and validation sets respectively using weighted training method.

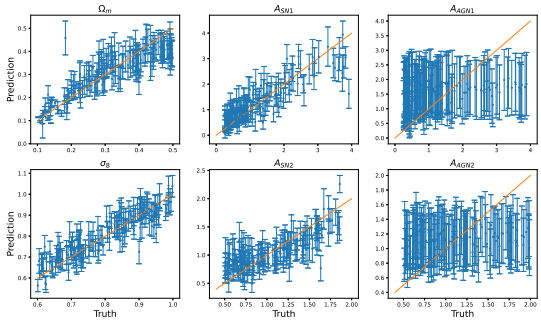

FIG. 2. Six parameters for a model trained on SIMBA with the original set of parameters $\vec{\theta}$, including σ_8.

the variation of the same parameter in the dataset. In Figure 2, we demonstrate that the astrophysical parameters A_{AGN1} and A_{AGN2} cannot be accurately determined from the SFRD, which aligns with the original findings of the CAMELS project [10].

Our calculations with the S_8 parameter, instead of σ_8, show that the variation for the set of parameters $\theta^{'}$ is smaller overall. The corresponding results are presented in Figure 3. This might indicate that we are dealing with degeneration between several parameters. Indeed, in Figure 4, we show the correlation between the restored variance of Ω_m and σ_8 parameters. However, this correlation almost disappears if we substitute σ_8 with S_8.

We also evaluate the Mean Squared Error Rescaled for the variance (MSER) for each data point in the training set:

$$\mathrm{MSER}(\{\theta_i\}) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\frac{(\theta_{i,j}-\mu_{i,j})^2}{\sigma_{i,j}^2}}\,. \qquad (6)$$

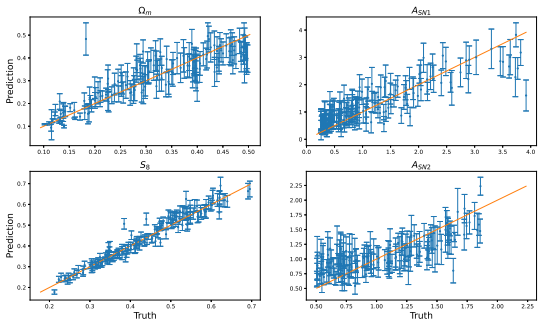

FIG. 3. Four parameters $\vec{\theta}^{'}$ including S_8, restored for SIMBA simulations.

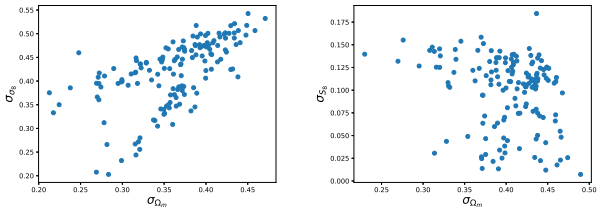

FIG. 4. Posterior variances of Ω_m vs σ_8 (left) and S_8 (right) for the test set data.

In this equation, the sum is taken over the test set. We anticipate MSER to be close to one, or ln(MSER) close to zero, for each of the six parameters. If it exceeds one, our model is biased toward the mean values, and if it is less than one, it is biased toward the standard deviation. These biases might arise from the structure of our loss function (4), which gives unequal weight to the first two moments. This imbalance can potentially be compensated for by the parameter λ, which will be discussed further. We demonstrate that the Neural Network trained without prior knowledge of the S_8 parameter exhibits a noticeable bias toward the standard deviations rather than the mean values of the posterior distribution. This finding is illustrated in the left part of Figure 5, where ln(MSER) for $(\hat{\sigma}, \lambda) = (2.5, 1)$, indicating no prior knowledge and no bias in the loss function, is negative. This deficiency can be addressed by either modifying prior knowledge (reducing $\hat{\sigma}$) or giving higher priority to the second moment in the loss function (increasing λ).

We also study a dependence of the quality of the predictions of our model on the parameters $\hat{\sigma}$ and λ. We provide a grid search on the following grids $\hat{\sigma} = \{0.025, 0.05, 0.125, 0.25, 2.5\}$ and $\lambda = \{0.5, 1.0, 2.0\}$. We summarize our findings in Figure 5. Our results showcase that the small $\hat{\sigma}$ leads to very poor fits.

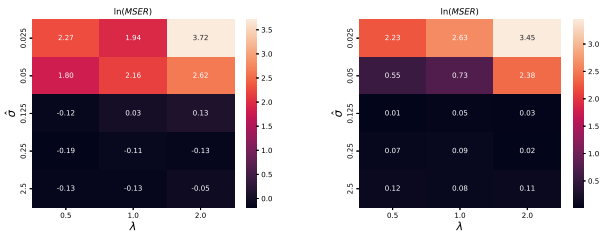

FIG. 5. Values of log(MSER) for the S_8 parameter calculated on the whole training set (left) and on a narrow subset near the prior value (right). Results are presented for a grid of λ and $\hat{\sigma}$ parameters, see Eqs. (4) and (5) respectively.

B. Robustness of the approach

The ability of predictive models to adapt to different underlying data distributions is of paramount importance. Models designed to understand complex cosmic phenomena must not only perform well on their native datasets but also exhibit robust generalization capabilities when exposed to alternative simulation environments. This section of the paper is dedicated to a comprehensive examination of the robustness and generalization capacities of two models with the same architecture but trained on distinct cosmological simulations: the IllustrisTNG and SIMBA datasets. The fundamental goal here is to evaluate how well these models, originally honed on their respective native data, can extrapolate their knowledge to datasets generated by a different cosmological simulation. This endeavor is laden with significance on multiple fronts. Firstly, it serves as a litmus test for understanding the limitations of our models. The discrepancy between the datasets necessitates us to scrutinize the models' adaptability and identify where they may excel or falter. Recognizing the boundaries of model generalization is not merely a scholarly pursuit; it has profound implications for the broader cosmological community and those who depend on these models for decision-making and further analysis. Secondly, by gauging the models' robustness in the context of different cosmological simulations, we gain insights into their potential for broader applicability. Models that can bridge the divide between datasets derived from various simulations hold the promise of more universally applicable cosmic insights and a deeper understanding of the underlying physics.

On the Figure 6, one may see the comparison of the performance of two of our models on datasets, which were produced by different from their training set, simulations. For this visualization, we have only showcased the set of cosmological parameters $(\Omega_{\rm m}, {\rm S}_8)$ without astrophysical. The reason for this is that SIMBA and IllustrisTNG have different subgrid physics models, and these parameters mean something different in each simulation. But in the case of cosmological parameters, things get better because, as one may notice, while the $\Omega_{\rm m}$ parameter is

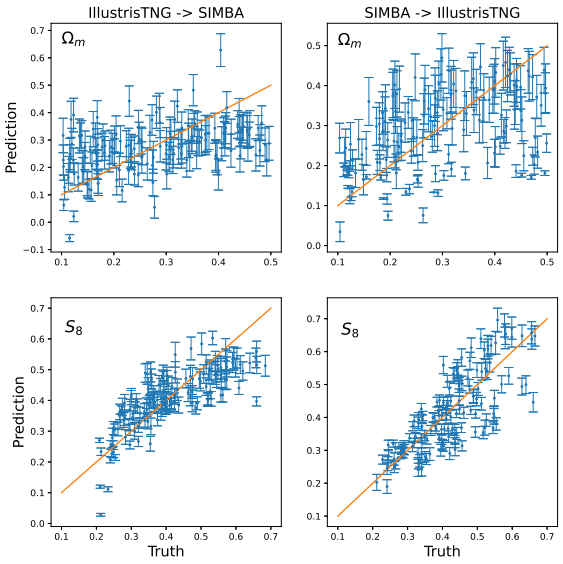

FIG. 6. Comparison of the performance of two the same models trained on different data on each other's samples. The left column shows truth/prediction plots for the model, trained on IllustrisTNG data run on SIMBA data. The right one highlights the opposite. For both parameters, it is evident that there is a notable disparity between the predictions and the ground truth when considering the $\Omega_{\rm m}$ parameter. In contrast, the performance is considerably enhanced when analyzing the ${\rm S}_8$ parameter.

hard for the models to be generalized, our ${\rm S}_8$ parameter might be predicted pretty accurately in some cases. This outcome is both intriguing and groundbreaking, as it marks a departure from prior research [10], where the σ_8 parameter exhibited a notable lack of generalization (it could not be predicted by a model that wasn't trained on specified simulation data).

C. Redshift threshold

For the possible improvement of results, we have performed a training on high redshifts. At $z > 2.5$, the universe was in its earlier stages, and many galaxies were actively forming stars. This era is characterized by intense star formation activity in newly formed galaxies. The rate at which stars are formed is directly affected by various cosmological parameters, including ${\rm S}_8$ [6].

By concentrating on high-redshift data, we have effectively focused on a cosmic epoch where the correlation between the star formation rate and cosmological parameters, such as S_8, is more pronounced. This is because the conditions in the early universe were more sensitive to changes in these parameters, leading to a stronger relationship between the two.

The decision to train the model specifically on high-redshift data allowed us to exploit the enhanced sensitivity of galaxy properties, like star formation rates, to changes in cosmological parameters. This, in turn, improved our ability to infer or constrain these parameters based on simulated data. As a result, for the MSER metric we have achieved slightly better values for pair of cosmological parameters - $\mathrm{MSER}^{\mathrm{rs}}_{\mathrm{S_8}} = 1.03$ vs $\mathrm{MSER}_{\mathrm{S_8}} = 1.1$, $\mathrm{MSER}^{\mathrm{rs}}_{\Omega_{\mathrm{m}}} = 1.28$ vs $\mathrm{MSER}_{\Omega_{\mathrm{m}}} = 1.36$ as well as for four astrophysical.

IV. CONCLUSION

In this project, we test the ability of a simple, fully connected perceptron to reconstruct the posterior means and variations for cosmological and astrophysical parameters from the star-formation rate density in SIMBA and IllustrousTNG simulations. We follow the logic of [4], which employed a moment's neural network and used the same observable. However, our model has significant differences:

- Instead of the amplitude of the matter power spectrum σ_8, we reconstruct $S_8 = \sigma_8 \left(\Omega_m\right)^{0.5}$.
- We incorporate a loss function with Gaussian weights that depend on S_8 for the samples, allowing us to include prior knowledge into our network.
- We vary the relative contribution of the loss function term corresponding to the dispersion.

In this work, we focus mostly on the SIMBA simulations. Our results showcase that the physically motivated parametrization may allow to break a strong degeneracy between posteriors of parameters. The predicted posterior mean value of S_8 aligns much better with the true values than σ_8. Simultaneously, the posterior dispersions for S_8 are much tighter than those for σ_8. We did not find any significant improvements when utilizing priors without the extension of the train set. Notably, our network successfully predicted the cosmological parameters via SFRD from IllustriousTNG simulations, despite differences in underlying physical models. Similar quality in cosmological inference with neural networks trained on different simulation codes was found when employing detailed hydrodynamical simulation fields (e.g., [10]) or halo catalogs [13].

This work could have a natural extension, for example, in applying a similar procedure to the high-redshift ultraviolet luminosity functions (UV LFs). At high redshifts, newly formed galaxies dominate, making the UV LFs highly dependent on the SFRD. Furthermore, it has been demonstrated that UV LFs can serve as a powerful probe for the clustering properties of matter (see, e.g., [14, 15]). However, previous studies employed MCMC methods, which are considerably more computationally expensive than the simulation-based inference approaches.

[1] R. Trotta, Bayesian Methods in Cosmology, ArXiv e-prints , arXiv:1701.01467 (2017), arXiv:1701.01467 [astro-ph.CO].

[2] U. von Toussaint, Bayesian inference in physics, Rev. Mod. Phys. **83**, 943 (2011).

[3] O. Lahav, Deep machine learning in cosmology: Evolution or revolution? (2023), arXiv:2302.04324 [astro-ph.CO].

[4] F. Villaescusa-Navarro *et al.* (CAMELS), The CAMELS project: Cosmology and Astrophysics with MachinE Learning Simulations, Astrophys. J. **915**, 71 (2021), arXiv:2010.00619 [astro-ph.CO].

[5] N. Jeffrey and B. D. Wandelt, Solving high-dimensional parameter inference: marginal posterior densities & Moment Networks, in *34th Conference on Neural Information Processing Systems* (2020) arXiv:2011.05991 [stat.ML].

[6] F. Villaescusa-Navarro *et al.* (CAMELS), The CAMELS Multifield Data Set: Learning the Universe's Fundamental Parameters with Artificial Intelligence, Astrophys. J. Supp. **259**, 61 (2022), arXiv:2109.10915 [cs.LG].

[7] R. Davé, D. Anglés-Alcázar, D. Narayanan, Q. Li, M. H. Rafieferantsoa, and S. Appleby, SIMBA: Cosmological simulations with black hole growth and feedback, MNRAS **486**, 2827 (2019), arXiv:1901.10203 [astro-ph.GA].

[8] P. F. Hopkins, A new class of accurate, mesh-free hydrodynamic simulation methods, Mon. Not. Roy. Astron. Soc. **450**, 53 (2015), arXiv:1409.7395 [astro-ph.CO].

[9] A. Pillepich *et al.*, Simulating Galaxy Formation with the IllustrisTNG Model, Mon. Not. Roy. Astron. Soc. **473**, 4077 (2018), arXiv:1703.02970 [astro-ph.GA].

[10] F. Villaescusa-Navarro *et al.*, Multifield Cosmology with Artificial Intelligence, (2021), arXiv:2109.09747 [astro-ph.CO].

[11] J. A. Peacock, *Cosmological Physics* (1999).

[12] P. A. R. Ade *et al.* (Planck), Planck 2015 results. XIII. Cosmological parameters, Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].

[13] R. Friedman and S. Hassan, HIGlow: Conditional Normalizing Flows for High-Fidelity HI Map Modeling, arXiv e-prints , arXiv:2211.12724 (2022), arXiv:2211.12724 [astro-ph.CO].

[14] N. Sabti, J. B. Muñoz, and D. Blas, Galaxy luminosity function pipeline for cosmology and astrophysics, Phys. Rev. D **105**, 043518 (2022), arXiv:2110.13168 [astro-ph.CO].

[15] N. Sabti, J. B. Muñoz, and D. Blas, New Roads to the Small-scale Universe: Measurements of the Clustering of Matter with the High-redshift UV Galaxy Luminosity Function, Astrophys. J. Lett. **928**, L20 (2022), arXiv:2110.13161 [astro-ph.CO].