Constraining the Reionization History using Bayesian Normalizing Flow (Mach. Learn.: Sci.Technol. 1(2020) 035014)

Héctor J. Hortúa, Riccardo Volpi, Luigi Malagò

Romanian Institute of Science and Technology, Cluj Napoca, Romania
email: hortua.orjuela@rist.ro

Abstract

In this work we present the use of Bayesian Neural Networks (BNNs) to predict the posterior distribution for four astrophysical and cosmological parameters from the 21cm signal. Besides achieving state-of-the-art prediction performances, the proposed methods provide accurate estimation of parameters uncertainties and infer correlations among them. Additionally, we demonstrate the advantages of Normalizing Flows (NF) combined with BNNs, being able to model more complex output distributions and thus capture key information as non-Gaussianities. Finally, we propose novel calibration methods employing Normalizing Flows after training, to produce reliable predictions, and we demonstrate the advantages of this approach both in terms of computational cost and prediction performances.

1 The 21cm signal

The 21cm signal from the neutral hydrogen in the intergalactic medium (IGM) is described through its brightness temperature contrast, δT_b, relative to the CMB [1]

$$\delta T_b \approx 3.3(1+\delta_m)x_{HI}\Big(\frac{T_S - T_\gamma(z)}{T_S}\Big)\Big(\frac{\omega_b h^2}{0.023}\Big)\Big(\frac{1+z}{\Omega_m h^2}\Big)^{\frac{1}{2}}\mathrm{mK}, \quad (1)$$

where T_S and $T_\gamma(z)$ are the gas spin and the CMB temperatures at redshift z respectively, δ_m the density contrast of baryons, and x_{HI} denotes the neutral fraction of hydrogen. The contrast density strongly depends on the cosmological parameters, while the HI ionized field parametrized by x_{HI} is determined via different astrophysical parameters. Therefore, 21cm observations have the potential not only to constrain fundamental cosmology, but also to provide another window into the properties of the IGM and the first galaxies and stars [1].

2 Variational Inference and BNN

We will focus on a variational inference approach which approximates the posterior distribution $p(\mathbf{w}|\mathcal{D})$ by an variational distribution $q(\mathbf{w}|\theta)$, depending on a set of parameters θ [2]. The objective can then be formalized as finding θ that makes q as close as possible to the true posterior, for instance by minimizing the KullBack-Leibler (KL) divergence between the two distributions [2]

$$\mathrm{KL}(q(\mathbf{w}|\theta)||p(\mathbf{w}|\mathcal{D})) \equiv \int_\Omega q(\mathbf{w}|\theta)\ln\frac{q(\mathbf{w}|\theta)}{p(\mathbf{w}|\mathcal{D})}d\mathbf{w}. \quad (2)$$

Using the Bayes theorem, one finds that minimizing Eq. (2) is equivalent to minimizing the following objective function

$$\mathrm{KL}(q(\mathbf{w}|\theta)||p(\mathbf{w})) - \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{D}}\int_\Omega q(\mathbf{w}|\theta)\ln p(\mathbf{y}|\mathbf{x},\mathbf{w})d\mathbf{w}. \quad (3)$$

If the network is minimized at $\hat{\theta}$, the probability distribution of y^* for a new input x^* can be written as [2]

$$q_{\hat{\theta}}(\mathbf{y}^*|\mathbf{x}^*) = \int_\Omega p(\mathbf{y}^*|\mathbf{x}^*,\mathbf{w})q(\mathbf{w}|\hat{\theta})d\mathbf{w}, \quad (4)$$

while the covariance of the variational predictive distribution, for a fixed x^* is [3, 4]

$$\mathrm{Cov}_{q_{\hat{\theta}}}(\mathbf{y}^*,\mathbf{y}^*|\mathbf{x}^*) \equiv \mathbb{E}_{q_{\hat{\theta}}}[\mathbf{y}^*\mathbf{y}^{*\mathrm{T}}|\mathbf{x}^*] - \mathbb{E}_{q_{\hat{\theta}}}[\mathbf{y}^*|\mathbf{x}^*]\mathbb{E}_{q_{\hat{\theta}}}[\mathbf{y}^*|\mathbf{x}^*]^{\mathrm{T}}. \quad (5)$$

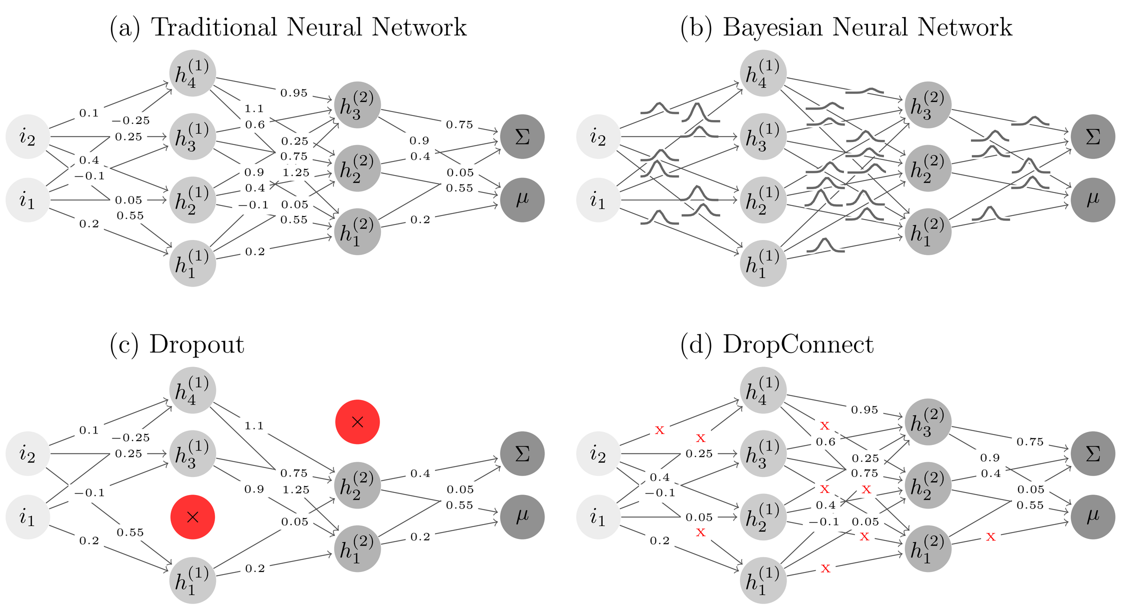

Figure 1: Representation of BNN [4].

3 Dataset

We have created 6000 brightness temperature images with resolution 1.5 Mpc through the semi-numerical code 21cmFast [5]. We varied two parameters corresponding to the cosmological context: the matter density parameter $\Omega_m \in [0.2, 0.4]$ and the rms linear fluctuation in the mass distribution on $8h^{-1}$Mpc $\sigma_8 \in [0.6, 0.8]$, and the other two parameters corresponding to the astrophysical context: the ionizing efficiency of high-z galaxies $\zeta \in [10, 100]$ and the minimum virial temperature of star-forming haloes $T^F_{vir} \in [3.98, 39.80] \times 10^4$K (hereafter represented in log10 units). For each set of parameters we have produced 20 images at different redshifts in the range $z \in [6, 12]$, and stacked these redshift-images into a single multi-channel tensor [4, 6].

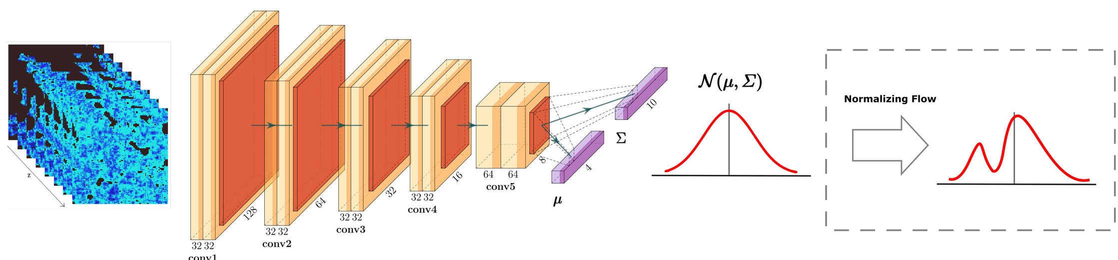

Figure 2: The images have size of $(128, 128, 20)$, where the channel stands for the redshift.

4 Implementation

We use Tensorflow [7] for building modified VGG16 architecture, along with tf.probabilities for adding a multivariate normal distribution in the last layer of the NN, and the NN is trained with Adam-optimizer. The last layer is dense with output corresponding to a multivariate Gaussian distribution. We trained the network both with and without the Normalizing Flows.

5 Metrics

We quantify the performance of the network by its high prediction of the parameters and accurate uncertainties. The high prediction is evaluated via coefficient of determination

$$R^2 = 1 - \frac{\sum_i(\bar{\mu}(x_i) - y_i)^2}{\sum_i(y_i - \bar{y})^2} \quad (6)$$

where $\bar{\mu}(x_i)$ is the prediction of the trained network , $\bar{y}$ is the average of the true parameters and the summations are performed over the entire test set. On the other hand, well calibrated networks means that y_i should fall in a $\beta\%$ confidence interval approximately $\beta\%$ of the time, where $\beta = \{68.3, 95.5, 99.7\}$ corresponding to 1, 2, and 3σ confidence levels of a normal distribution.

6 Calibration

1. Fine-tuning: Regularization for the scale of the variational distribution in Flipout.
2. Using NF (figure1), following two possible paths:
   (a) Retrain the best model found so far by including NF in the output of the network to minimize NLL.
   (b) Calibrate the network with a post-processing calibration approach, by fine-tuning the last layer of the network and minimizing again the NLL transformed by NF.

At the end, we will compare the resulting networks.

7 Results

We consider different kinds of Normalizing Flows acting on the output distribution of a BNN: the inverse autoregressive Flow(IAF), Masked Autoregressive Flow (MAF) and non-volume preserving flows (NVP). We observed that the R^2 are comparable for all methods, but the NLL is higher for the IAF, which means that this method tends to recover better accuracy in the uncertainties.

The improvement of the NLL leads to better calibrated networks, validated thought the coverage probabilities as we observe in the following table:

	IAF (NLL=-3.8)				MAF (NLL=-3.73)			
	σ_8	Ω_m	ζ	T^F_{vir}	σ_8	Ω_m	ζ	T^F_{vir}
R^2	0.94	0.98	0.87	0.98	0.94	0.98	0.87	0.98
C.L. 68.3%	66.0	64.0	69.2	65.4	64.7	63.7	69.1	65.0
C.L. 95.5%	94.0	94.0	95.0	94.0	93.3	94.2	95.1	94.0
C.L. 99.7%	99.2	99.2	99.5	99.6	99.0	99.3	99.3	99.4

7.1 Parameter constraint contours

In order to show the parameter intervals and contours from the Epoch of Reionization dataset, we choose randomly one example from the test set. The two-dimensional posterior distribution of the parameters are shown in Fig. 3 and the parameter 95% intervals are given by:

	σ_8	Ω_m	ζ	T^F_{vir}
IAF	$0.670^{+0.037}_{-0.033}$	$0.375^{+0.021}_{-0.021}$	$82.00^{+20.00}_{-10.00}$	$5.142^{+0.072}_{-0.070}$
MAF	$0.667^{+0.031}_{-0.029}$	$0.382^{+0.015}_{-0.016}$	$84.00^{+10.00}_{-10.00}$	$5.179^{+0.066}_{-0.068}$
Example true value	0.652	0.372	88.847	5.096

We can observe that after calibration, the contours produced by MAF becomes wider solving the underestimation found during training. Moreover, the contours of MAF and IAF applied in calibration overlap, and they are smaller compared with the base experiment, while Flows applied during training produce better results only for IAF.

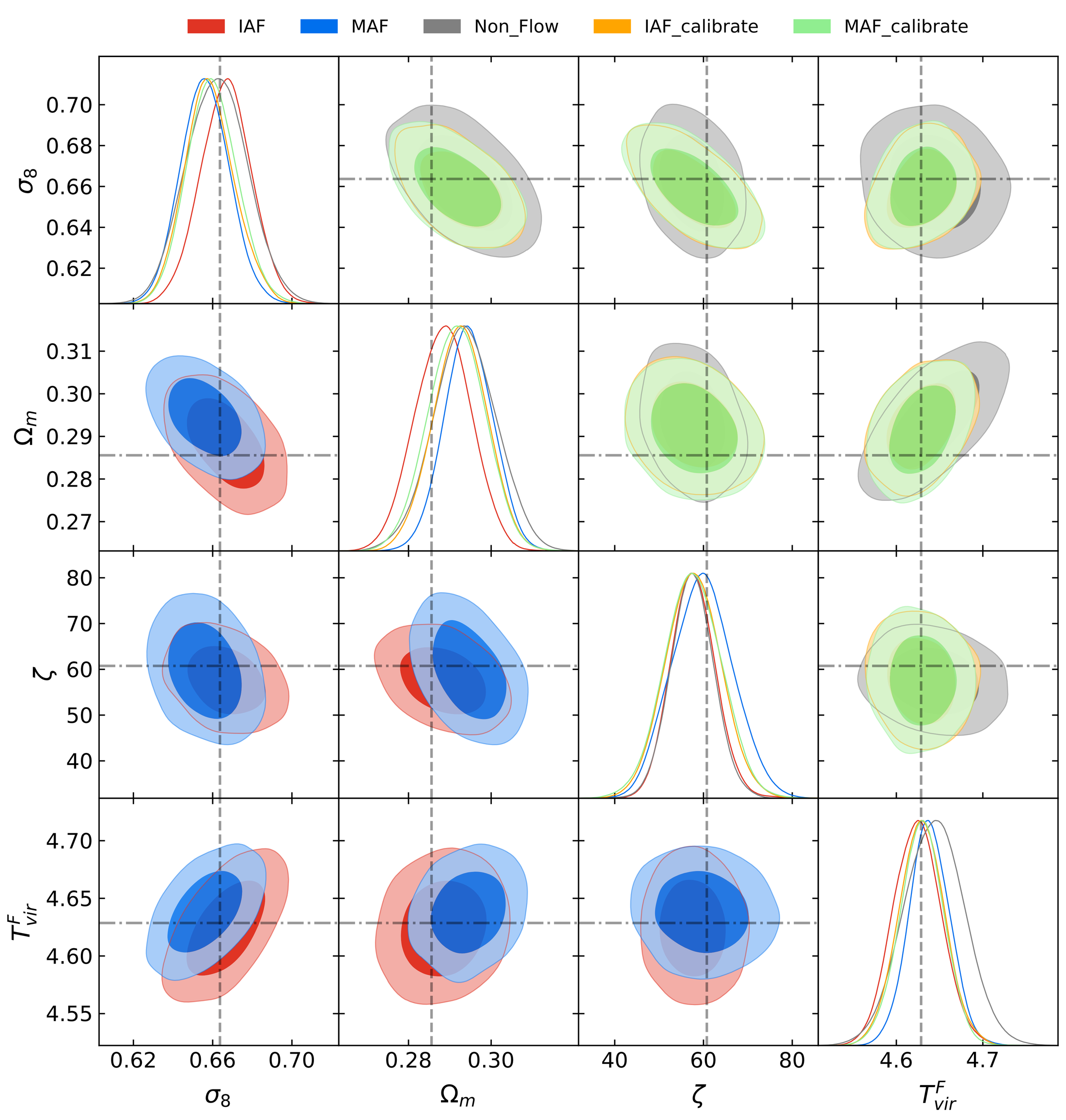

Figure 3: 68% and 95% contours from one example of our synthetic 21cm dataset.

8 Conclusions

We presented the first study using BNNs and Normalizing Flows to obtain credible estimates for cosmological parameters from 21cm signals. These methods offer alternative ways different from MCMC to make inference and recover the information in the 21cm observations.

References

[1] J. R. Pritchard, A. Loeb, Rep.Prog.Phys.75,086901(2012)
[2] Kendall, Alex and Gal, Yarin, NIPS, 5574-5584(2017), Y. Gal and Z. Ghahramani, ICLR 2016
[3] Y. Kwon, J.H. Won, B. Joon Kim, and M. Paik, ICMI with deep learning, 13(2018).
[4] Hortua et.al. Phys. Rev. D 102, 103509 (2020)
[5] A. Mesinger, S. Furlanetto, ApJ.669,2,663(2007)
[6] P. La Plante et.al., ApJ. 880(2):110, 2019.
[7] https://www.tensorflow.org/, https://www.tensorflow.org/probability