UNIVERSITÀ DEGLI STUDI
DI GENOVA

INFN
GENOVA
Istituto Nazionale di Fisica Nucleare
Sezione di Genova

The NFlikelihood:
Unsupervised Machine Learning LHC likelihoods with Normalizing Flows.

Humberto Reyes-González
University of Genoa

In collab. with Riccardo Torre

IRN Terascale, Grenoble.
26/04/2023

Introduction

- **Likelihood functions (full statistical models)** parametrise the full information of an LHC analysis; wether it is New Physics (NP) search or an SM measurement.
- Their **preservation** is a key part of the **LHC legacy.**

Usage:
- Resampling
- Reinterpretation with different statistical approaches.
- Reinterpretation in the context of different NP models.
- …

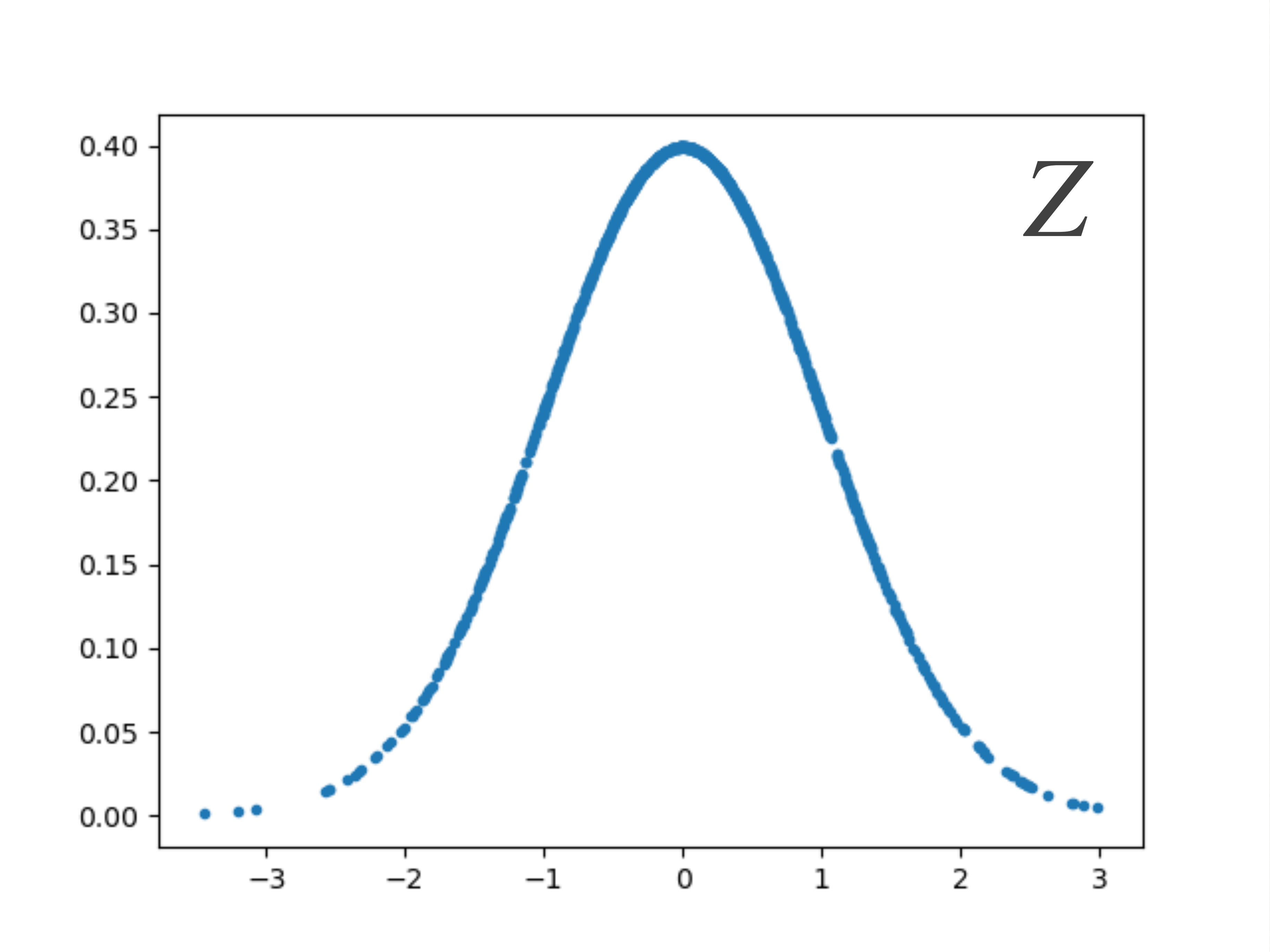
Challenges:
- LHC likelihoods are often high-dimensional complex distributions.
- We want precise descriptions that can be efficiently reinterpreted.

Important steps forward:
- ATLAS started publishing full likelihoods of NP searches ATL-PHYS-PUB-2019-029.
- Release of the pyhf package to construct statistical models 10.21105/joss.02823, L Heinrich, M Feickert, G Stark
- Theorists have started profiting from this arXiv:2012.08192, arXiv:2206.14870 SModelS/MadAnalys collab.
- Supervised learning with DNN likelihood arxiv:1911.03305 A Coccaro, M. Perini, L Silvestrini, R Torre

Our approach: Unsupervised Learning with Normalizing Flows

Intro to NFs.

BASIC PRINCIPLE:
Following the change of variables formula, perform a series of **bijective, continuous, invertible** transformations on a *simple* probability density function (pdf) to obtain a *complex* one.

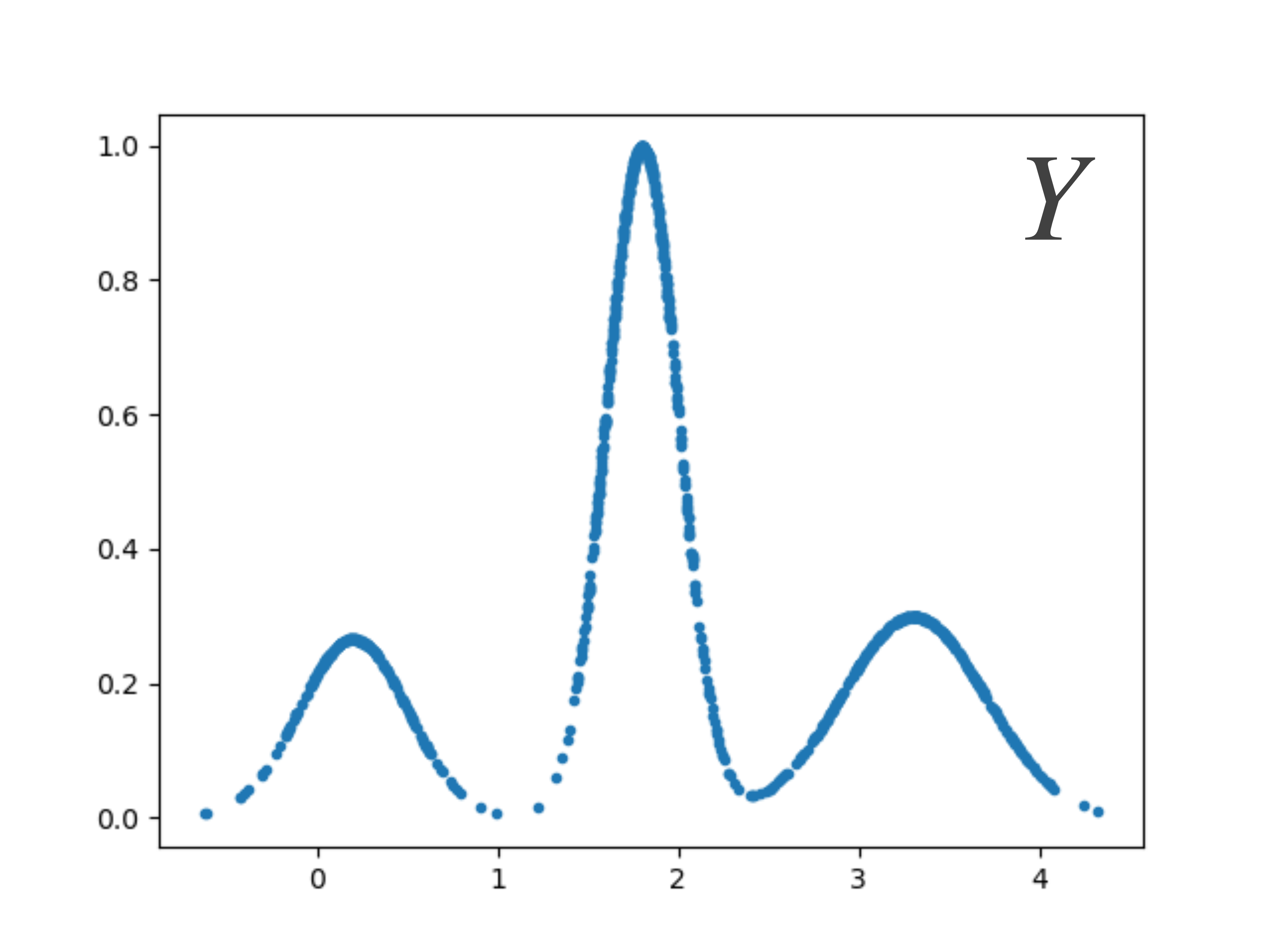


Normalizing direction

$$Z = f(Y)$$

Generative direction

$$Y = g(Z)$$



See review
Ivan K. et. al.
arXiv:1908.09257

Choosing the transformations

THE OBJECTIVE:

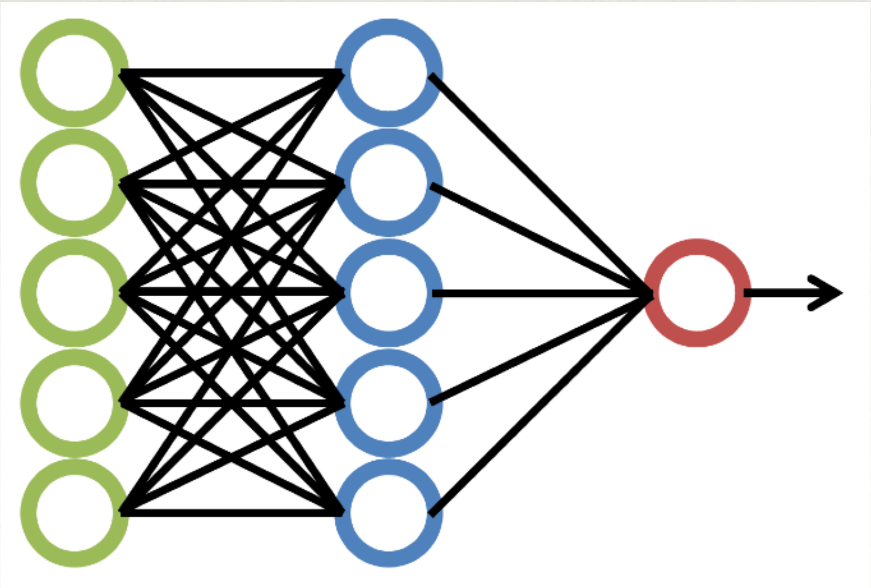
To perform the right transformations to accurately estimate the complex underlying distribution of some observed data.

THE RULES OF THE GAME:

- The transformations must be invertible
- They should be sufficiently expressive
- And computationally efficient (including Jacobian)

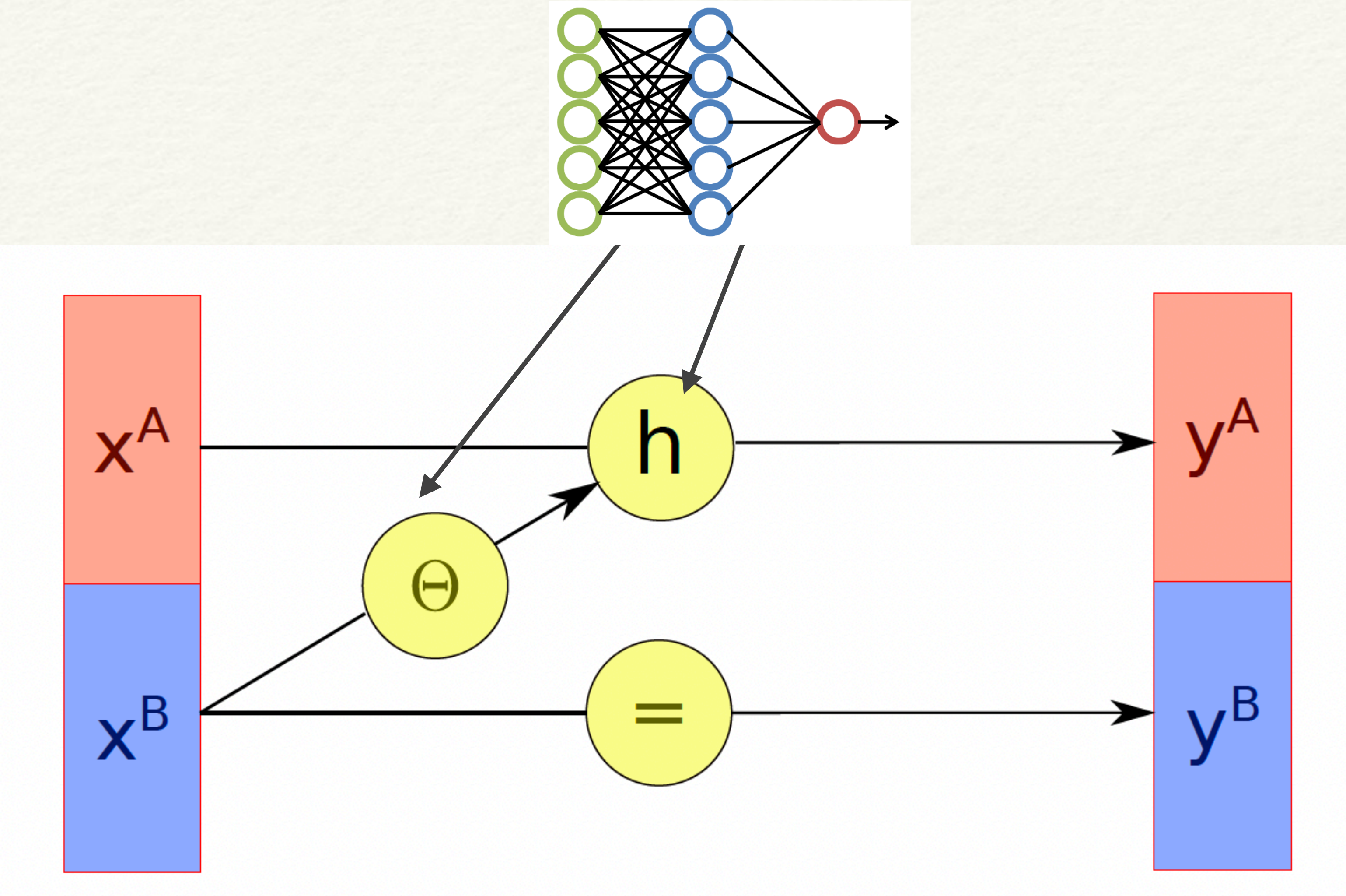
THE STRATEGY

Let ***Neural Networks*** learn the parameters of ***Autoregressive Normalizing Flows.***

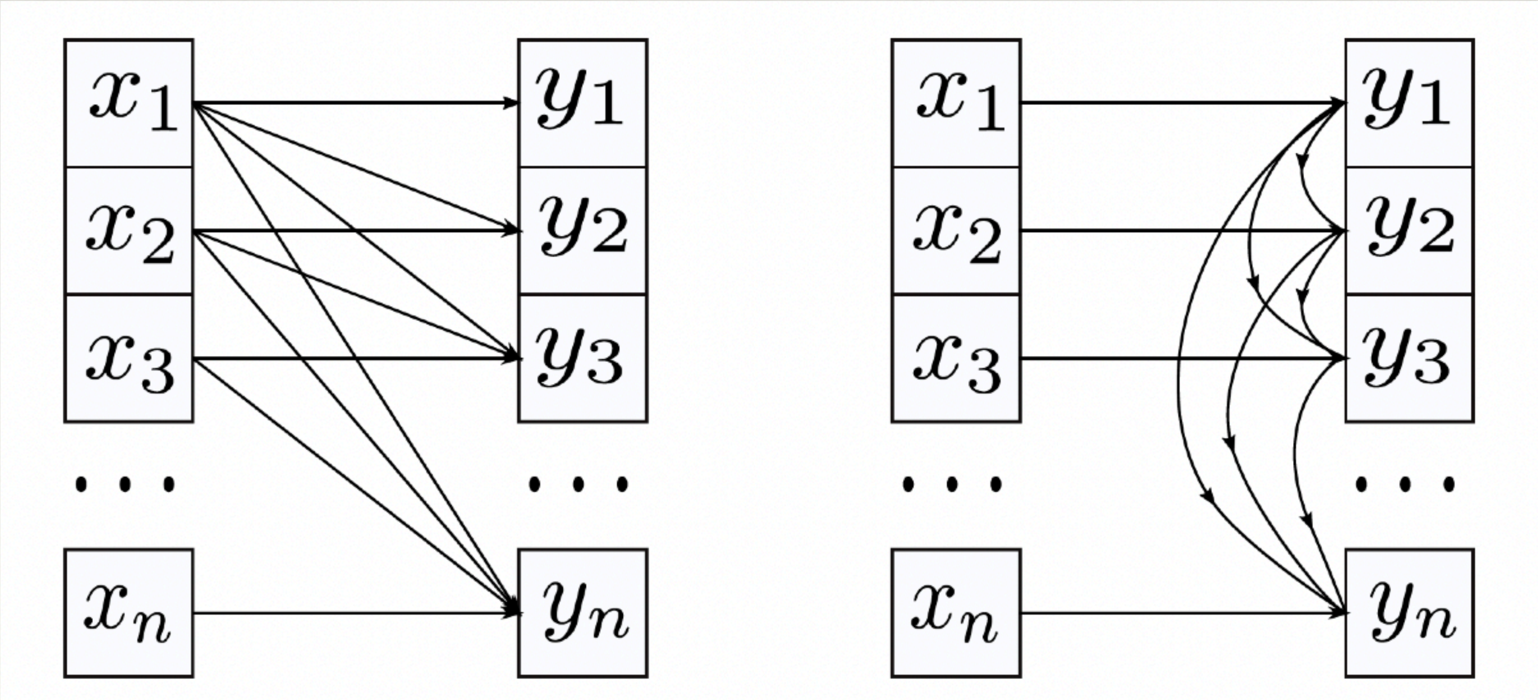


Autoregressive Flows

- Dimension x^i is transformed with bijectors trained with $y_{1:i-1}$
- Bijector parameters are trained with Autoregressive NNs.
- The Jacobian J is triangular.
- **Jacobian is easily computed!**
- **Direct sampling OR density estimation.**
- **Expressive transformations.**



arXiv:1908.09257



arXiv:1908.09257

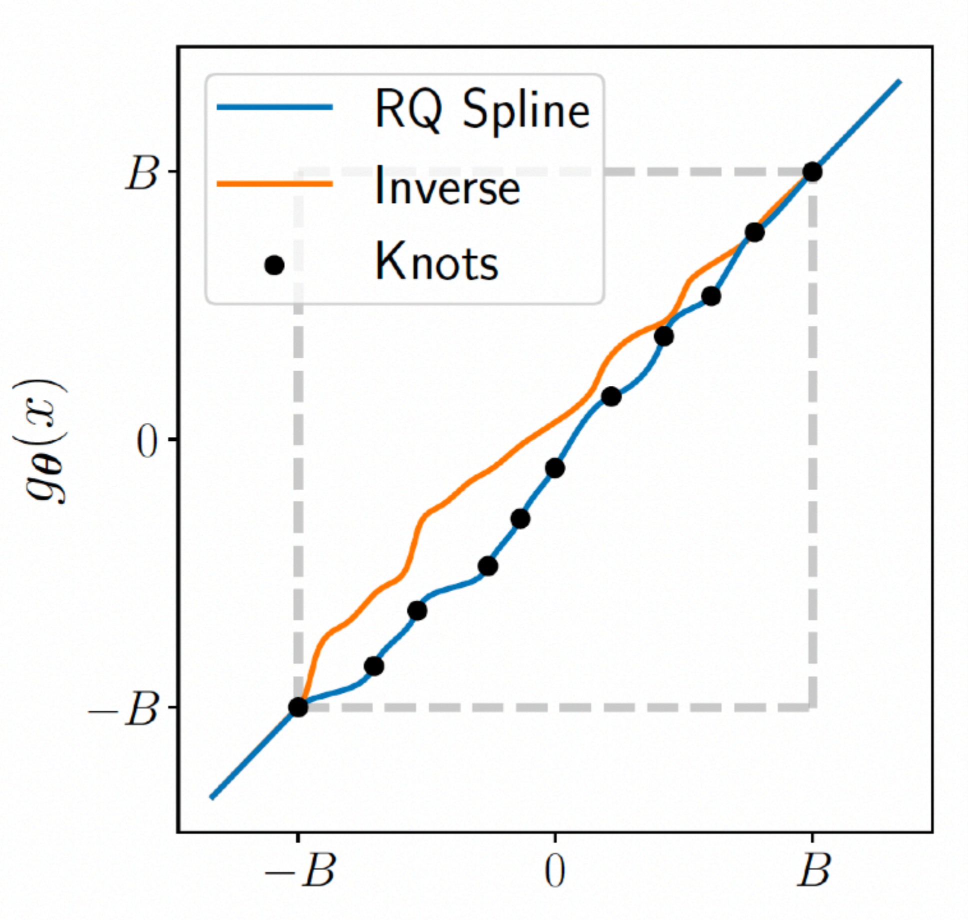
The loss function:
$-\log(p_{AF}(real_{dist}))$

Autoregressive Flows

MAF	A-RQS
Masked Autoregressive Flow	Neural Spline Flows
arXiv:1705.07057	arXiv:1906.04032

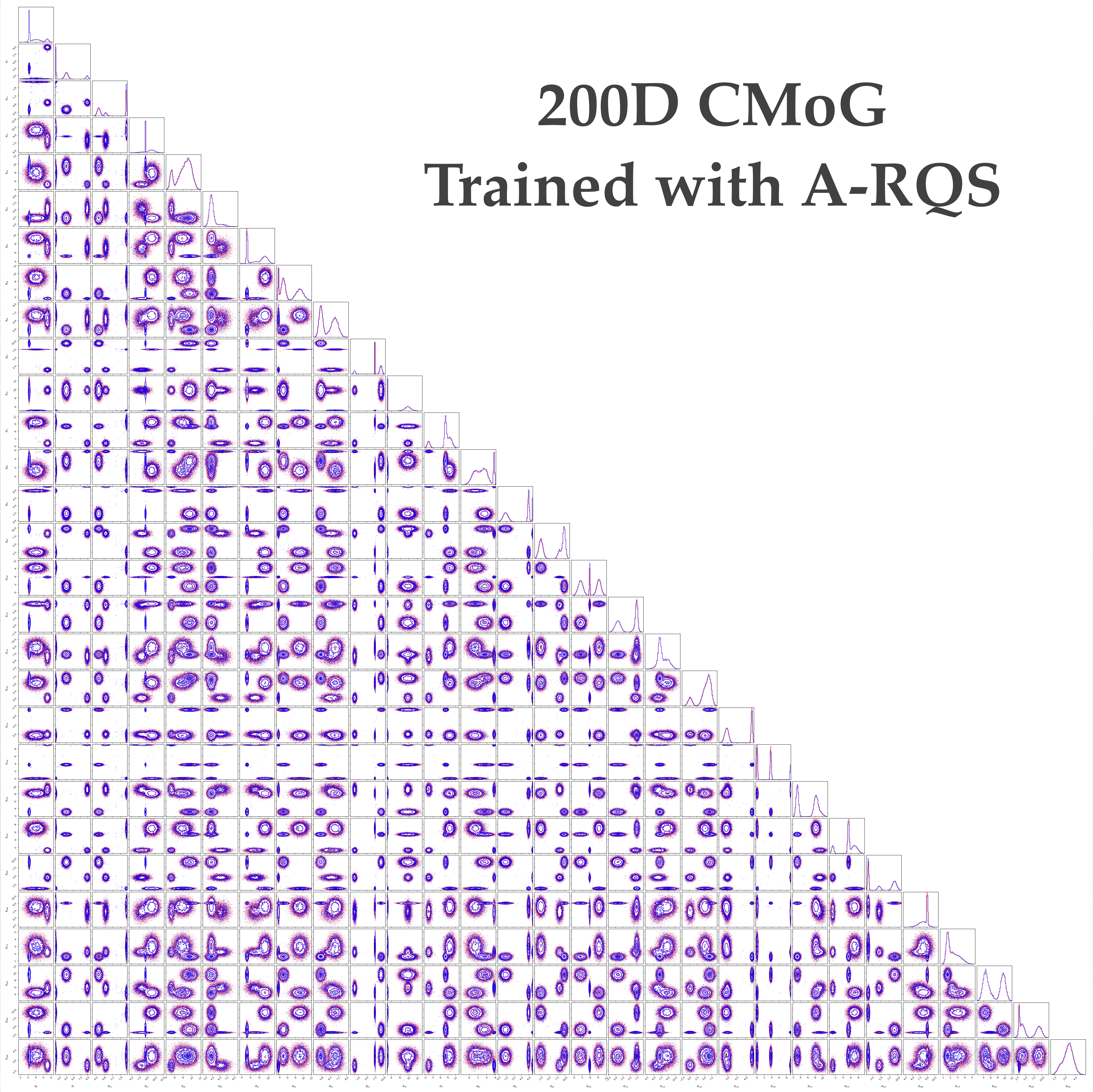
Affine ← **BIJECTORS** → Rational Quadratic Spline

$$y(x; \mu, b) = \mu \cdot x + b$$



arXiv:1906.04032

Normalizing Flows in High Dimensions



NFs scale very well to High Dimensions:

arXiv:2202.09188,arXiv:2302.12024 (Coccaro, Letizia, Torre, HRG)

Great potential in HEP, Astro, Lattice ….

- Numerical integration (arXiv:2001.05486, arXiv:2001.05478)
- Unfolding (arXiv:2006.06685)
- Calorimeter shower simulation (arXiv:2106.05285,arXiv:2211.15380)
- Event generation (arXiv:2001.10028, arXiv:2110.13632)
- Anomaly detection (arXiv:2001.04990)
- Astro/Cosmo (arXiv:2104.12789,arXiv:2105.0332)
- Lattice (arXiv:2101.08176, arXiv:2210.03139)

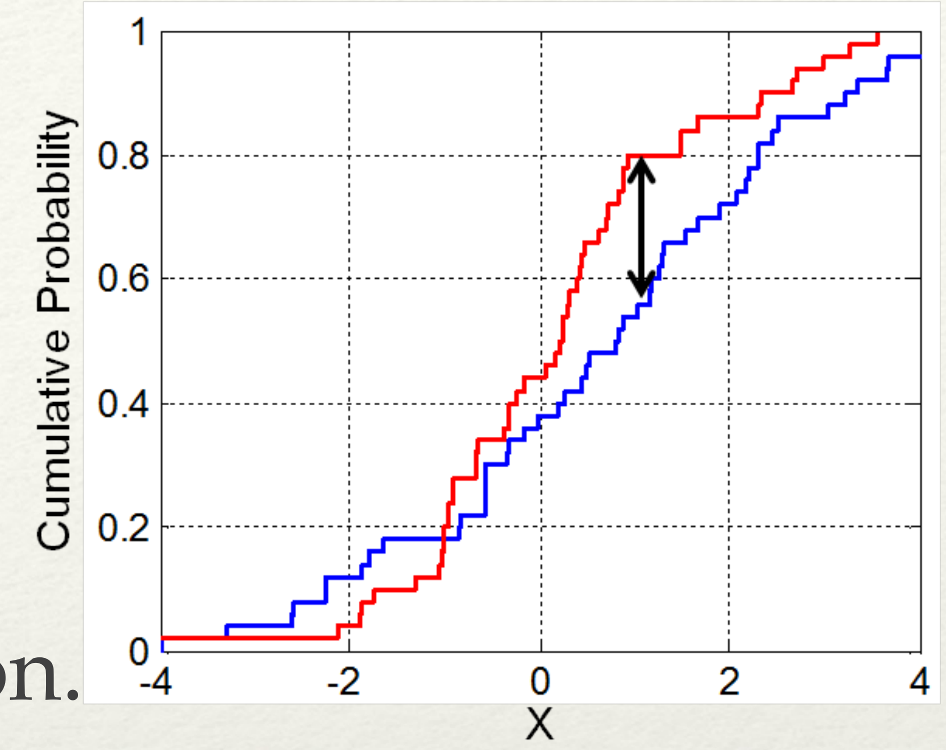
Non-exhaustive list…

better type 'Normalizing Flows' on InspireHEP and see for yourself!

Evaluation metrics.

- Two-sample 1D Kolgomonov - Smirnov test (ks test):

$$D_{n,m} = \sup_x |F_n(x) - F_m(x)|$$



https://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test

-Computes the p-value for two sets of 1D samples coming from the same *unknown* distribution.
-We average over ks test estimations and compute the median over dimensions.
-**Optimal value 0.5**

-HDPIe:

Highest Posterior Density Interval (HDPI) on a posterior density for fixed probability. We computed the **HPDI relative error width (HPDIe)** at the σ, 2σ and 3σ intervals between each 1D marginal of the true and predicted distributions. At each dimension, we computed the median of relative errors between each corresponding interval. The result is summarized as the median over all dimensions.

The NFLikelihood.

(HRG, R. Torre, arXiv:2305.xxxxx))

LHC likelihoods in a nutshell

Bayes theorem:

$$P(\Theta, x) = P_x(x\,|\,\Theta)\pi_\Theta(\Theta) = P_\Theta(\Theta\,|\,x)\pi_x(x)$$

LHC Statistical model:

$$P(\mu, \theta; \text{data}) = \Pi_{k=1}^{n_c} P[n_i; \mu\epsilon_{i,k}(\vec{\theta})N_{S,i,k}(\vec{\theta}) + B_{i.k}(\vec{\theta})]\Pi_{j=1}^{n_{syst}} G(\theta_j^{obs}; \theta_j; 1)$$

(Observed) data

Nuisance parameters (uncertainties)

(Auxiliary) data

Parameters of Interest (signal strength, etc.)

Test Statistic:

$$t(\mu) = -2\log\frac{L(\mu; \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta}(\hat{\mu}))}$$

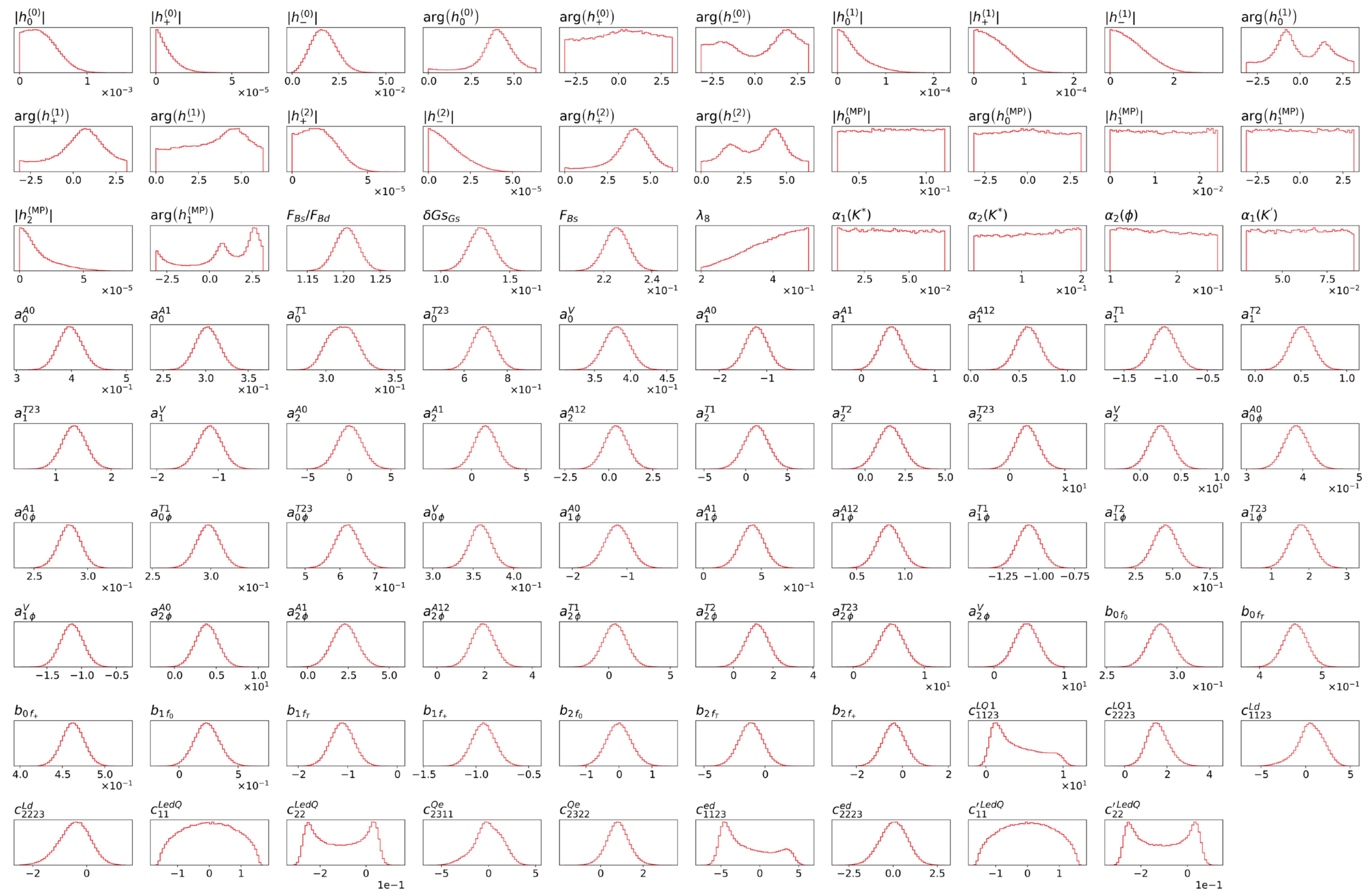
Best- fits:

$$L(\hat{\mu}; \hat{\theta})$$

Where μ are observables

best-fit $\theta(\mu)$

Even $P(\Theta \mid x = \text{obs})$ can be challenging…



CHALLENGES:

- High-dimensionality
- Truncated distributions
- Multimodal dimensions
- "Noisy" dimensions
- Correlations
- Wide/different ranges

Example Likelihoods

$P(\Theta | x = \text{obs})$

LHC-like new physics search Likelihood.

- 1 parameter of interest (signal strength)
- 94 nuisance parameters.
- Ref. arXiv:1809.05548

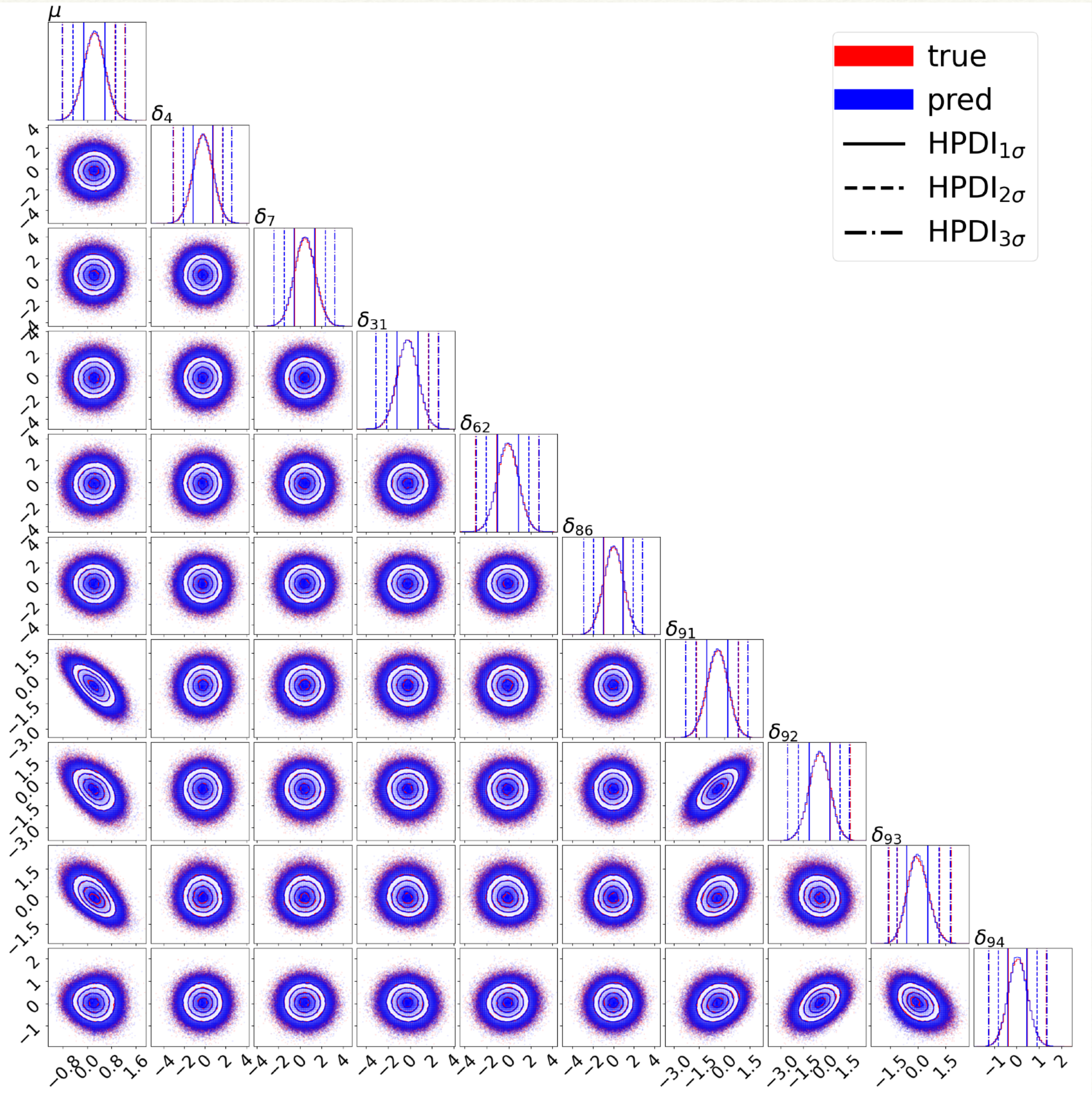
ElectroWeak fit Likelihood

- EW observables.
- Including recent measurements of top mass (CMS) and W mass (CDF).
- 8 parameters of interest (Wilson coefficients of SMEFT operators)
- 32 nuisance parameters.
- Ref. arXiv:2204.04204

Flavor fit likelihood.

- Flavor observables related to $b \to sl^+l^-$ transitions
- 12 parameters of interest (Wilson coefficients of SMEFT operators)
- 77 nuisance parameters.
- Ref. arXiv:1903.09632

LHC-like new physics search Likelihood.



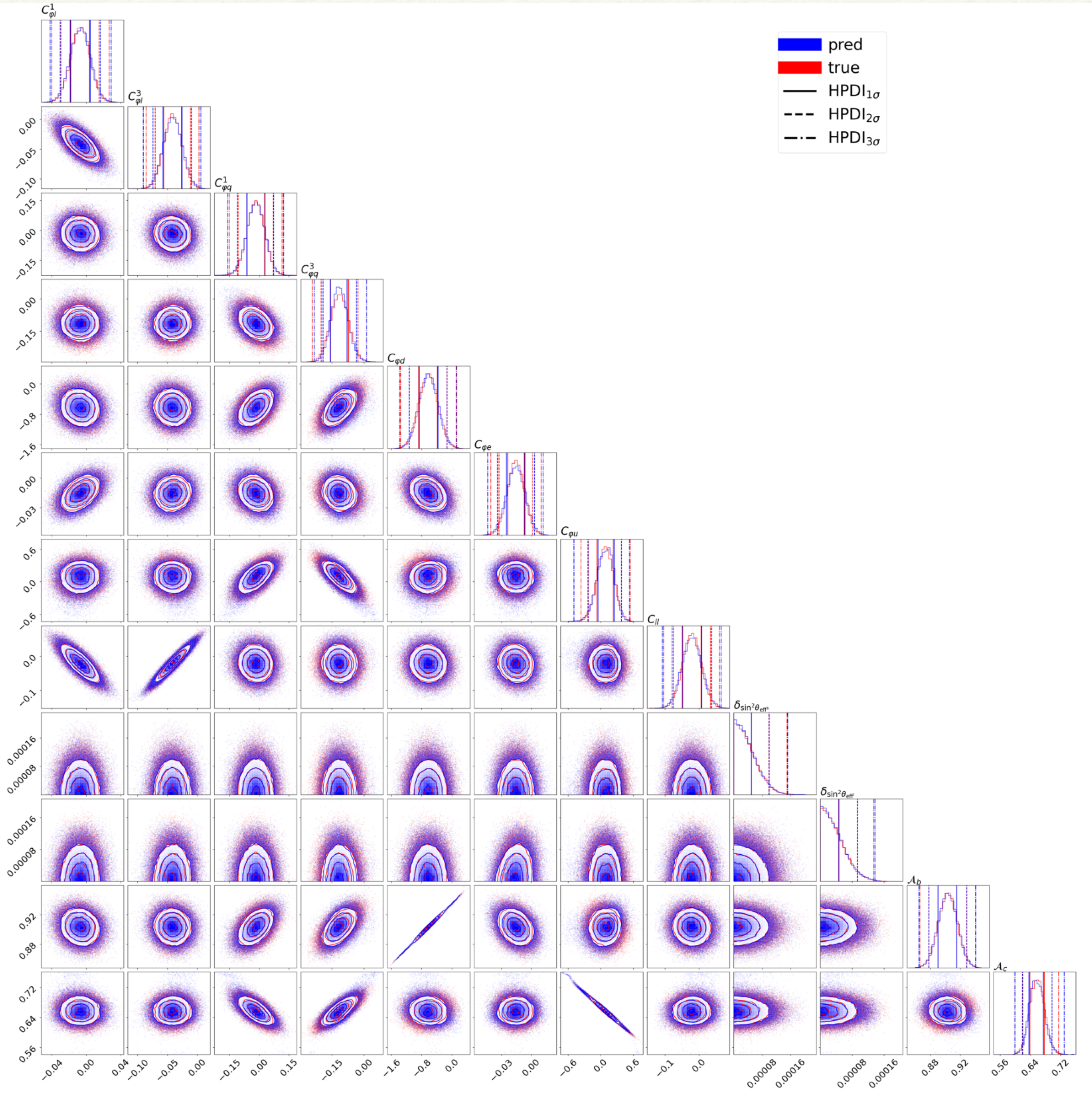
Test sample : 300k

Hyperparameters for Toy-Likelihood								
# of samples	hidden layers	# of bijec.	algorithm	spline knots	range	L1 factor	patience	max # of epochs
$\mathbf{2 \cdot 10^5}$	3×64	MAF	2	-	-	0	20	200

Results for Toy-Likelihood							
# of samples	Median KS-test	Median W-distance	Frobenius norm	HPDIe$_{1\sigma}$	HPDIe$_{2\sigma}$	HPDIe$_{3\sigma}$	time (s)
$\mathbf{2 \cdot 10^5}$	0.496	$5.39 \cdot 10^{-3}$	0.217	$2.179 \cdot 10^{-2}$	$9.738 \cdot 10^{-3}$	$1.059 \cdot 10^{-2}$	300

Results for Toy-Likelihood POI				
POI	KS-test	HPDIe$_{1\sigma}$	HPDIe$_{2\sigma}$	HPDIe$_{3\sigma}$
μ	0.5127	0.02184	0.009431	0.01584

ElectroWeak fit Likelihood



Test sample : 100k

Hyperparameters for the EW-Likelihood

# of samples	hidden layers	# of bijec.	algorithm	spline knots	range	L1 factor	patience	# of epochs
$\mathbf{2 \cdot 10^5}$	2	3×128	A-RQS	4	-6	0	20	800

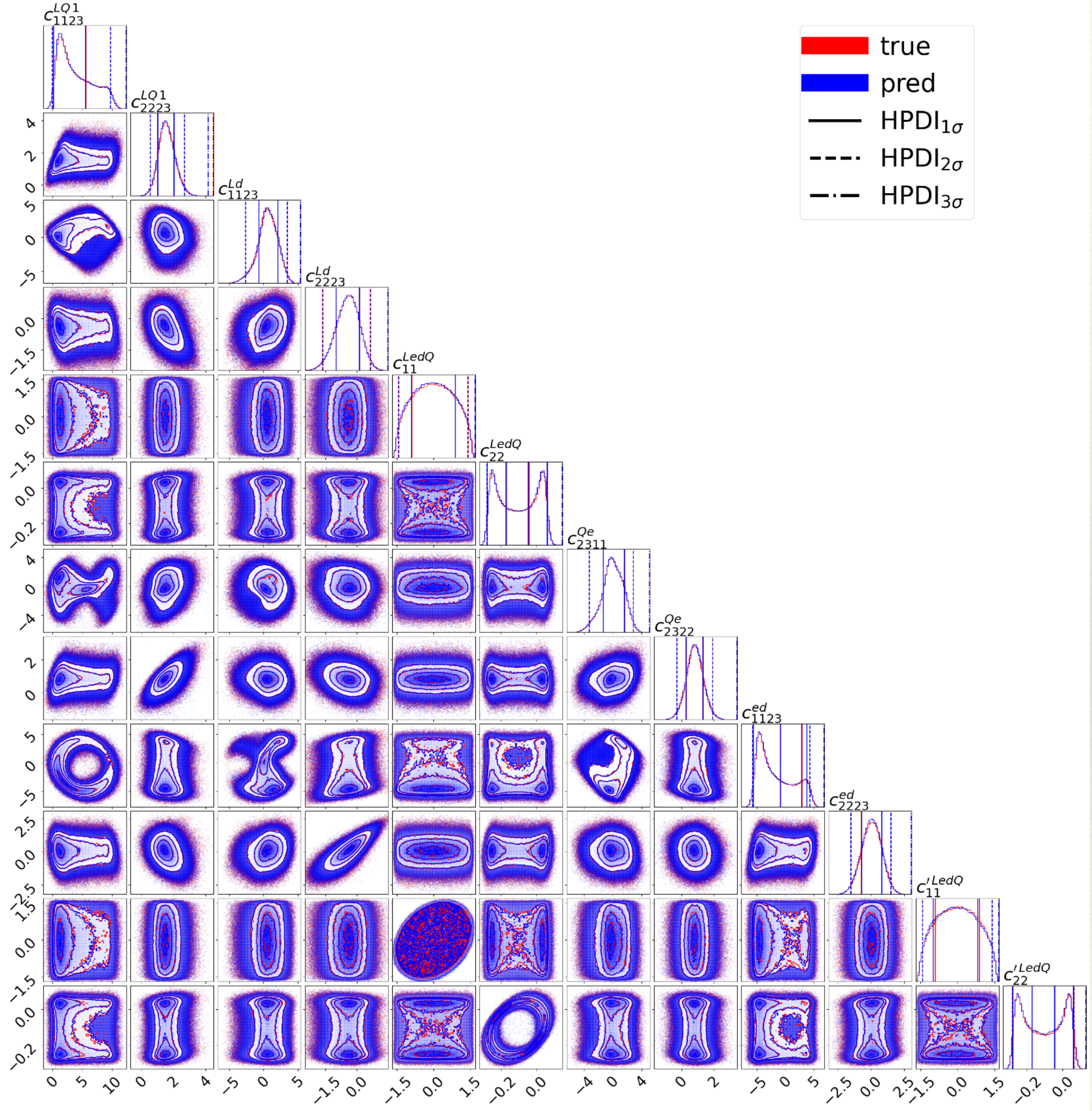
Results for the EW-Likelihood

# of samples	Median KS-test	Median W-distance	Frobenius norm	HPDIe$_{1\sigma}$	HPDIe$_{2\sigma}$	HPDIe$_{3\sigma}$	time (s)
$\mathbf{2 \cdot 10^5}$	.489	$5.59 \cdot 10^{-5}$	1.154	$8.57 \cdot 10^{-4}$	$8.88 \cdot 10^{-4}$	$2.93 \cdot 10^{-3}$	7255

Results for EW-Likelihood

POI	KS-test	HPDIe$_{1\sigma}$	HPDIe$_{2\sigma}$	HPDIe$_{3\sigma}$
$c^1_{\varphi l}$	0.4662	0.06708	0.06345	0.07149
$c^3_{\varphi l}$	0.3741	0.02977	0.0945	0.7034
$c^1_{\varphi q}$	0.4251	0.04999	0.03841	0.04914
$c^3_{\varphi q}$	0.2404	0.05101	0.1361	0.01995
$c_{\varphi d}$	0.495	0.02929	0.01275	0.05298
$c_{\varphi e}$	0.2719	0.06732	0.7391	0.1454
$c_{\varphi u}$	0.2643	0.1912	0.01737	0.1953
c_{ll}	0.4288	0.1558	0.0752	0.06005

Flavor fit Likelihood



Hyperparameters for the Flavor-Likelihood

# of samples	hidden layers	# of bijec.	algorithm	spline knots	range	L1 factor	patience	max # of epochs
10^6	3×1024	2	A-RQS	8	-5	$1e^{-}4$	50	12000

Results for the Flavor-Likelihood

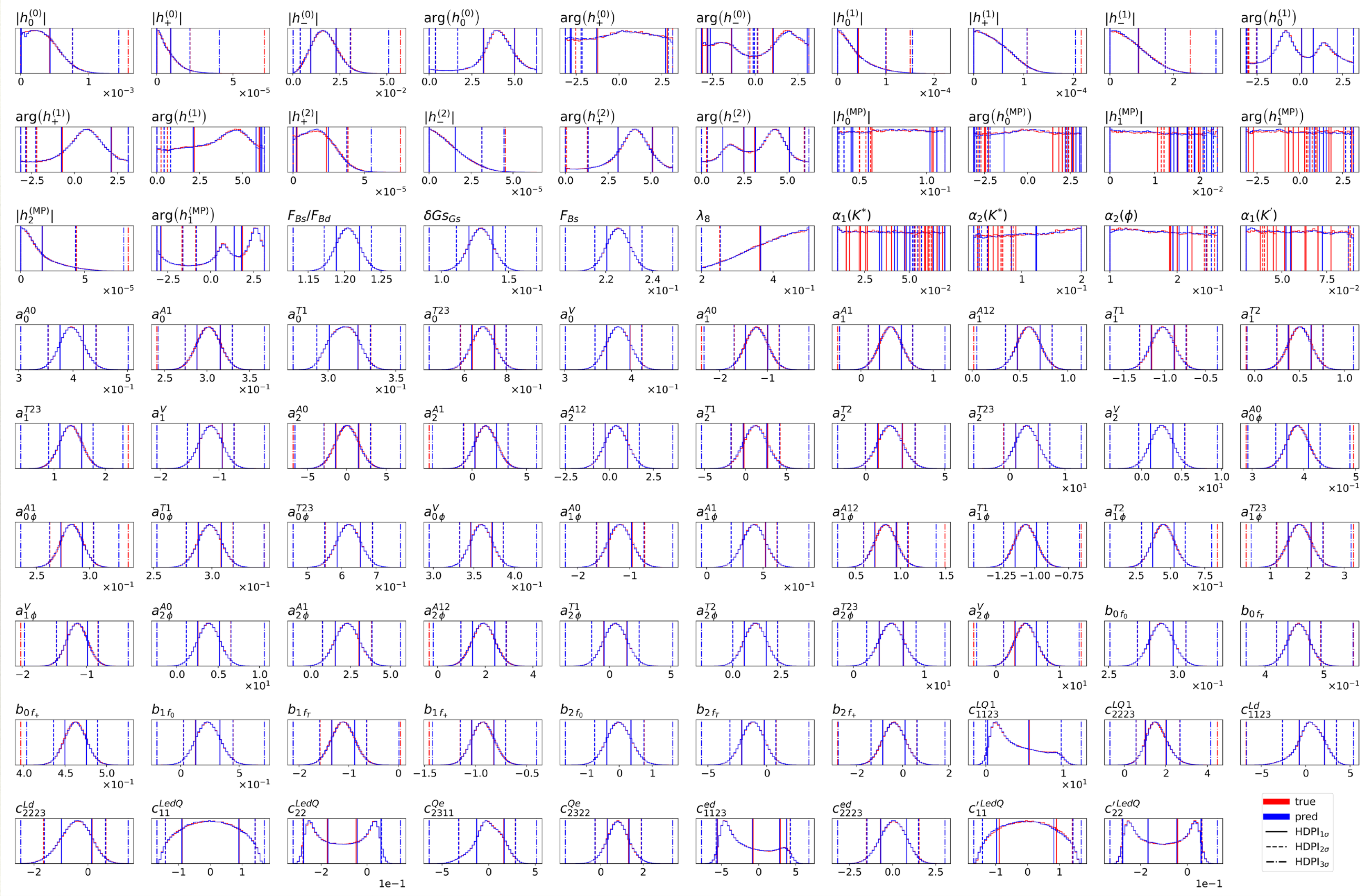
# of samples	Median KS-test	Median W-distance	Frobenius norm	$\mathrm{HPDIe}_{1\sigma}$	$\mathrm{HPDIe}_{2\sigma}$	$\mathrm{HPDIe}_{3\sigma}$	time (s)
10^6	0.455	$2.71 \cdot 10^{-3}$	0.2801	$8.67 \cdot 10^{-3}$	$7.346 \cdot 10^{-3}$	$1.42 \cdot 10^{-7}$	20212

Results for Flavor-Likelihood POIs

POI	KS-test	$\mathrm{HPDIe}_{1\sigma}$	$\mathrm{HPDIe}_{2\sigma}$	$\mathrm{HPDIe}_{3\sigma}$
$c^{LQ\,1}_{1123}$	0.4346	0.007251	1.83e-05	4.731e-08
$c^{LQ\,1}_{2223}$	0.4736	0.01249	0.00162	0.03575
c^{Ld}_{1123}	0.486	0.01466	0.006628	0.002338
c^{Ld}_{2223}	0.4138	0.0513	0.02446	2.398e-08
c^{LedQ}_{11}	0.5362	0.00738	0.004683	5.387e-08
c^{LedQ}_{22}	0.5161	0.02799	0.001639	2.155e-09
c^{Qe}_{2311}	0.4476	0.01389	0.007458	1.419e-07
c^{Qe}_{2322}	0.382	0.02132	0.02496	0.0004609
c^{ed}_{1123}	0.4789	0.04076	0.00333	5.602e-08
c^{ed}_{2223}	0.4436	0.008685	0.016	1.502e-08
c'^{LedQ}_{11}	0.3203	0.09194	0.007041	8.011e-08
c'^{LedQ}_{22}	0.4157	0.03001	0.008749	4.374e-08

Test sample : 500k

Flavor fit Likelihood



Conclusions

- Normalizing Flows show great capacity of learning complex high dimensional functions. Specially, the A-RQS.
- NFs can accurately and efficiently model LHC likelihoods.

Outlook

- Paper in preparation (2305.xxxxx).
- Check out N4HEP project: https://github.com/NF4HEP
- Deeper look into evaluation metrics, e.g. classifier likelihood ratio-test.
- Study NFs for data augmentation, uncertainties of NFs, framework agnostic NF models.
- Extended and systematic usage of NFlikelihoods.
- Learning full statistical models with Conditional NFs.

THANK YOU!