Inference of astrophysical parameters with a conditional invertible neural network

T Bister[1], M Erdmann[1], U Köthe[2] and J Schulte[1]

[1] RWTH Aachen University, III. Physikalisches Institut A, Otto-Blumenthal-Str., 52074 Aachen, Germany

[2] University of Heidelberg, Interdisciplinary Center for Scientific Computing, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

E-mail: josina.schulte@rwth-aachen.de

Abstract. Conditional Invertible Neural Networks (cINNs) provide a new technique for the inference of free model parameters by enabling the creation of posterior distributions. With these distributions, the mean parameter values, their uncertainties and the correlations between the parameters can be estimated. In this contribution, we summarize the functionality of cINNs, which are based on normalizing flows, and present the application of this new method to a scenario from astroparticle physics. We show that it is possible to constrain properties of the currently unknown sources of ultra-high-energy cosmic rays and compare the posterior distributions obtained with the network to those acquired using the classic Markov Chain Monte Carlo method.

1. Introduction

Often, measurements in physics can be described by a mathematical or simulation-based model with free parameters. In this scenario, the aim is to reconstruct these parameters using the measurements. In the Bayesian formulation of parameter inference, the parameters θ of a model are constrained by posterior distributions $p(\theta|y)$ given the observation y. They can be related to the likelihood $\mathcal{L}(y|\theta)$ and a prior distribution $p(\theta)$ using Bayes theorem $p(\theta|y) \propto \mathcal{L}(y|\theta)p(\theta)$. The likelihood describes the probability of the observation given specific model parameters, whereas the prior distribution includes prior knowledge of the parameter space. The correct modeling of the posterior distribution is crucial for Bayesian parameter estimation. Unfortunately, usually, when the model's likelihood is not analytically calculable, accordingly, the posterior is not tractable. Hence, one resorts to techniques for approximation, like Markov chain Monte Carlo (MCMC) [1]. The MCMC technique is based on efficient sampling, which is mathematically shown to converge to the correct posterior distributions in the limit of infinite samples. A caveat is the theoretical limit of infinite sampling steps with which the convergence is only ensured. Alternative methods are also developed, ranging from different levels of expressiveness of the approximate posterior. Among these are variational methods, like variational inference [2] and normalizing flows [3, 4].

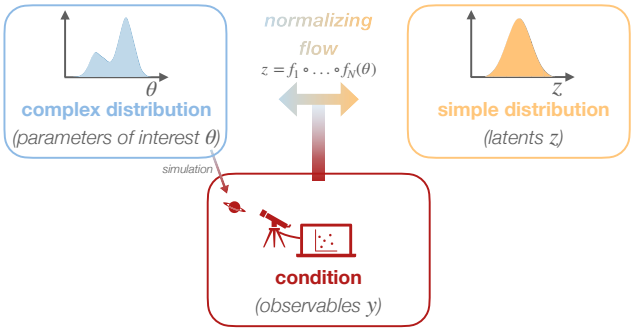

Figure 1: Schematic representation of a (conditional) normalizing flow. The transformation between a complex distribution *(left)* and a simple distribution *(right)* can be represented by a normalizing flow. If the complex distribution is a *conditional* probability distribution, a conditional normalizing flow can be used, using the condition as additional information. In Bayesian parameter inference, the complex distribution corresponds to the posterior distribution of the model parameters θ under the condition of the observables y. Note that, in general, the observables are related to the model parameters via simulations.

2. From normalizing flows to conditional Invertible Neural Networks

The aim of this work is the exploration of a new technique for the determination of posterior distributions, the so-called conditional Invertible Neural Network (cINN), which is based on the concept of normalizing flows.

2.1. Normalizing flows

In recent years, normalizing flows arose as a concept usable for the approximation and modeling of complex probability distributions given samples from that distribution [3]. With this, e.g., the generation of additional samples from a complex distribution is enabled. The basis is the transformation of a simple probability distribution with so-called latent variables z to the more complex distribution of the parameters of interest θ; see the upper part of Fig. 1 for a schematic representation. The transformation is employed by a composition of differentiable and invertible mappings, and the simple probability distribution for the latent space can be chosen freely, whereby often a standard normal distribution is adopted. Advantages of normalizing flows are their high expressive power as well as fast and efficient computation due to a tractable Jacobian [4].

2.2. Conditional normalizing flows

Additionally, it is possible to reconstruct conditional probability distributions with normalizing flows by extending the concept to a conditional setup. This procedure allows the approximation of posterior distributions $p(\theta|y)$ of the parameters of interest θ conditioned on an observation y. Here, the parameters of interest are mapped to the latent space, following a fixed and known distribution $p(z)$, with the observables y as additional information, see Fig. 1. There are different ways to implement a conditional normalizing flow: examples of architectures are masked autoregressive flows [5], variational autoencoders and cINNs [6, 7, 8]. The network structure of cINNs is based on affine coupling layers, introduced in [9, 10], of which several chained together constitute the total network. After the training, the network can be evaluated in the inverse direction given specific observables as well as samples from the known simple probability distribution, which is enforced on the latent space. With this, the network returns samples from the wanted posterior distributions of the parameters of interest. cINNs have already been applied to different problems, see, e.g., [8, 11, 12]. For the implementation, we use

the *Framework for Easily Invertible Architectures* [13], which is based on pytorch [14]. For a more detailed description of the network architecture and training procedure, see [15].

3. Scenario from astroparticle physics

The sources of ultra-high-energy cosmic rays are still unknown and, therefore, subject to ongoing research. Currently, observatories on Earth, like the Pierre Auger Observatory [16], measure three important quantities of extensive air showers induced by the cosmic rays reaching the Earth's atmosphere: the direction of arrival, the energy and the depth of shower maximum, which correlates with the initial cosmic-ray mass. These observables contain information about the origin of cosmic rays that is distorted by propagation effects such as energy loss and scattering as well as detector effects. The goal of this work is the reconstruction of original characteristics of the sources in a simple model using two of the three essential observables, the energy as well as the depth of shower maximum, by means of inversion with the cINN. The procedure is analogous to the analysis in [17], where traditional techniques were employed for the analysis. The underlying model includes homogeneously distributed identical sources, which isotropically emit cosmic rays. The emitted cosmic rays follow an energy spectrum of the form

$$J_{\mathrm{inj}}(E_{\mathrm{inj}}, A_{\mathrm{inj}}) = J_0 \cdot a(A_{\mathrm{inj}}) \Big(\frac{E_{\mathrm{inj}}}{10^{18}\ \mathrm{eV}}\Big)^{-\gamma} \cdot \begin{cases} 1 & Z_{\mathrm{inj}} R_{\mathrm{cut}} < E_{\mathrm{inj}} \\ \exp\left(1 - \frac{E_{\mathrm{inj}}}{Z_{\mathrm{inj}} R_{\mathrm{cut}}}\right) & Z_{\mathrm{inj}} R_{\mathrm{cut}} \geq E_{\mathrm{inj}} \end{cases} \tag{1}$$

with the spectrum normalization J_0, the fraction $a(A_{\mathrm{inj}})$ of injected cosmic rays with mass A_{inj} below the cutoff and the initial energy E_{inj}. The steepness of this power law is given by the spectral index γ until the start of the exponential cutoff at the maximum rigidity R_{cut}, whereby rigidity denotes energy per charge. We model this scenario using the CRPropa 3 software [18] to include propagation effects, namely interactions with the background radiation, nuclear decay and adiabatic energy losses, which affect and thus distort the emitted energy spectrum as well as the mass distribution until detection on Earth. We use five elements at the sources, namely protons, helium, nitrogen, silicon and iron. In total, we have six free model parameters θ: four representative values for the elemental fractions to always satisfy the side condition of sum to unity [19], as well as the spectral index γ and the maximum rigidity R_{cut} describing the form of the injected energy spectrum, see eq. 1.

3.1. Training of the cINN

We need to create training data to learn the connection between the source parameters θ and the observables y on Earth. For that, the aforementioned six free source parameters are varied using a flat prior for the values representing the elemental fractions, whereas the spectral parameters γ and R_{cut} are constrained by the region found in the previous analysis in [17].

The chosen observables are modeled to mimic the measurements at the Pierre Auger Observatory. For the energy spectrum on Earth, 17 logarithmic bins in the detected energy range $E_{\mathrm{det}} = 10^{18.7}\,\mathrm{eV}$ and $E_{\mathrm{det}} = 10^{20.4}\,\mathrm{eV}$ are used. For better distinction of characteristic features of the spectrum, it is multiplied with E_{det}^3. The depths of shower maximum X_{max} distributions are binned in two dimensions into 25 X_{max} bins between $550\,\mathrm{g/cm^2}$ and $1050\,\mathrm{g/cm^2}$ and similar energy bins as the spectrum, using a combined bin above $10^{19.6}\,\mathrm{eV}$. An example observation is shown in Fig. 2 where the source parameters are chosen to match the best-fit values of [17] in the CRPropa 3-based setup, namely a spectral index γ of 0.87, a rigidity cut-off of $10^{18.62}\,\mathrm{V}$, 88% nitrogen and 12% silicon. Bin-wise modifications of the modeled observation following a Poisson distribution are added to mimic the statistical fluctuations of the measurements. These modified observables are then used during the training, differing randomly in each epoch.

In total, we use 10^6 training sets and 10^5 validation sets and train the network for 13 hours until convergence.

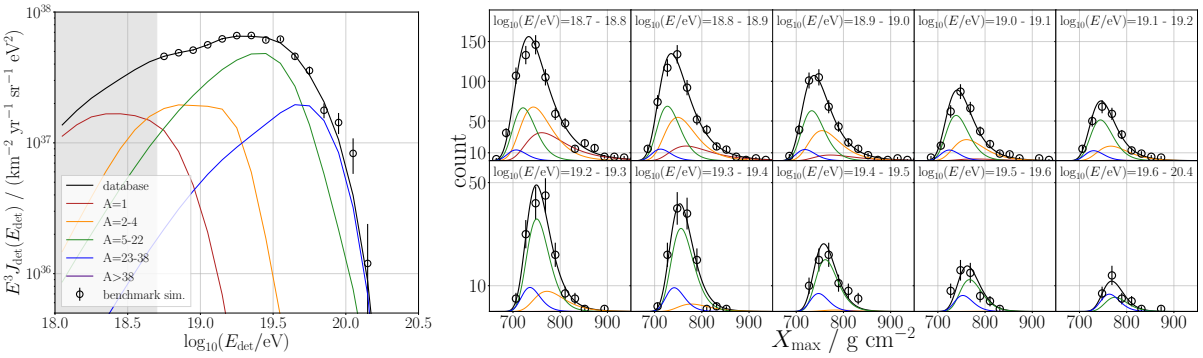

Figure 2: Observables on Earth (energy spectrum multiplied with $E^3_{\rm det}$ on the left and depth of shower maximum distributions on the right) corresponding to the best-fit source parameter values of [17]. The curves correspond to the simulations without additional statistical fluctuations *(database)*, the points show one realization of the resulting bin contents with the added statistical fluctuations *(benchmark simulation)*, also depicted with corresponding error bars. The colored curves show the contribution of different elemental groups to the total observed flux and the depth of shower maximum distributions, respectively.

4. cINN reconstruction quality

To evaluate the cINN reconstruction quality, we use the network to create posterior distributions for a test dataset of size 10^4. This extensive test of the method's accuracy is enabled by the speed of the posterior creation using the trained network. To assess the quality of the parameter estimator, we compare the mean of the posterior distribution of the parameters to the true value of the corresponding simulation. In Fig. 3, the posterior mean value of the spectral index and the rigidity cutoff are shown as a function of the true value. The reconstructed values lie close to the grey line indicating perfect agreement for both parameters. Quantitatively, the normalized root mean square error (NRMSE) [15] is below 2% for both parameters, affirming the good agreement between the estimator of the parameters from the posterior distributions obtained with the cINN and the true value of the simulations. Additionally, the accuracy of the widths of the posteriors as an uncertainty measure can be evaluated. We use the calibration error $e_{\rm cal}$ as in [11], defined as the difference between a confidence interval q and the fraction of test datasets where the true value of the parameter lies within this q-confidence interval of the cINN posterior distribution. The calibration error is calculated over a range of confidence intervals between 0.01 and 0.99 in steps of 0.01, and we report the median of the absolute values. The median calibration error is close to zero if the widths of the cINN posterior distributions are appropriate. We reach a value of 0.002 for both parameters, validating the accuracy of the widths and their use for uncertainty estimation. For an evaluation of the reconstruction of the elemental fractions, see [15].

For a further approval of the method, we compare the cINN posteriors for the specific benchmark scenario shown in Fig. 2 to posterior distribution obtained with a more traditional sampling technique, a Markov chain Monte Carlo. Here, sequential Monte Carlo sampling [20] in the *PyMC3* framework [21] is applied; see [15] for details. In Fig. 4, the results for the posterior distributions obtained with both methods for the two parameters γ and $R_{\rm cut}$ are shown. Generally, the posterior distributions including the correlation look very similar for the cINN and the MCMC. The uncertainties on both parameters estimated with the 68% highest posterior density interval are nearly identical. Further, the true simulation value lies approximately 1σ away from the mean value of the distribution for both parameters for both methods.

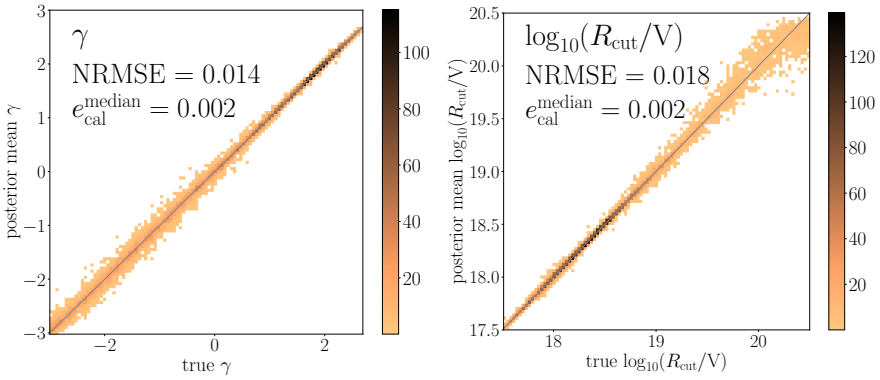

Figure 3: $2D$ histograms of the source parameters γ *(left)* and R_{cut} *(right)* using 10^4 test datasets. On the vertical axis the posterior mean value of the parameter is given, on the horizontal axis is the true value of the simulation. Perfect agreement is depicted by the gray straight line.

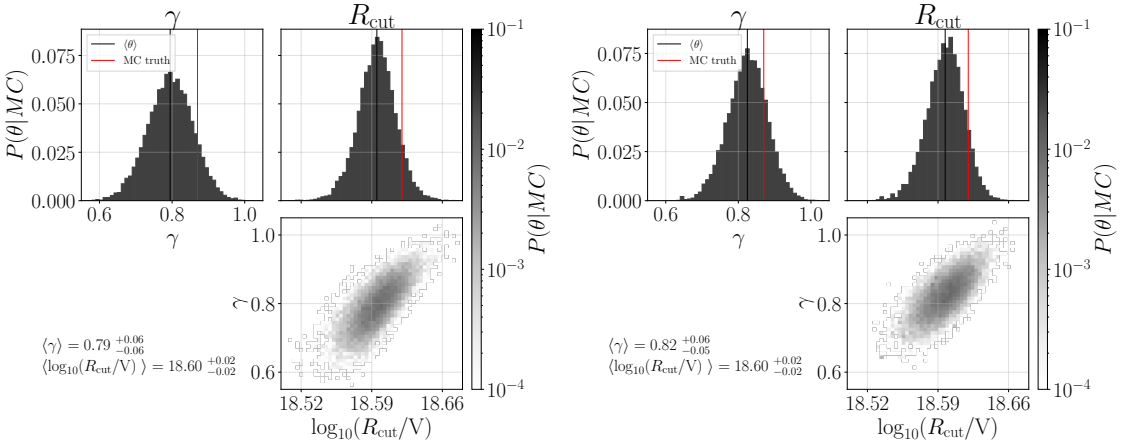

Figure 4: Posterior distributions for the source parameters γ and R_{cut} for the *benchmark simulation* obtained with the cINN *(left)* and the MCMC method *(right)*. The true value of the simulation is depicted with the red line and the mean of the posterior distribution with the black line.

5. Conclusion

We presented conditional Invertible Neural Networks as a technique based on normalizing flows usable for the approximation of posterior distributions of parameters conditioned on an observation. With these, not only an estimator of the parameter using, e.g., the mean of the distribution can be obtained, but also an uncertainty and correlations between parameters can be unveiled. We adopted this approach to reconstruct parameters of astrophysical sources of ultra-high-energy cosmic rays using simulated data of the observables energy and depth of shower maximum. We first evaluated the reconstruction quality of the method based on a test dataset regarding the mean and the widths of the posterior distributions finding appropriate accuracy. Additionally, the cINN method was compared to the more traditional approach of MCMC sampling on a single scenario where a very good agreement between the posterior distributions was acquired. Concluding, the conditional Invertible Neural Network is well-suited to perform Bayesian parameter estimation as a computationally efficient and reliable method. For more details and an extended description of the method and the application, see [15].

Acknowledgments
This work is supported by the Ministry of Innovation, Science, and Research of the State of North Rhine-Westphalia, and by the Federal Ministry of Education and Research (BMBF).

References
[1] Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A and Rubin D B 2013 *Bayesian Data Analysis* 3rd ed (CRC Press)
[2] Blei D M, Kucukelbir A and McAuliffe J D 2017 *Journal of the American Statistical Association* **112** 859–77 URL http://dx.doi.org/10.1080/01621459.2017.1285773
[3] Kobyzev I, Prince S J and Brubaker M A 2021 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43** 3964–79 URL http://dx.doi.org/10.1109/TPAMI.2020.2992934
[4] Papamakarios G, Nalisnick E, Rezende D J, Mohamed S and Lakshminarayanan B 2021 Normalizing Flows for Probabilistic Modeling and Inference URL http://jmlr.org/papers/v22/19-1028.html
[5] Papamakarios G, Pavlakou T and Murray I 2018 Masked Autoregressive Flow for Density Estimation (*arXiv* 1705.07057)
[6] Ardizzone L, Kruse J, Wirkert S, Rahner D, Pellegrini E W, Klessen R S, Maier-Hein L, Rother C and Köthe U 2019 Analyzing Inverse Problems with Invertible Neural Networks (*arXiv* 1808.04730)
[7] Radev S T, Mertens U K, Voss A, Ardizzone L and Köthe U 2020 BayesFlow: Learning complex stochastic models with invertible neural networks (*arXiv* 2003.06281)
[8] Ardizzone L, Lüth C, Kruse J, Rother C and Köthe U 2019 Guided Image Generation with Conditional Invertible Neural Networks (*arXiv* 1907.02392)
[9] Dinh L, Sohl-Dickstein J and Bengio S 2017 *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings* (OpenReview.net) (*arXiv* 1605.08803) URL https://openreview.net/forum?id=HkpbnH9lx
[10] Kingma D P and Dhariwal P 2018 Glow: Generative Flow with Invertible 1x1 Convolutions (*arXiv* 1807.03039)
[11] Ksoll V F, Ardizzone L, Klessen R, Koethe U, Sabbi E, Robberto M, Gouliermis D, Rother C, Zeidler P and Gennaro M 2020 *Monthly Notices of the Royal Astronomical Society* **499** 5447–85 URL http://dx.doi.org/10.1093/mnras/staa2931
[12] Radev S T, Graw F, Chen S, Mutters N T, Eichel V M, Bärnighausen T and Köthe U 2021 *PLOS Computational Biology* **17** e1009472 URL http://dx.doi.org/10.1371/journal.pcbi.1009472
[13] Ardizzone L *et al.* FrEIA [software] URL https://github.com/VLL-HD/FrEIA
[14] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L and Lerer A 2017
[15] Bister T, Erdmann M, Köthe U and Schulte J 2021 Inference of cosmic-ray source properties by conditional invertible neural networks (submitted to *EPJ* C) (*arXiv* 2110.09493)
[16] The Pierre Auger Collaboration 2015 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **798** 172–213 URL http://dx.doi.org/10.1016/j.nima.2015.06.058
[17] The Pierre Auger Collaboration 2017 *Journal of Cosmology and Astroparticle Physics* (04) 38 URL http://dx.doi.org/10.1088/1475-7516/2017/04/038
[18] Batista R A, Dundovic A, Erdmann M, Kampert K H, Kuempel D, Müller G, Sigl G, Vliet A v, Walz D and Winchen T 2016 *Journal of Cosmology and Astroparticle Physics* 038–038 URL http://dx.doi.org/10.1088/1475-7516/2016/05/038
[19] Onn S and Weissman I 2011 *Annals of Operations Research* **189** 331–42 URL https://doi.org/10.1007/s10479-009-0567-7
[20] Minson S E, Simons M and Beck J L 2013 *Geophysical Journal International* **194** 1701–26 URL https://doi.org/10.1093/gji/ggt180
[21] Salvatier J, Wiecki T V and Fonnesbeck C 2016 *PeerJ Computer Science* **2** e55