THE USE OF GANS IN THE SEARCH FOR NEW RESONANCES AT THE LHC FOR SEMI-SUPERVISED MACHINE LEARNING TECHNIQUES

PRESENTED BY
BENJAMIN LIEBERMAN

TOPIC HIGHLIGHTS

PRESENTATION OUTLINE

1. Introduction to semi-supervised classification and the quantifying of uncertainty generated in training models
    - Semi-supervised Machine Learning Classification BSM
    - Quantifying over-training in physics classification
        - Zy Study and Dataset
        - Semi-Supervised DNN and Response
        - Fitting to DNN output invariant mass
        - Significance of fake signals generated
2. Look Elsewhere Effect and need for AI driven data generation
3. Generative Adversarial Network
    - Scaling datasets using WGANs
    - Training and Results of WGANs
4. Conclusions

MACHINE LEARNING SEMI-SUPERVISED CLASSIFICATION

FULL SUPERVISION

TRAINING DATASETS

Sample 1: Labelled Background dataset

Sample 2: Labelled Signal dataset

CHARACTERISTICS

- Excellent classification of signal from background based on well defined physics in training datasets.
- Results however are biased to characteristics of given training set.

$$f_{full} = argmin_{f:\mathbb{R}^n \to [0,1]} \sum_{i=1}^{N} \ell(y_i, \hat{y}_i),$$

SEMI SUPERVISION

TRAINING DATASETS

Sample 1: Labelled Background dataset

Sample 2: Unlabelled (Signal + Background) dataset

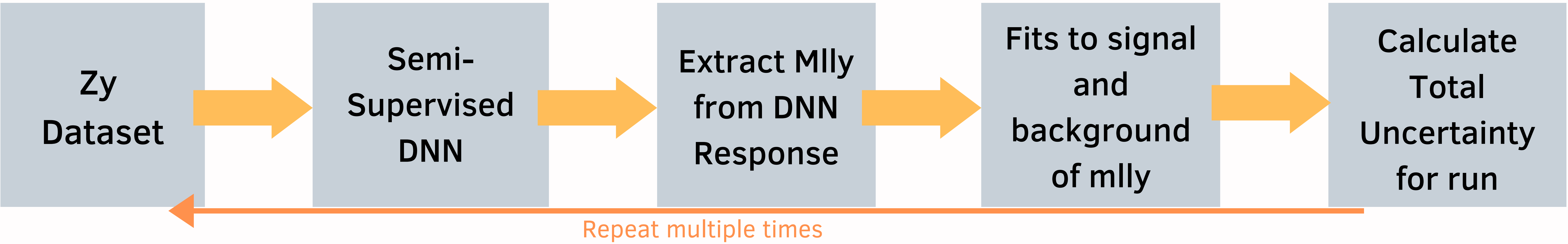
CHARACTERISTICS

- Classification of signal from background with well guided training samples.
- Provide classification of datasets that are not as well defined without limiting results by currently understood physics.

$$f_{semi} = argmin_{f:\mathbb{R}^n \to [0,1]} \sum_{K} \ell \left(\frac{1}{|K|} \sum_{i \in K} \hat{y}_i, y_K \right),$$

QUANTIFYING OVERTRAINING IN SEMI-SUPERVISION

When using machine learning classification models in particle physics, the extent of uncertainty generated due to overtraining must be quantified in order to validate results. To investigate the extent of false signals generated, due to over-training, by semi-supervised classification models, the following methodology is implemented.



ATLAS full simulation to be used in final study. Fast simulation data used for current analysis

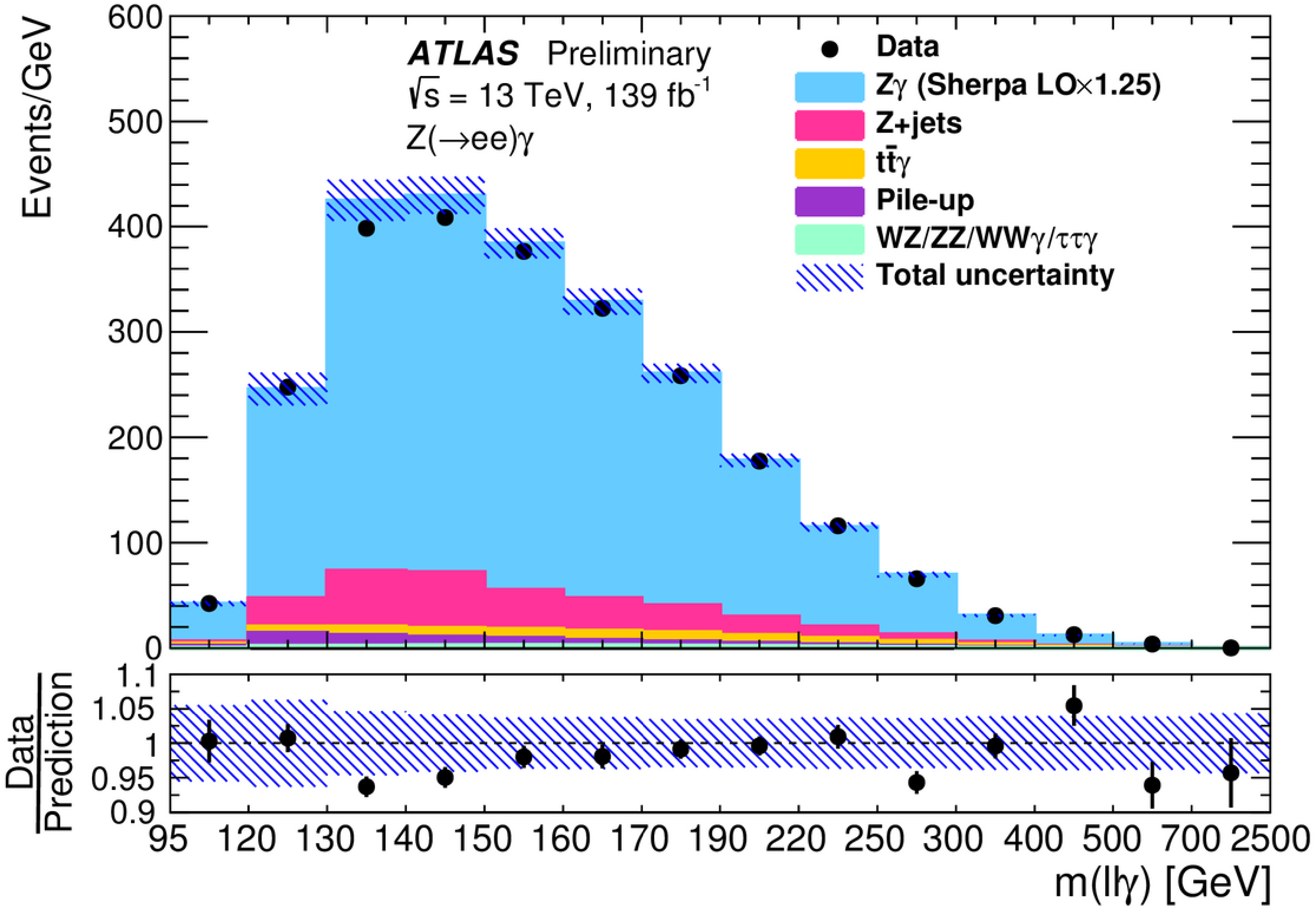
Model trained using a batch of background data (sample 1 and 2 containing Zy dataset).

Extracting batches of different numbers of events from the DNN response distribution.

Functional fits to mass window (signal) and side-band (background) regions of mlly distributions of each batch.

Calculating significance of false signals exposed through mlly fits.

Z-GAMMA FINAL STATE MONTE CARLO DATASET



Measurement of Zγ→ℓ+ℓ−γ differential cross-sections in pp collisions at s√=13 TeV with the ATLAS detector

Zγ→(ℓ+ℓ−)γ

Mass Range based on invariant mass of di-lepton system with gamma, mℓℓγ:

130-170GeV

Zy final state cuts:

- Number of leptons >= 2
- Dilepton (muons or electrons) have opposite charge (ℓ+ℓ−)
- Number of photons (gamma) >= 1

Features used to Train DNN:

$$m_{\ell\ell\gamma},\ \Phi_{\ell\ell\gamma},\ \eta_{\ell\ell\gamma},\ Pt_{\ell\ell\gamma},$$

$$E_{\ell\ell\gamma},\ m_{\ell\ell},\ \Phi_{\ell\ell},\ \eta_{\ell\ell},$$

$$Pt_{\ell\ell},\ E_{\ell\ell},\ \Delta R_{\ell\ell},$$

$$E_T^{miss},\ \Phi_{E_T^{miss}},\ \Delta\Phi_{\ell\ell},$$

$$\Delta\Phi(E_T^{miss}, Z\gamma),\ Pt_{\ell\ell}/m_{\ell\ell\gamma},$$

$$N_j,\ N_{cj}$$

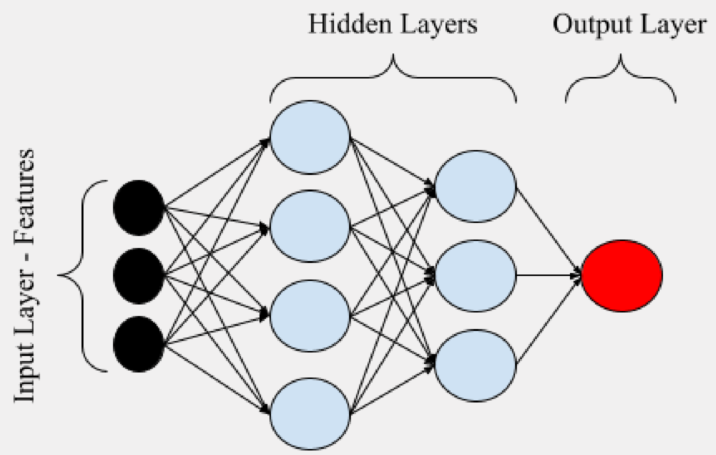
Data Simulation

Monte Carlo Zy Data generated using Madgraph5 with NNPDF3.0 parton distribution function. Parton level generation is done using Pythia and detector level simulation is done using Delphes(v3)

OVERTRAINING ANALYSIS METHODOLOGY

SEMI-SUPERVISED DNN CLASSIFIER

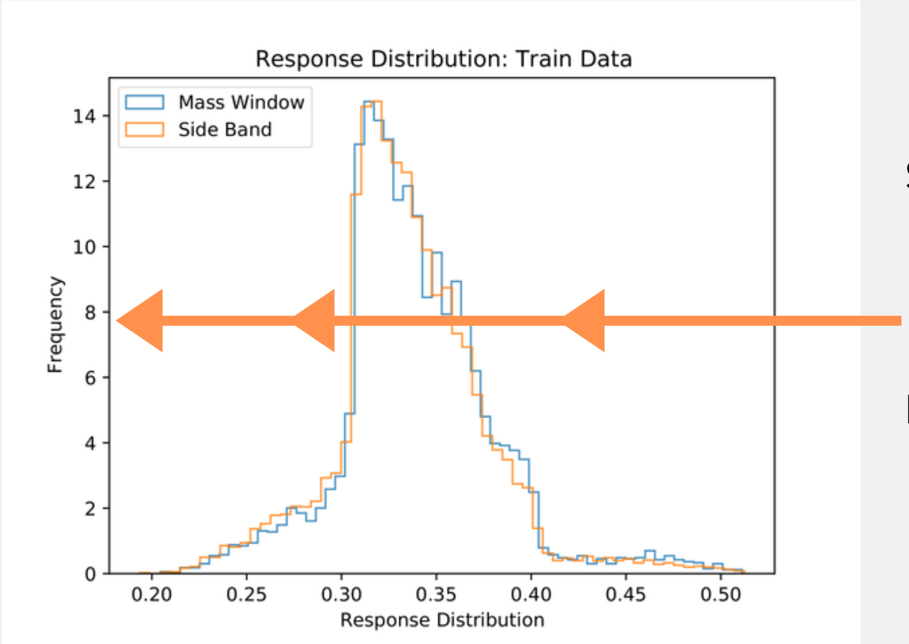
- The Deep Neural Network (DNN) classifier was selected and optimised to classify using both the full supervision and semi-supervision frameworks.
- A learning rate of 1·10−3 is used with a learning decay of 3·10−4. The model is run for 8 epochs using a batch size of 1.



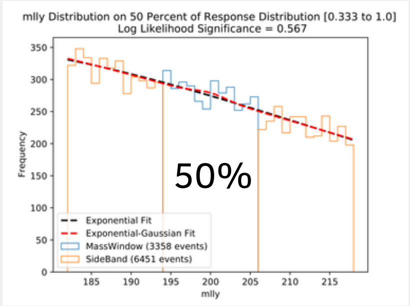
Layer	Number of Nodes	Activation Function
Input Layer	360	Relu
Hidden Layer 1	180	Relu
Hidden Layer 2	180	Relu
Hidden Layer 3	90	Relu
Hidden Layer 4	180	Relu
Output Layer	1	Sigmoid

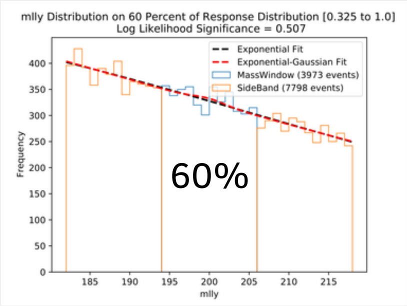
Table showing DNN structure used in the analysis.

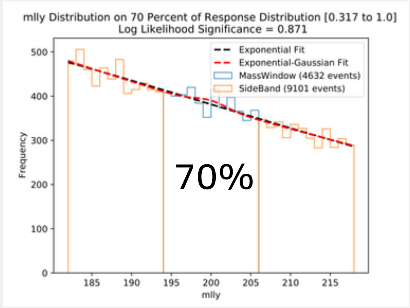
EXTRACTING MLLY DISTRIBUTIONS FROM DNN RESPONSE DISTRIBUTION

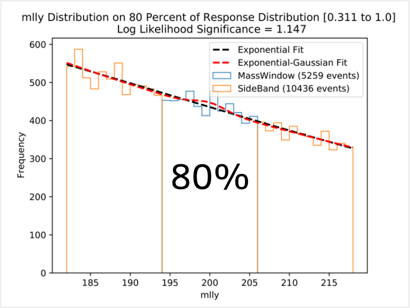


Event batches selected from the Response Distribution maximum, 1.0, to minimum, 0.0.









FITTING TO SIGNAL AND BACKGROUND REGIONS OF INVARIANT MASS

An exponential, f(x), and exponential + gaussian function, g(x), are applied to the mass window and sideband regions of the mlly distribution respectively.

Side-band (background) region:

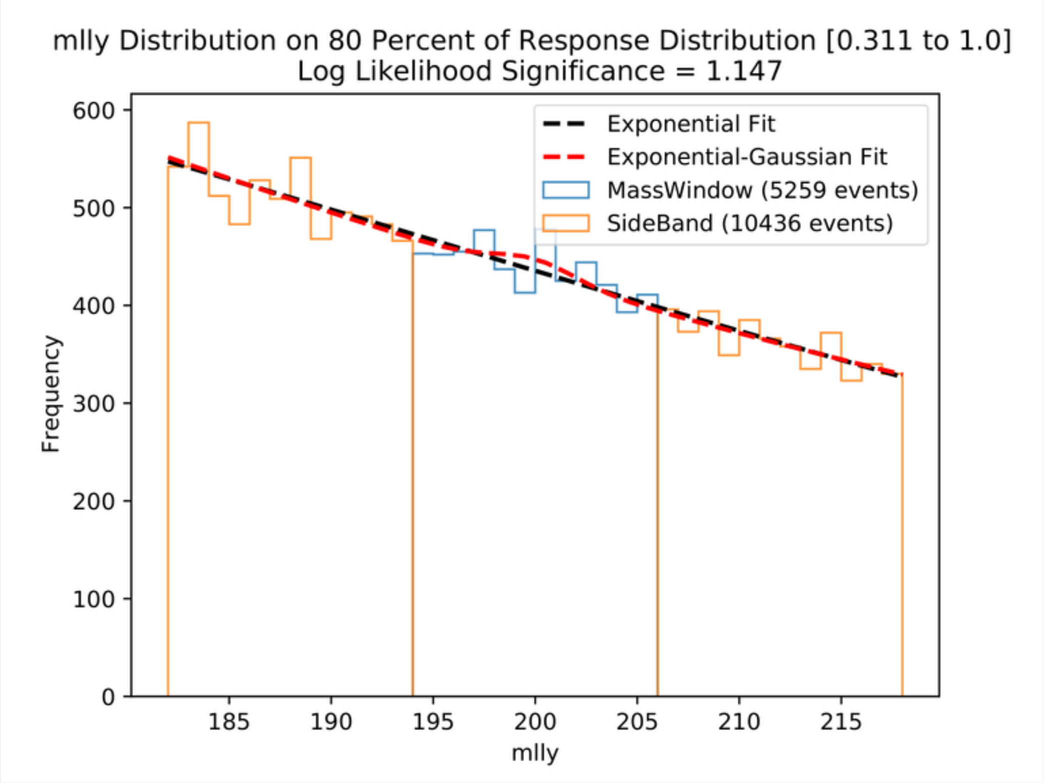
$$f(x) = n_0 * e^{ax+bx^2}$$

Exponential Function

Mass-window (signal) region:

$$g(x) = n_0 * e^{ax+bx^2} + n_1 * e^{\frac{(x-\mu)^2}{2\sigma}},$$

Exponential + Gaussian Function



QUANTIFYING UNCERTAINTY SIGNIFICANCE OF FALSE SIGNALS GENERATED

Probability of event following functional fits:

$$p_X(x_i) = e^{-\lambda \frac{\lambda^{x_i}}{x_i!}}$$

Poisson probability density function

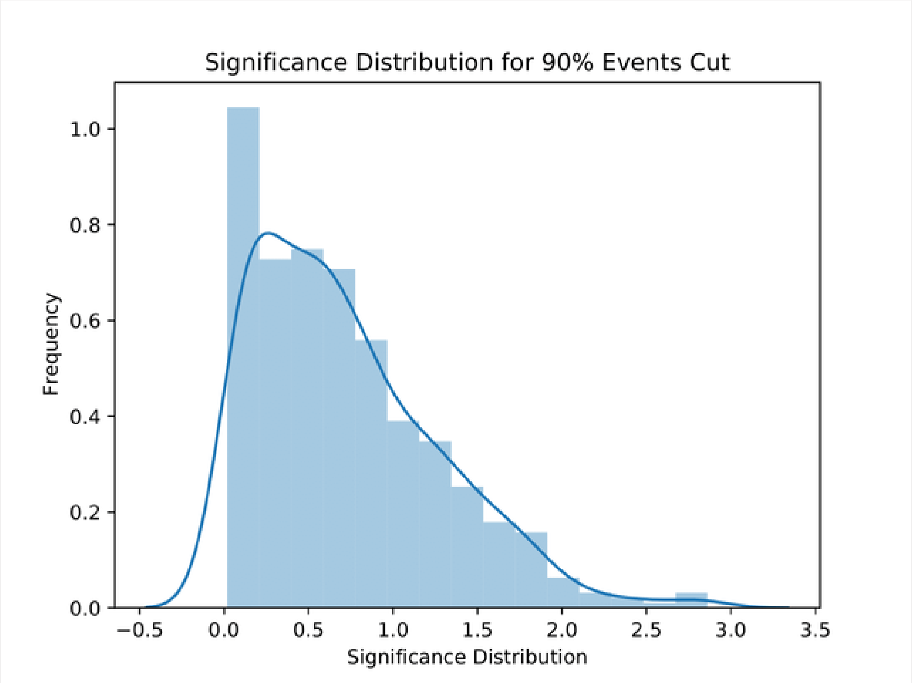
Log likelihood function:

$$\ln L(\lambda; x_1, x_2, ..., x_n) = -n\lambda - \sum_{i=1}^{n} \ln(x_i!) + \ln(\lambda) \sum_{i=1}^{n} x_i$$

The log-likelihood function, ln(L):

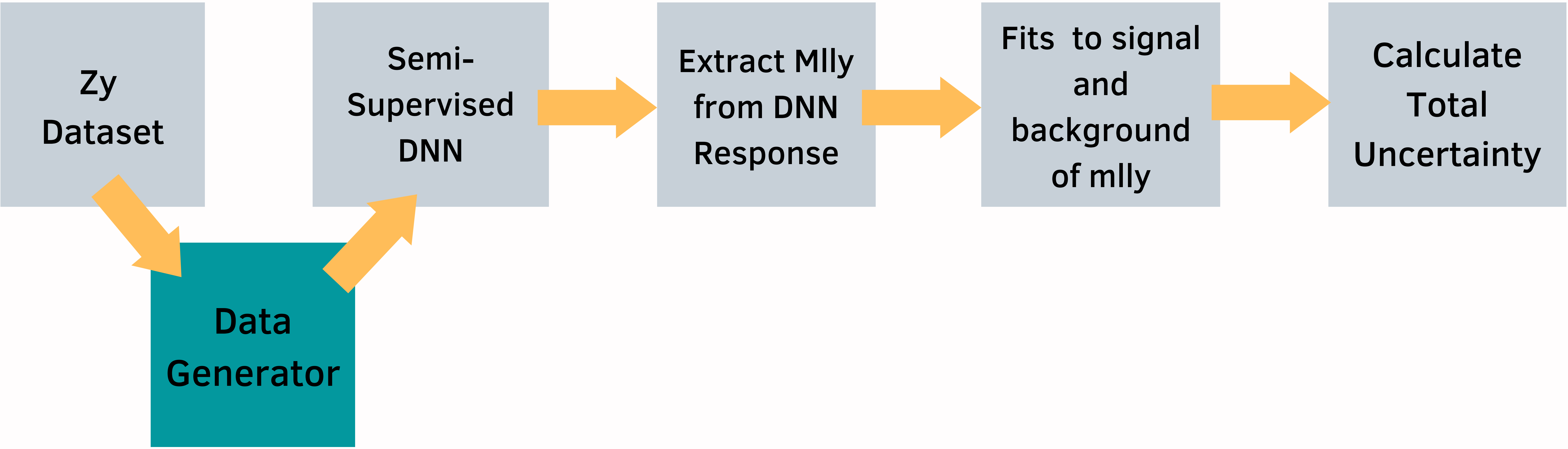
Significance between ln(L) of signal and background:

$$S_k = \sqrt{2 \cdot (\ln L_{eg} - \ln L_e)},$$



NEED FOR FREQUENTIST STUDY - LOOK ELSEWHERE EFFECT

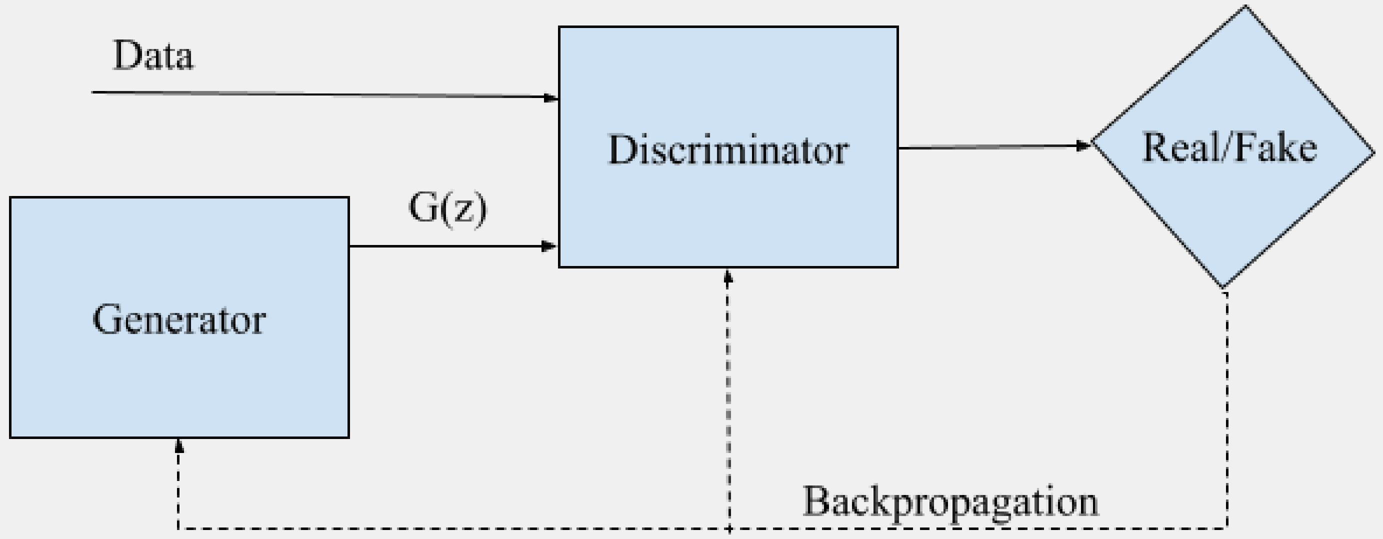
In searches for resonances within a given mass range, the significance of observing a local excess of events, must consider the probability of observing the same excess elsewhere within the range.



GENERATIVE ADVERSARIAL NETWORKS

Loss Function: $E_x[log(D(x))] + E_z[log(1 - D(G(z)))]$

GAN training: minimizing Generator loss and maximise Discriminator loss

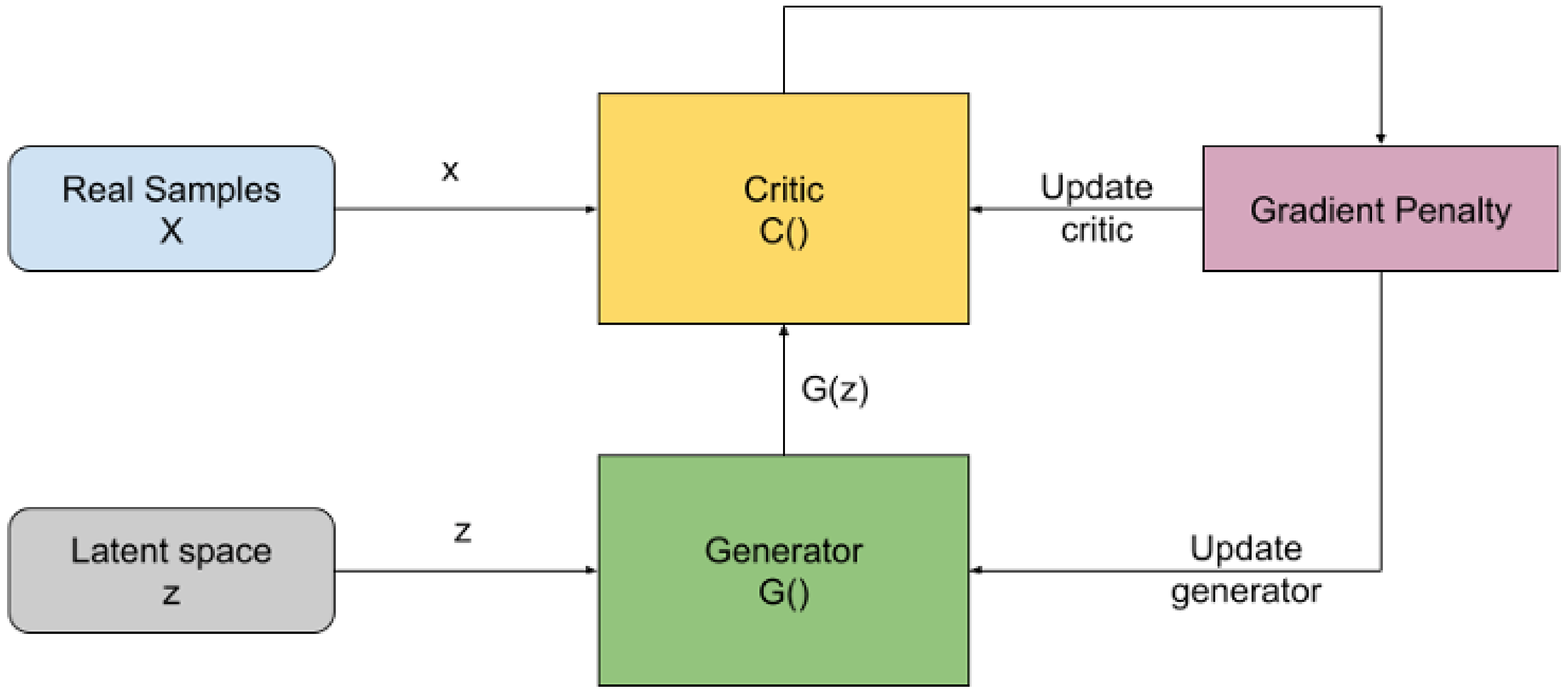


WASSERSTEIN GAN WITH GRADIENT PENALTY

$Loss_{generator} = {}_{min}\ E[\ C(G(z))\ ]$

$Loss_{critic} = {}_{max}\ E[C(x) - C(G(z))\ ] + \lambda * GP$

$GP = E\ [\ (\ ||\ \nabla_x C(x)\ ||_2 - 1\ )^2\ ]$



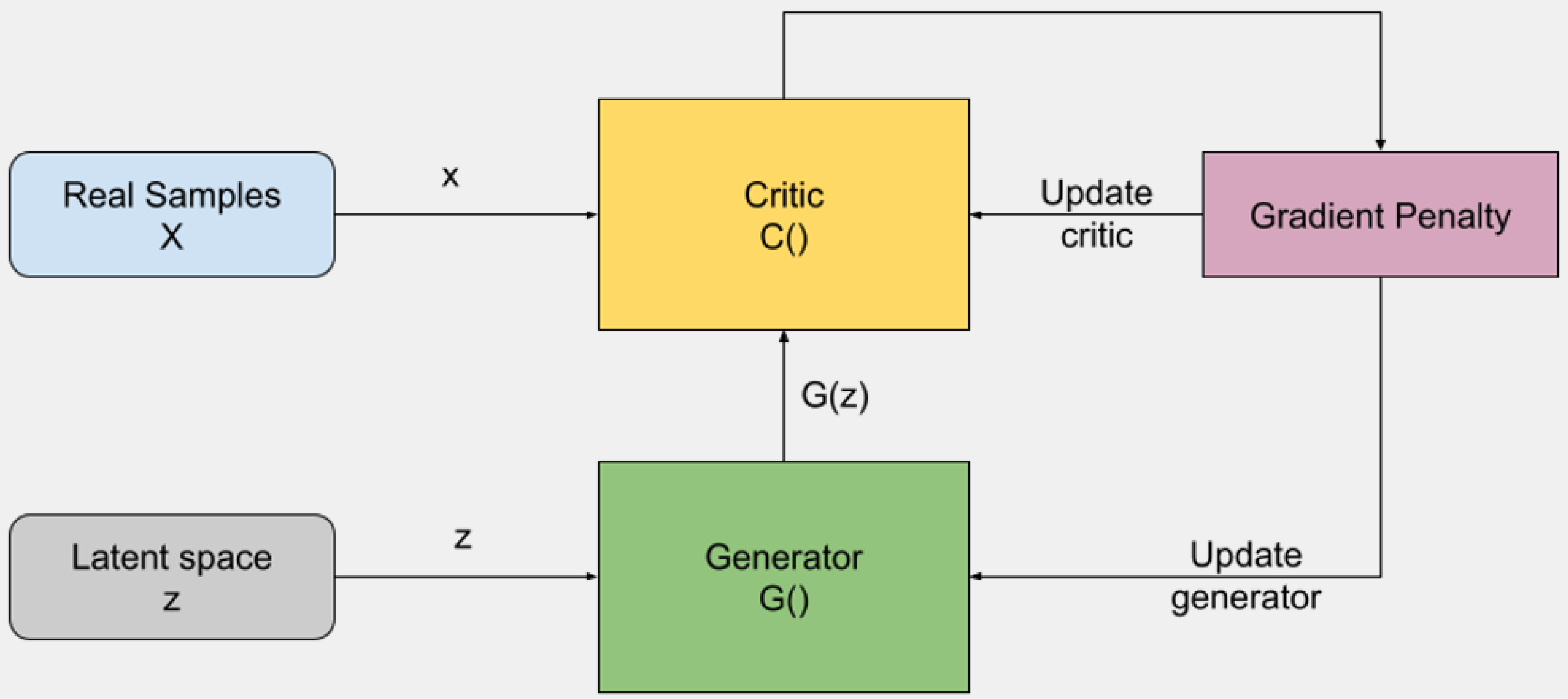
WASSERSTEIN GAN WITH GRADIENT PENALTY

<u>Generator Model Architecture:</u>

```
(0): Linear(in_features=16, out_features=512, bias=True)
(1): ReLU(inplace=True)
(2): Linear(in_features=512, out_features=1024, bias=True)
(3): ReLU(inplace=True)
(4): Linear(in_features=1024, out_features=512, bias=True)
(5): ReLU(inplace=True)
(6): Linear(in_features=512, out_features=18, bias=True)
```

<u>Critic Model Architecture:</u>

```
(0): Linear(in_features=18, out_features=512, bias=True)
(1): ReLU(inplace=True)
(2): Linear(in_features=512, out_features=512, bias=True)
(3): ReLU(inplace=True)
(4): Linear(in_features=512, out_features=512, bias=True)
(5): ReLU(inplace=True)
(6): Linear(in_features=512, out_features=1, bias=True)
```



<u>Hyper-parameters:</u>

latent dimension = 16
batch size = 128
learning rate = 1e-4
lamda (gp weight) = 0.001

data transform = minmax scaler

$$Loss_{generator} = {}_{min}\ E[\,C(G(z))\,]$$

$$Loss_{critic} = {}_{max}\ E[C(x) - C(G(z))\,] + \lambda * GP$$

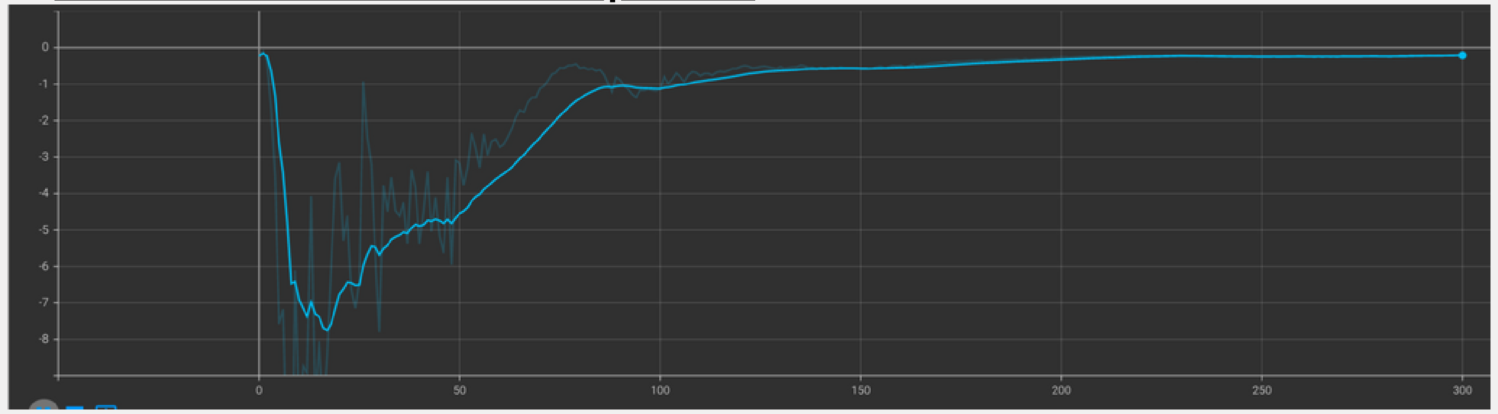
$$GP = E\,[\,(\,||\,\nabla_x C(x)\,||_2 - 1\,)^2\,]$$

<u>Update Ratio</u>

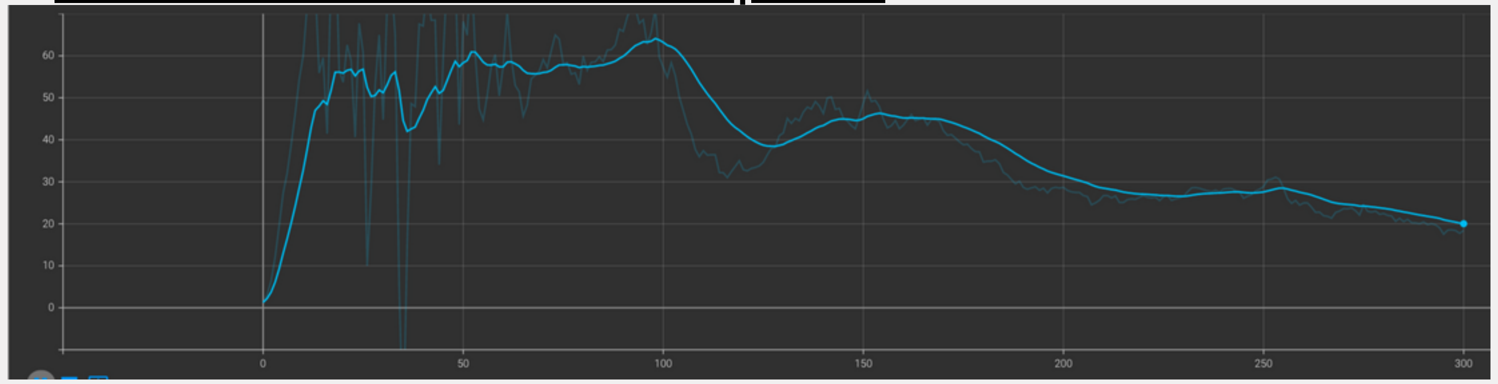
Critic : Generator
5 : 1

WGAN TRAINING

Critic Loss at each Epoch:



Generator Loss at each Epoch:

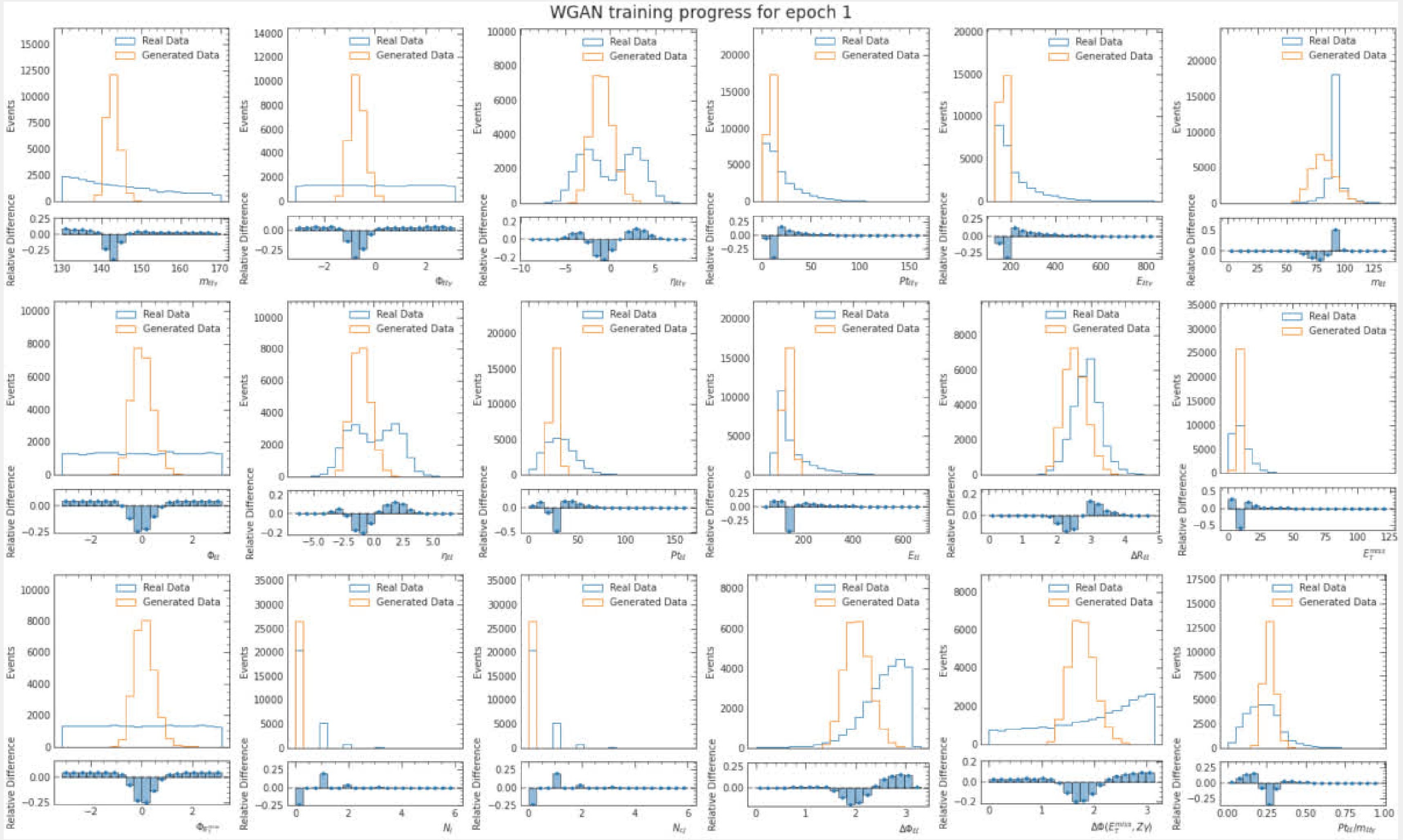


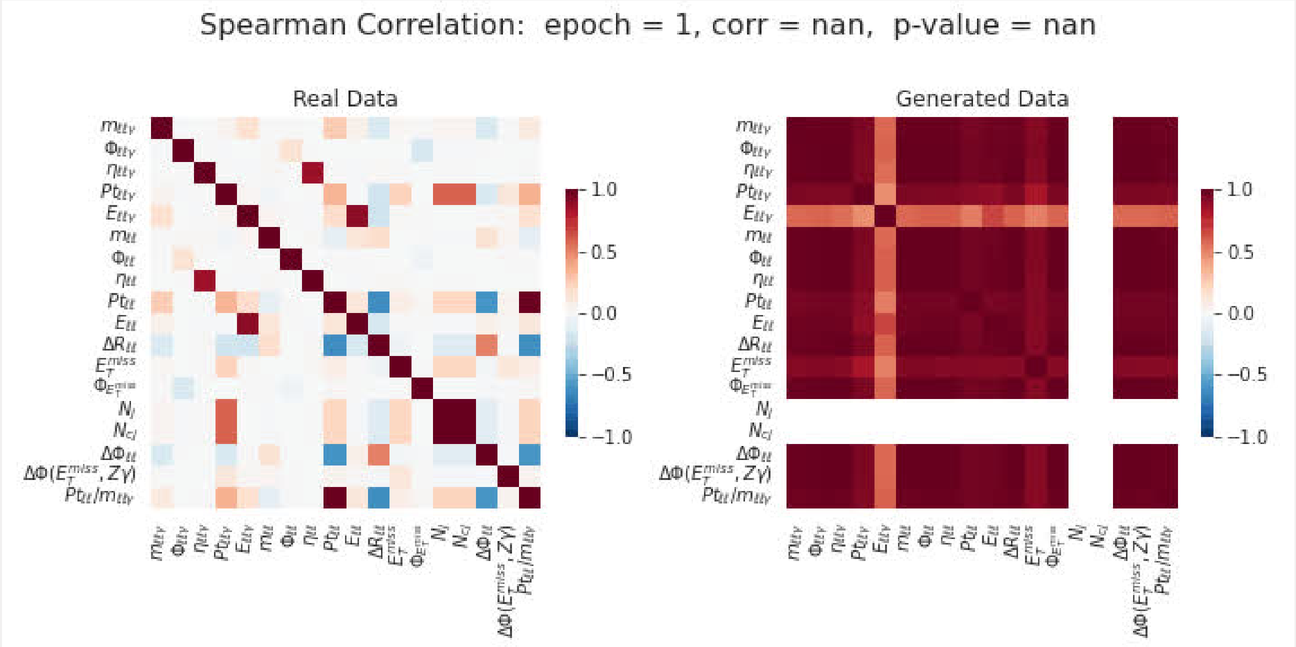
Batch size Hyper-parameter:

- increasing batch size - slows training but difficulty converging
- decreasing batch size - speeds up training but reduces success at very small sizes

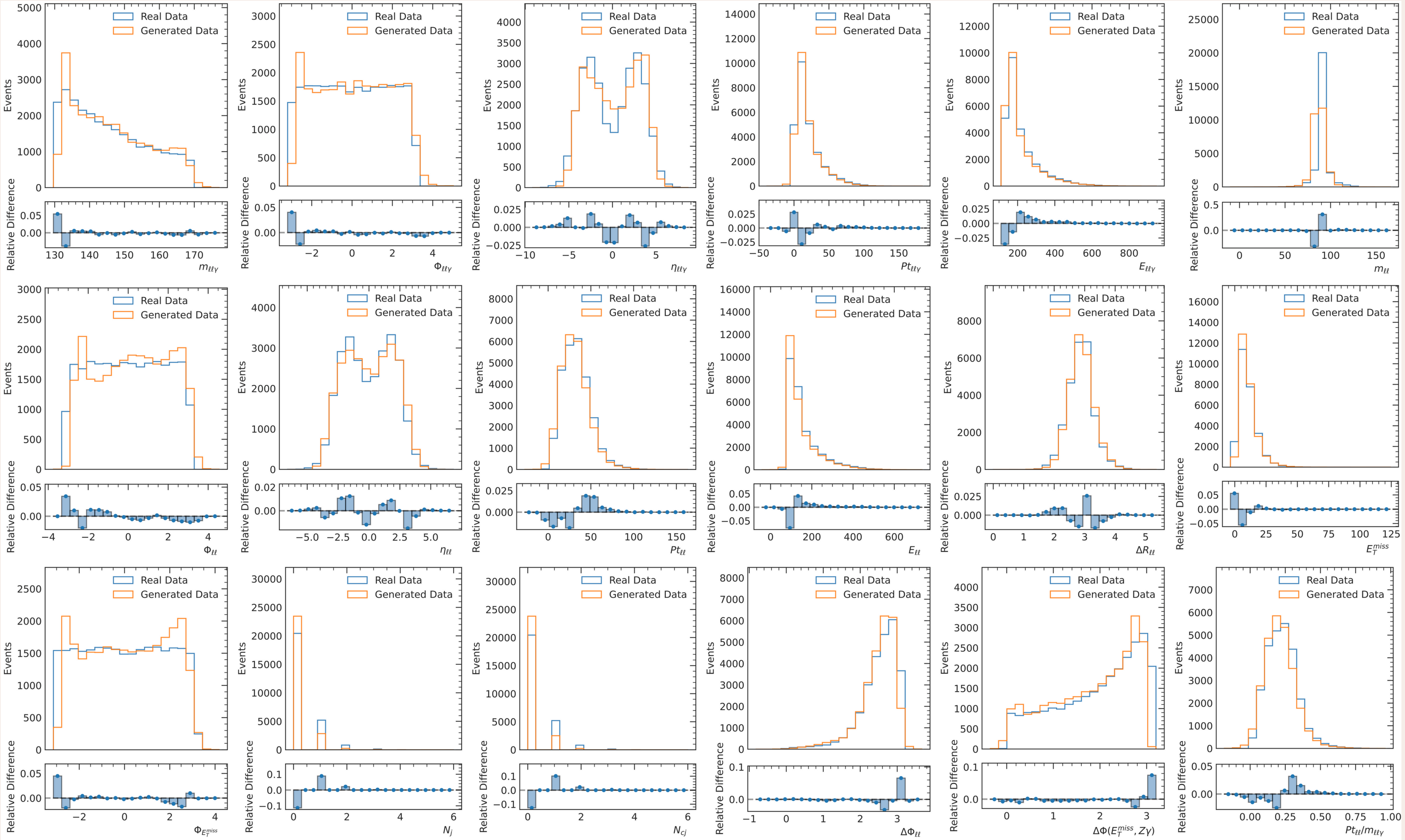
Model built using pytorch and trained using google colaboratory GPU. 500 epochs trained in ±3 hours.

TRAINING GIFS

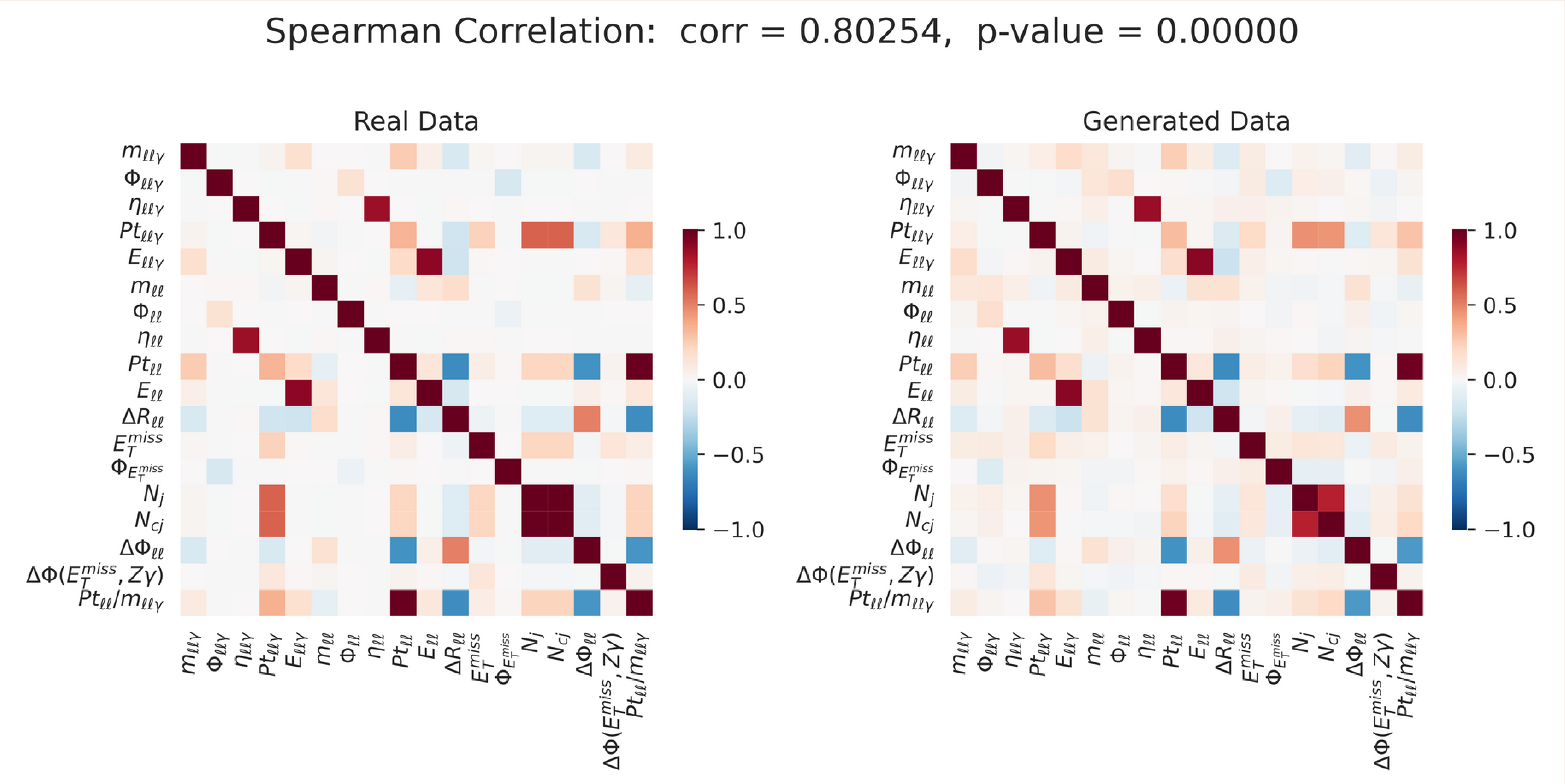




WGAN RESULTS - FEATURE DISTRIBUTIONS



WGAN RESULTS - FEATURE CORRELATION



CONCLUSIONS

WGAN FOR USE IN SCALING SUB-ATOMIC PARTICLE PHYSICS DATASETS

The model is able to a reasonable extent produce a synthetic dataset that follows the feature distributions and event-wise feature correlation of the training dataset.

FUTURE WORK

Further optimisation of generator and critic network architecture for improved results. Completing frequentist analysis of semi-supervised uncertainty quantification with inclusion of GANs.

REFERENCES

1. Von Buddenbrock, S., Chakrabarty, N., Cornell, A.S. et al. "Phenomenological signatures of additional scalar bosons at the LHC". Eur. Phys. J. C 76, 580 (2016). https://doi.org/10.1140/epjc/s10052-016-4435-8
2. ATLAS Collaboration Collaboration, "Measurement of Zγ → l+l−γ differential √s = 13 TeV with the ATLAS detector, tech. rep., CERN, Geneva", 2019. All figures including auxiliary figures are available at cross-sections in pp collisions at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-034.
3. Y. Hernandez, M. Kumar, A. S. Cornell, S.-E. Dahbi, Y. Fang, B. Lieberman, B. Mellado, K. Monnakgotla, X. Ruan, and S. Xin, "The anomalous production of multi-leptons and its impact on the measurement of wh production at the lhc", The European Physical Journal C 81 (Apr, 2021) 365
4. S. eddine Dahbi, J. Choma, B. Mellado, G. Mokgatitswane, X. Ruan, T. Celik, and B. Lieberman, "Machine learning approach for the search of resonances with topological features at the large hadron collider", 2021. International Journal of Modern Physics A.
5. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A., “Improved Training of Wasserstein GANs”, 2017.
6. Backes M, Butter A, Plehn T and Winterhalder R, "How to GAN Event Unweighting". 2021 SciPost Phys. 10 089