Bayesian Stokes inversion with Normalizing flows

C. J. Díaz Baso
Institute for Solar Physics, Dept. of Astronomy, Stockholm University
AlbaNova University Centre, SE-10691 Stockholm, Sweden
carlos.diaz@astro.su.se

A. Asensio Ramos
Instituto de Astrofísica de Canarias
C/Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain
Departamento de Astrofísica, Universidad de La Laguna
E-38206 La Laguna, Tenerife, Spain
andres.asensio@iac.es

J. de la Cruz Rodríguez
Institute for Solar Physics, Dept. of Astronomy, Stockholm University
AlbaNova University Centre, SE-10691 Stockholm, Sweden
jaime@astro.su.se

Abstract

Stokes inversion techniques are very powerful methods for obtaining information on the thermodynamic and magnetic properties of solar and stellar atmospheres. Most of the existing inversion codes are designed for finding the optimum solution to the nonlinear inverse problem. However, to obtain the location of potentially multimodal solutions, degeneracies, and the uncertainties of each parameter from the inversions, algorithms such as Markov chain Monte Carlo require to evaluate the model thousand of times. Variational methods are a quick alternative by approximating the posterior distribution by a parametrized distribution. In this study, we explore a highly flexible variational method, known as normalizing flows, to return Bayesian posterior probabilities for solar observations. We illustrate the ability of the method using a simple Milne-Eddington model and a complex non-LTE inversion. The training procedure need only be performed once for a given prior parameter space and the resulting network can then generate samples describing the posterior distribution several orders of magnitude faster than existing techniques.

1 Introduction

Through the analysis of spectra and their polarization, we have been able to infer the properties of the solar and stellar atmospheres. To infer the stratification of physical properties as a function of depth, we compare the emergent spectra given by a solar model using the radiative transfer theory with observations. This process is commonly known as spectropolarimetric inversion and nowadays is routinely used in solar physics. The traditional way for finding the optimum solution is the use of a gradient search minimization algorithm to drive the solution in the direction of the minimum difference between the forward calculated spectrum and the observed one.

To have complete knowledge of the parameter space (the location of the global minimum if it exists, whether there are degeneracies or multiple solutions that can equally reproduce the observations, and

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

to have a proper estimation of the uncertainty in the solution), the posterior probability distribution has to be calculated [2]. However, computing the posterior turns out to be complex and one has to rely on efficient stochastic sampling techniques such as Markov Chain Monte Carlo (MCMC; [19]) which require many forward calculations and are therefore computationally very costly. This is especially relevant in non-local thermodynamic equilibrium (NLTE) cases, for which the forward problem is computationally heavy and highly non-linear. Carrying out an inference of a standard observation with millions of pixels requires the use of supercomputers running parallelized inversion codes for many hours [e.g., 14]. A very promising alternative method for Bayesian inference is variational inference, where the true distribution of the solution is approximated. In this study, we use an advanced method for approximating probability distributions known as normalizing flows [26, 22, 7, 13] and perform Bayesian inference for the physical atmospheric parameters from spectroscopic data.

2 Normalizing flows

Normalizing flows can approximate the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$ by transforming a simple probability distribution $p_z(\mathbf{z})$ into a complex one by applying an invertible and differentiable transformation $\boldsymbol{\theta} = \mathbf{f}(\mathbf{z})$. In practice, we can construct a flow-based model by implementing $\mathbf{f} = \mathbf{f}_\phi$ with a neural network with parameters ϕ and take the base distribution $p_z(\mathbf{z})$ to be simple, typically a multivariate standard normal distribution. When these transformations are conditioned on the observations, they can approximate the posterior distribution of any observation. More precisely, the resulting probability distribution after the transformation is computed by applying the change of variables formula from probability theory [23]:

$$p_\phi(\boldsymbol{\theta}|\mathbf{x}) = p_z(\mathbf{z}) \left| \det \frac{\partial \mathbf{f}_\phi}{\partial \mathbf{z}} \right|^{-1} = p_z(\mathbf{f}_\phi^{-1}(\boldsymbol{\theta})) \left| \det \frac{\partial {\mathbf{f}_\phi}^{-1}}{\partial \boldsymbol{\theta}} \right| \tag{1}$$

where the first factor represents the probability density for the base distribution (p_z) evaluated at $\mathbf{f}_\phi^{-1}(\boldsymbol{\theta})$ and the second factor is the absolute value of the Jacobian determinant and accounts for the change in the volume due to the transformation. Normalizing flows can be trained by minimizing the negative log-likelihood with respect to the parameters of the flows ϕ for each pair of example $(\boldsymbol{\theta_i}, \mathbf{x_i})$ in the dataset of size D: $\mathcal{L}_\phi = -\frac{1}{D}\sum_{i=1}^{D} \log p_\phi(\boldsymbol{\theta_i}|\mathbf{x_i})$. This loss is typical in "neural posterior estimation", one of the goals of simulation-based inference [3]. Among the different families of transformations, we use a widely-used transformation known as "coupling spline flows" which has been demonstrated to be effective at representing complex densities, fast to train, and fast to evaluate [7, 20, 8]. It consists on a family of differentiable functions that are invertible and very expressive based on monotonically increasing splines. We use the implementation of normalizing flows in `PyTorch` [21] available in `nflows` [9]. We have used an architecture that is a concatenation of blocks of an invertible linear transformation using the LU-decomposition [16] with a rational-quadratic spline transform [RQ, 8] where a residual network [ResNet, 11] is used to calculate the parameters of the splines[1]. For the first simple case, a flow with 5 coupling transformations, 5 residual blocks, and 32 neurons per layer was enough. In the second case, we need at least 15 coupling layers, 10 residual blocks, and 64 neurons per layer. We trained the models for 500 epochs with a batch size of 100. We have used a learning rate of 10^{-4} and the Adam optimizer [15], reserving 10% of our training set for validation.

3 Simple case: Milne-Eddington atmosphere

As a first example, we show the capabilities of the normalizing flows in a case where the forward model is fast enough to allow comparison with a MCMC method. For this case, we choose the Milne-Eddington solution [1] of the radiative transfer problem as a baseline. Focusing only on Stokes I, $\boldsymbol{\theta}$ is five-dimensional: the macroscopic velocity v_{LOS}, the Doppler width Δv_{D}, the line-to-continuum opacity ratio η_0, and the two parameters of the source function S_0 and S_1. The normalizing flow is trained using 10^6 training pairs $(\boldsymbol{\theta_i}, \mathbf{x_i})$ by drawing $\boldsymbol{\theta_i}$ from an extended uniform prior for all the variables. We have simulated the photospheric Fe I 6301.5Å line using the Milne-Eddington model[2] and a Gaussian noise of $\sigma = 8 \cdot 10^{-3}$ in continuum units. Once the normalizing flow is trained

[1] `https://github.com/cdiazbas/bayesflows`

[2] `https://github.com/aasensio/milne`

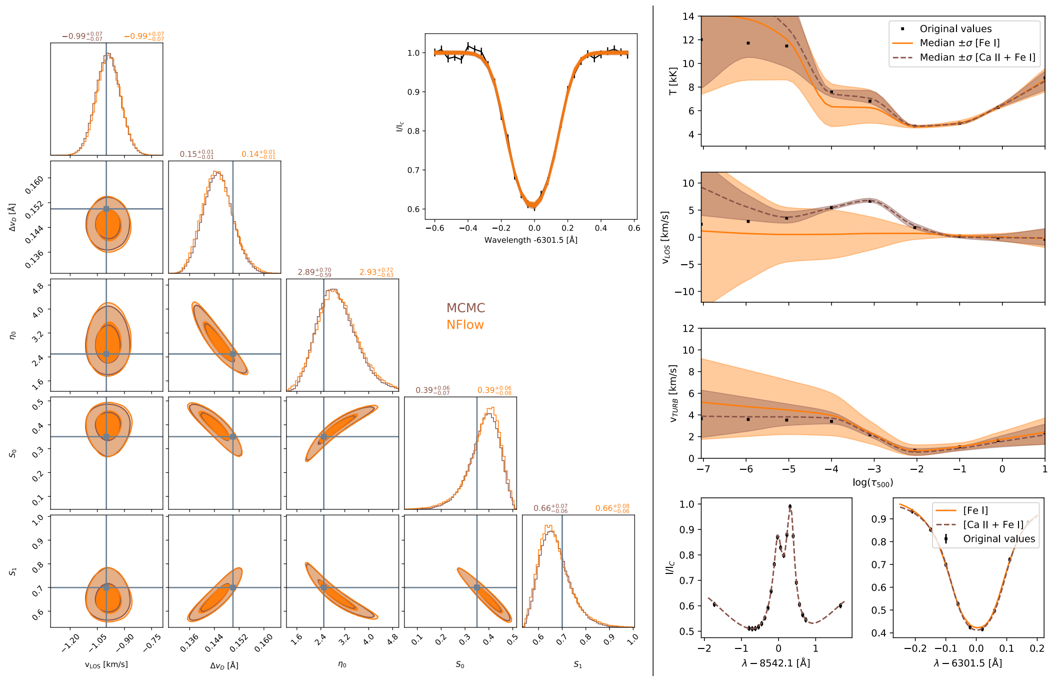

Figure 1: Left panel: Joint and marginal posterior distributions for the Milne-Eddington model. Contours are drawn at 1 and 2 sigmas. Right panel: Atmospheric stratification inferred in the NLTE case. The orange solution is inferred only using the Fe I line while the brown solution also uses the Ca II profile. The shaded regions mark the corresponding 68% confidence interval.

after ~12 hours, we carry out Bayesian inference for arbitrary observations. Given that the flows are almost 2 orders of magnitude faster than the MCMC, the training for inferring thousands of pixels is already worth it. To verify the accuracy of the posterior inference we compare the result against a MCMC computed with the `emcee` sampler [10] using a Gaussian likelihood. According to the results shown in the left panel of Fig. 1, both distributions are clearly in very close agreement. For strongly degenerate parameters (such as Δv_D and η_0) we recover the typical joint banana-shaped posterior, while for highly correlated parameters we find ridge-shaped distributions (like between S_0 and S_1). Given these strong degeneracies, an inversion method based on single-point estimations using neural networks will not perform correctly as there are multiple solutions for the same profile.

4 Complex case: NLTE with stratification

In this second example, we have simulated a case in which we have simultaneously observed the photospheric Fe I 6301.5Å line and the chromospheric Ca II 8542Å line. This configuration is commonly used to study events occurring both in the photosphere and the chromosphere [28, 14, 6, 29]. It is also one of the most common configurations on the Swedish 1-m Solar Telescope [SST, 24, 25]. To create a diverse set of samples of solar-like stratifications and intensity profiles, we have created 10^6 new stratifications by perturbing the solar stratifications inferred in [6]. The density and gas pressure stratifications are computed by assuming hydrostatic equilibrium (HE) and the spectra were synthesized using the NLTE code STiC [4, 5]. We have then degraded each spectral line to the spectral resolution of the CRISP instrument and with Gaussian noise with a standard deviation of 10^{-2} in units of the continuum intensity. The dimensionality of this case is much higher and depends on the height grid and the number of physical parameters.

We train two different models to capture the difference when more spectral lines are included. The first model only uses observations of the Fe I line, while the second model uses both the Fe I and Ca II lines together. The trained models are evaluated on a profile with strong emission in the chromospheric line (right panel of Fig. 1). The top three panels show the stratification with optical depth of the temperature, line-of-sight velocity, and turbulent velocity. The solid orange and dashed

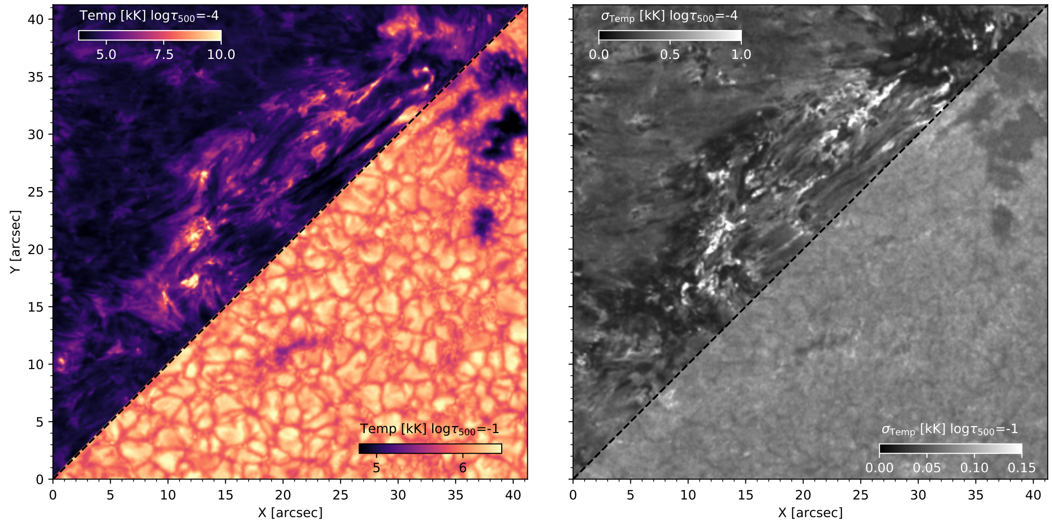

Figure 2: Atmospheric structure of the FOV as inferred from the inversion. The left panel shows the temperature at two layers for half of the FOV. The right panel shows the associated uncertainty for the same layers.

brown lines show the median value estimated from the posterior distribution when considering only the photospheric line or both lines, respectively. As expected, the inference that considers both lines can recover with high accuracy the whole stratification, whereas using only the Fe I line yields a model where only the photosphere is recovered, with a large uncertainty towards the upper atmosphere. This result shows that the normalizing flow is able to learn the range of sensitivity of each spectral line just by looking at the examples of the database.

5 Performance and validation

We have quantified how the accuracy of our models depends on the size of the training set. To this end, we use the fact that, when the models extracted from the posterior distribution are used to re-synthesize the line profiles, they should be distributed according to the assumed sampling distribution (a standard deviation of $8{\cdot}10^{-3}$ and 10^{-2} in units of the continuum intensity for the ME and NLTE case, respectively). These calculations show that the average error of the normalizing flow model decreases asymptotically with the size of the dataset, reaching the expected error with 10^3 examples for the Milne-Eddington case while 10^6 in the complex high-dimensional NLTE case.

We have also explored two possible procedures to reduce the effect of the size of the training set. The first procedure relies on compression to reduce the dimensionality of the forward model by using an autoencoder [12] and it shows that a more compact representation helps the normalizing flow to train faster and perform better. The second procedure reuses the samples from the normalizing flow and reweights them using importance sampling to produce a better approximation to the posterior distribution. This requires to have access to the forward model which is time-consuming in NLTE cases, but a pre-trained neural network that works as an emulator of the forward model alleviates this problem.

Finally, we have also tested the trained normalizing flows on large fields of view. For that, we have chosen the observations analyzed in [17] observed with the SST on 2016-09-19 at around 09:30 UT. We have applied the neural network to a field of view of approximately 42×42 arcseconds (around $5 \cdot 10^5$ pixels). Spectra from individual pixels are analyzed independently. The normalizing flow was able to produce the posterior distribution in a few tens of minutes, whereas a standard technique would have required several days only for a single-point estimate. Figure 2 shows in the left and right panels the mean stratification and standard deviation. The lower half of each panel shows the temperature at the photosphere, and the upper half provides a view of the chromosphere. The uncertainties tend to increase from the photosphere to the chromosphere. The magnitude and uncertainty are correlated since our spectral lines are less sensitive to higher chromospheric temperatures.

6 Conclusions

In summary, normalizing flows can accurately infer the posterior distribution of a solar model atmosphere (parameters, correlations, and uncertainties) from the interpretation of observed photospheric and chromospheric lines. Once the normalizing flow model is trained, the inference is extremely fast. We have also shown that the quality of the approximate posterior distribution depends on the size of the training set and that applying dimensionality reduction techniques makes the normalizing flow performs better. As a natural extension of this work, we plan to include the four Stokes parameters to infer the magnetic properties of our target of interest. Rapid parameter estimation is critical if complex forward models are used to analyze a large amount of data that the existing and next generation of telescopes such as the Daniel K. Inouye Solar Telescope [DKIST; 27] and the European Solar Telescope [EST; 18] will produce.

References

[1] L. H. Auer, J. N. Heasley, and L. L. House. The determination of vector magnetic fields from Stokes profiles. *Sol. Phys.*, 55(1):47–61, Nov. 1977. doi: 10.1007/BF00150873.

[2] W. M. Bolstad. *Understanding Computational Bayesian Statistics*. Wiley Series in Computational Statistics. 1 2012. ISBN 9780470046098. doi: 10.1002/9780470567371.

[3] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *arXiv e-prints*, art. arXiv:1911.01429, Nov. 2019.

[4] J. de la Cruz Rodríguez, J. Leenaarts, and A. Asensio Ramos. Non-LTE Inversions of the Mg II h & k and UV Triplet Lines. *ApJ*, 830(2):L30, Oct. 2016. doi: 10.3847/2041-8205/830/2/L30.

[5] J. de la Cruz Rodríguez, J. Leenaarts, S. Danilovic, and H. Uitenbroek. STiC: A multiatom non-LTE PRD inversion code for full-Stokes solar observations. *A&A*, 623:A74, Mar. 2019. doi: 10.1051/0004-6361/201834464.

[6] C. J. Díaz Baso, J. de la Cruz Rodríguez, and J. Leenaarts. An observationally constrained model of strong magnetic reconnection in the solar chromosphere. Atmospheric stratification and estimates of heating rates. *A&A*, 647:A188, Mar. 2021. doi: 10.1051/0004-6361/202040111.

[7] L. Dinh, D. Krueger, and Y. Bengio. NICE: Non-linear Independent Components Estimation. *arXiv e-prints*, art. arXiv:1410.8516, Oct. 2014.

[8] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Neural Spline Flows. *arXiv e-prints*, art. arXiv:1906.04032, June 2019.

[9] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. nflows: normalizing flows in pytorch, Nov. 2020. URL `https://doi.org/10.5281/zenodo.4296287`.

[10] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP*, 125 (925):306, Mar. 2013. doi: 10.1086/670067.

[11] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. *arXiv e-prints*, art. arXiv:1512.03385, Dec. 2015.

[12] G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, July 2006. doi: 10.1126/science.1127647.

[13] D. Jimenez Rezende and S. Mohamed. Variational Inference with Normalizing Flows. *arXiv e-prints*, art. arXiv:1505.05770, May 2015.

[14] S. Kianfar, J. Leenaarts, S. Danilovic, J. de la Cruz Rodríguez, and C. José Díaz Baso. Physical properties of bright Ca II K fibrils in the solar chromosphere. *A&A*, 637:A1, May 2020. doi: 10.1051/0004-6361/202037572.

[15] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, art. arXiv:1412.6980, Dec. 2014.

[16] D. P. Kingma and P. Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions. *arXiv e-prints*, art. arXiv:1807.03039, July 2018.

[17] J. Leenaarts, J. de la Cruz Rodríguez, S. Danilovic, G. Scharmer, and M. Carlsson. Chromospheric heating during flux emergence in the solar atmosphere. *A&A*, 612:A28, Apr. 2018. doi: 10.1051/0004-6361/201732027.

[18] S. A. Matthews, M. Collados, M. Mathioudakis, and R. Erdelyi. *The European Solar Telescope (EST)*, volume 9908 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 990809. 2016. doi: 10.1117/12.2234145.

[19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21:1087–1092, June 1953. doi: 10.1063/1.1699114.

[20] T. Müller, B. McWilliams, F. Rousselle, M. Gross, and J. Novák. Neural Importance Sampling. *arXiv e-prints*, art. arXiv:1808.03856, Aug. 2018.

[21] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv e-prints*, art. arXiv:1912.01703, Dec. 2019.

[22] O. Rippel and R. Prescott Adams. High-Dimensional Probability Estimation with Deep Density Models. *arXiv e-prints*, art. arXiv:1302.5125, Feb. 2013.

[23] W. Rudin. *Real and complex analysis*. Tata McGraw-Hill, New Delhi, third edition, 2006.

[24] G. B. Scharmer, K. Bjelksjo, T. K. Korhonen, B. Lindberg, and B. Petterson. The 1-meter Swedish solar telescope. In S. L. Keil and S. V. Avakyan, editors, *Innovative Telescopes and Instrumentation for Solar Astrophysics*, volume 4853 of *Proc. SPIE*, pages 341–350, Feb. 2003. doi: 10.1117/12.460377.

[25] G. B. Scharmer, G. Narayan, T. Hillberg, J. de la Cruz Rodriguez, M. G. Löfdahl, D. Kiselman, P. Sütterlin, M. van Noort, and A. Lagg. CRISP Spectropolarimetric Imaging of Penumbral Fine Structure. *ApJ*, 689: L69, Dec. 2008. doi: 10.1086/595744.

[26] E. G. Tabak and C. V. Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013.

[27] A. Tritschler, T. R. Rimmele, S. Berukoff, R. Casini, S. C. Craig, D. F. Elmore, R. P. Hubbard, J. R. Kuhn, H. Lin, J. P. McMullin, K. P. Reardon, W. Schmidt, M. Warner, and F. Woger. DKIST: Observing the Sun at High Resolution. In *18th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, volume 18 of *Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, pages 933–944, Jan. 2015.

[28] G. J. M. Vissers, J. de la Cruz Rodríguez, T. Libbrecht, L. H. M. Rouppe van der Voort, G. B. Scharmer, and M. Carlsson. Dissecting bombs and bursts: non-LTE inversions of low-atmosphere reconnection in SST and IRIS observations. *A&A*, 627:A101, July 2019. doi: 10.1051/0004-6361/201833560.

[29] R. Yadav, C. J. Díaz Baso, J. de la Cruz Rodríguez, F. Calvo, and R. Morosin. Stratification of physical parameters in a C-class solar flare using multiline observations. *A&A*, 649:A106, May 2021. doi: 10.1051/0004-6361/202039857.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Section 3 and Section 4.
   (b) Did you describe the limitations of your work? [Yes] See Section 5.
   (c) Did you discuss any potential negative social impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] There are no potential negative societal impacts.
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A] Our results are not purely theoretical.
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] It is included in our repository: `https://github.com/cdiazbas/bayesflows`
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 2

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A] The difference between multiple training normalizing flows is well below the width of the distributions.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A] There is not a big difference between a CPU and a standard GPU. The training takes around 12-24 hours.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

   (a) If your work uses existing assets, did you cite the creators? [Yes]

   (b) Did you mention the license of the assets? [Yes] It is indicated in each source repository.

   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Links of the codes we have used to generate the database are included.

   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] Any database can be generated randomly from physical models.

   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] There is no personal information in the dataset.

5. If you used crowdsourcing or conducted research with human subjects...

   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]