ttH process data analysis

Kristijonas Silius

DAM project

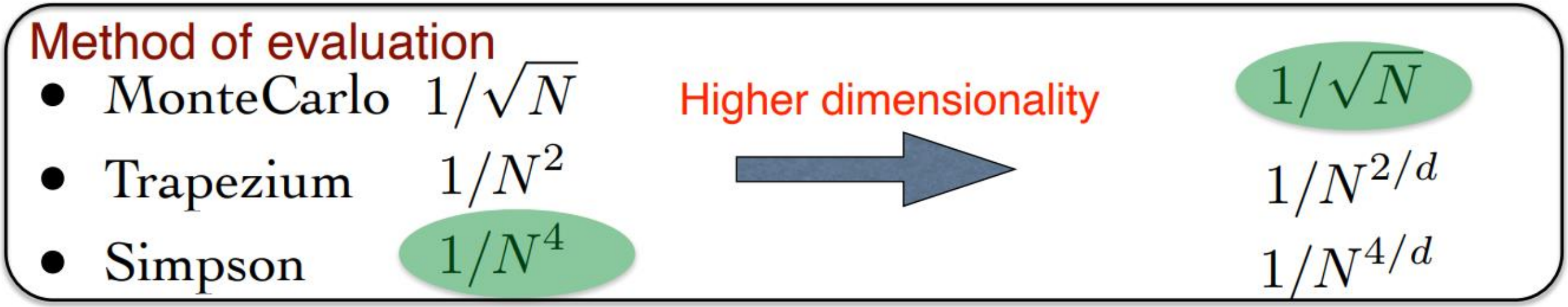
2024

Goals

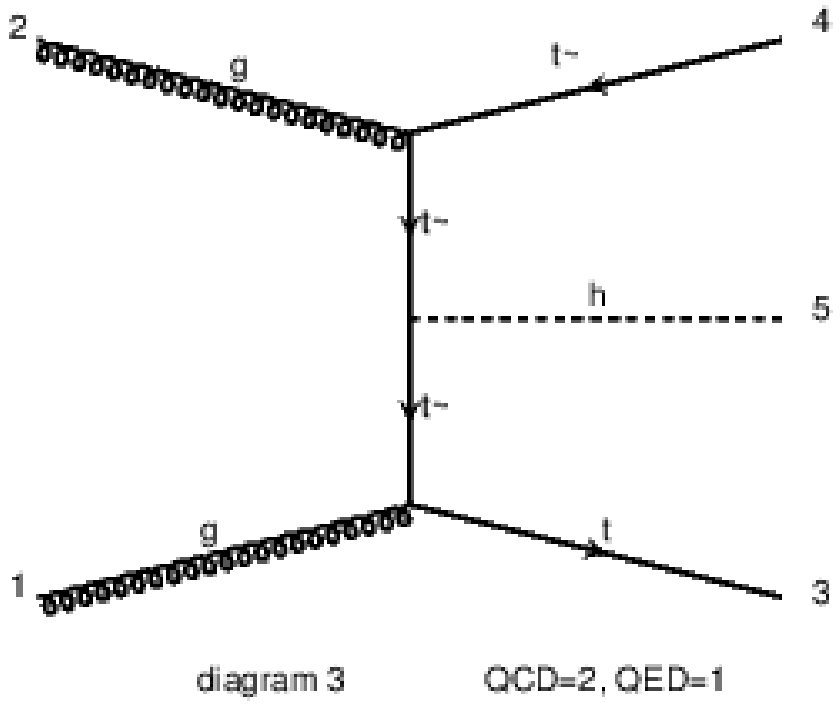
- Generate a particle collision dataset using madgraph software
- Calculate relevant parameters (transverse momentum, pseudorapidity, invariant mass)
- Find correlating parameters
- Visualize the data
- Train a simple regression model to predict invariant mass values

DATASET - generate events via MadGraph5

Monte Carlo event generator – solving phase space integral to generate particle event data



Process: p p > t t~ h



Parse generated event data from .lhe file (looks similar to an .html)

unweighted_events.html

home > ksilius > Downloads > MG5_aMC_v3_5_1 > ttH_first_try > Events > run_01 > unweighted_events.html > LesHouchesEvents > header

```
18293  5      1 +3.9984600e-01 2.75585100e+02 7.54677100e-03 1.11412500e-01
18294        21 -1    0    0  501  502 +0.0000000000e+00 +0.0000000000e+00 +2.0110498484e+03 2.0110498484e+03 0.0000000000e+00 0.0000e+00 -1.0000e+00
18295        21 -1    0    0  502  503 -0.0000000000e+00 -0.0000000000e+00 -1.4769733205e+02 1.4769733205e+02 0.0000000000e+00 0.0000e+00 1.0000e+00
18296         6  1    1    2  501    0 -3.7735007045e+01 -2.0046702125e+02 +1.4039547900e+03 1.4292057268e+03 1.7300000000e+02 0.0000e+00 -1.0000e+00
18297        -6  1    1    2    0  503 +3.3812676473e+01 -4.8419716203e+01 -1.0704529752e+01 1.8311568192e+02 1.7300000000e+02 0.0000e+00 1.0000e+00
18298        25  1    1    2    0    0 +3.9223305718e+00 +2.4888673745e+02 +4.7010225617e+02 5.4642577173e+02 1.2500000000e+02 0.0000e+00 0.0000e+00
18299  <mgrwt>
18300  <rscale>  2 0.24870661E+03</rscale>
18301  <asrwt>0</asrwt>
18302  <pdfrwt beam="1">  1       21 0.30939226E+00 0.27558507E+03</pdfrwt>
18303  <pdfrwt beam="2">  1       21 0.22722668E-01 0.27558507E+03</pdfrwt>
18304  <totfact> 0.49496147E+02</totfact>
18305  </mgrwt>
18306  </event>
18307  <event>
18308  5      1 +3.9984600e-01 2.13899100e+02 7.54677100e-03 1.14207900e-01
18309        21 -1    0    0  501  502 +0.0000000000e+00 +0.0000000000e+00 +5.0161909268e+01 5.0161909268e+01 0.0000000000e+00 0.0000e+00 -1.0000e+00
18310        21 -1    0    0  502  503 -0.0000000000e+00 -0.0000000000e+00 -2.8568169682e+03 2.8568169682e+03 0.0000000000e+00 0.0000e+00 -1.0000e+00
18311         6  1    1    2  501    0 -1.3137873502e+02 -1.0791654017e+02 -8.1821751897e+02 8.5341388552e+02 1.7300000000e+02 0.0000e+00 1.0000e+00
18312        -6  1    1    2    0  503 +1.5543991920e+01 +9.1428228556e+01 -1.7446685610e+03 1.7556759736e+03 1.7300000000e+02 0.0000e+00 1.0000e+00
18313        25  1    1    2    0    0 +1.1583474310e+02 +1.6488311612e+01 -2.4376897902e+02 2.9788901836e+02 1.2500000000e+02 0.0000e+00 0.0000e+00
18314  <mgrwt>
18315  <rscale>  2 0.20935397E+03</rscale>
18316  <asrwt>0</asrwt>
18317  <pdfrwt beam="1">  1       21 0.77172148E-02 0.21389909E+03</pdfrwt>
18318  <pdfrwt beam="2">  1       21 0.43951042E+00 0.21389909E+03</pdfrwt>
18319  <totfact> 0.60282610E+02</totfact>
18320  </mgrwt>
18321  </event>
18322  <event>
18323  5      1 +3.9984600e-01 2.65030000e+02 7.54677100e-03 1.10415800e-01
18324         2 -1    0    0  502    0 +0.0000000000e+00 +0.0000000000e+00 +1.8382460758e+02 1.8382460758e+02 0.0000000000e+00 0.0000e+00 1.0000e+00
18325        -2 -1    0    0  501    0 -0.0000000000e+00 -0.0000000000e+00 -4.4707280611e+02 4.4707280611e+02 0.0000000000e+00 0.0000e+00 -1.0000e+00
18326         6  1    1    2  502    0 -5.2627408539e+00 +9.3434213215e+01 -1.0231714423e+02 2.2171027636e+02 1.7300000000e+02 0.0000e+00 1.0000e+00
18327        -6  1    1    2    0  501 +8.2560277376e+01 -1.8703887716e+01 -2.2516132727e+01 1.9391237982e+02 1.7300000000e+02 0.0000e+00 1.0000e+00
18328        25  1    1    2    0    0 -7.7297536522e+01 -7.4730325499e+01 -1.3841492158e+02 2.1527475750e+02 1.2500000000e+02 0.0000e+00 0.0000e+00
18329  <mgrwt>
18330  <rscale>  2 0.26503005E+03</rscale>
18331  <asrwt>0</asrwt>
18332  <pdfrwt beam="1">  1        2 0.28280709E-01 0.26503005E+03</pdfrwt>
18333  <pdfrwt beam="2">  1       -2 0.68780430E-01 0.26503005E+03</pdfrwt>
```

Parsed .csv file

This example has 10k events

Possible to generate a new batch with more

top

Energy px py pz top~ higgs

new_data.csv

home > ksilius > Documents > ttH > new_data.csv

```
9979  565.0405292,-79.062803729,27.966495275,531.32734528,324.29393593,162.96034962,-92.669709569,200.23437832,658.9238438,-83.897545891,64.703214293,638.22435536
9980  570.60602103,100.75455507,176.39648062,-504.37588413,283.47845415,-4.8261910656,103.79597702,199.08324139,380.81255991,-95.928364009,-280.19245764,-204.16449604
9981  609.88833843,-20.76930727,-287.43289358,-508.90642845,672.02863399,-28.782780489,112.2822023,-638.9505016,1522.0273891,49.55208776,175.15069128,-1505.924699
9982  548.62954775,287.1058308,-65.801038382,-429.30856728,1193.6559773,-165.32620129,27.457396722,-1169.1017623,2139.2848243,-121.7796295,38.34364166,-2131.8100399
9983  250.41420619,-116.8759829,22.478687589,-136.42942456,212.10595935,53.478963242,-28.36444758,-106.74922293,236.05977555,63.397019656,5.8857599903,-189.85624394
9984  389.68046447,-100.67149755,42.426129699,331.6430875,217.38131241,99.692771737,-73.121392583,45.169106658,557.95899014,0.9787258088,30.695262884,542.9089036
9985  301.35907165,174.41151827,173.43881897,-19.694883376,386.57305738,-95.018975478,-273.31347576,189.1583119,398.54425805,-79.392542794,99.874656787,356.27854652
9986  472.14825953,212.01090408,-79.148940994,376.53924187,903.17688574,-207.7001406,200.68381944,838.08480666,213.50946415,-4.3107634824,-121.53487845,123.17459934
9987  1010.1033895,54.627826275,72.821068232,991.00592836,430.49769992,66.195654309,-175.88117397,-346.530255,202.63124502,-120.82348058,103.06010574,-14.653416046
9988  778.12292773,-192.50427103,155.50292282,717.15217164,783.94279258,116.45167571,-77.574127578,-751.7037741,324.14934824,76.052595324,-77.928795238,-278.55144154
9989  227.56513524,144.45182902,31.293951237,3.3538760787,475.44430526,-193.86765354,-120.64369014,379.44527978,161.44354727,49.415824526,89.349738905,-3.7039706274
9990  409.90542112,15.141811649,17.604112061,370.88310159,845.62373728,23.766204315,-87.769581278,822.72849299,555.7729859,-38.908015963,70.165469217,535.55726592
9991  479.19353567,53.924830395,-36.908190509,442.0716489,371.49107266,-133.21439593,-53.225474938,295.79991649,322.71285621,79.289565535,90.133665447,272.22724829
9992  267.30165462,-60.620188935,10.963573628,-194.23225095,960.32435505,99.324764959,21.153422619,-939.13842999,159.15496359,-38.704576025,-32.116996247,-84.709838765
9993  384.43545906,214.32119949,42.774864621,264.76094234,284.09306999,-214.00495827,-60.945159119,35.601093178,127.16677146,-0.3162412239,18.170294498,14.701297673
9994  232.20841723,79.610690956,101.52943874,85.706825798,295.09175621,-75.552950174,-44.402433185,-222.41924429,147.34955161,-4.0577407822,-57.127005555,-52.980471274
9995  368.7609152,-33.362034537,240.98423411,-216.49292399,476.19118165,178.43947467,-176.23365422,-365.96460823,270.81756778,-145.07744014,-64.750579893,-180.21391118
9996  175.17444418,-7.1504305085,23.808294597,-11.795013627,239.78050985,56.71161657,26.20159294,-153.82770224,184.97650808,-49.561186062,-50.009887536,116.76475721
9997  236.24341913,-27.348847221,-105.67486341,118.1812882,467.38502312,2.4984903352,190.81912535,390.00202408,177.81361893,24.850356886,-85.144261949,-90.142096135
9998  174.59153448,-15.155283324,-11.626444522,13.723960309,205.42358155,-51.822473595,27.790106294,93.872195447,142.73852657,66.97775692,-16.163661772,-1.4153030631
9999  535.98931047,121.35214193,-167.26455416,463.30526385,1103.3408127,-64.903526247,94.387372892,1083.6560824,198.93102261,-56.448615684,72.877181272,124.30213995
10000 221.52956029,-5.8229915438,28.963893781,135.17962757,323.91136976,4.873082544,64.072638951,266.19640393,158.46352768,0.94990899974,-93.036532732,-28.7922013
10001
```

Conventions and meanings

- Calculations and measurements done using $c = 1$ and $\hbar = 1$
- t – top quark
- t2 – anti-top quark
- h – Higgs boson
- Eta – η – pseudorapidity - spatial coordinate describing the angle between a particle and the beam axis
- p_T – transverse momentum
- M_inv – invariant mass

Analysis start

- Generated new batch with 500 000 data points
- After calculations, data set looks like this [500000 rows x 22 columns]

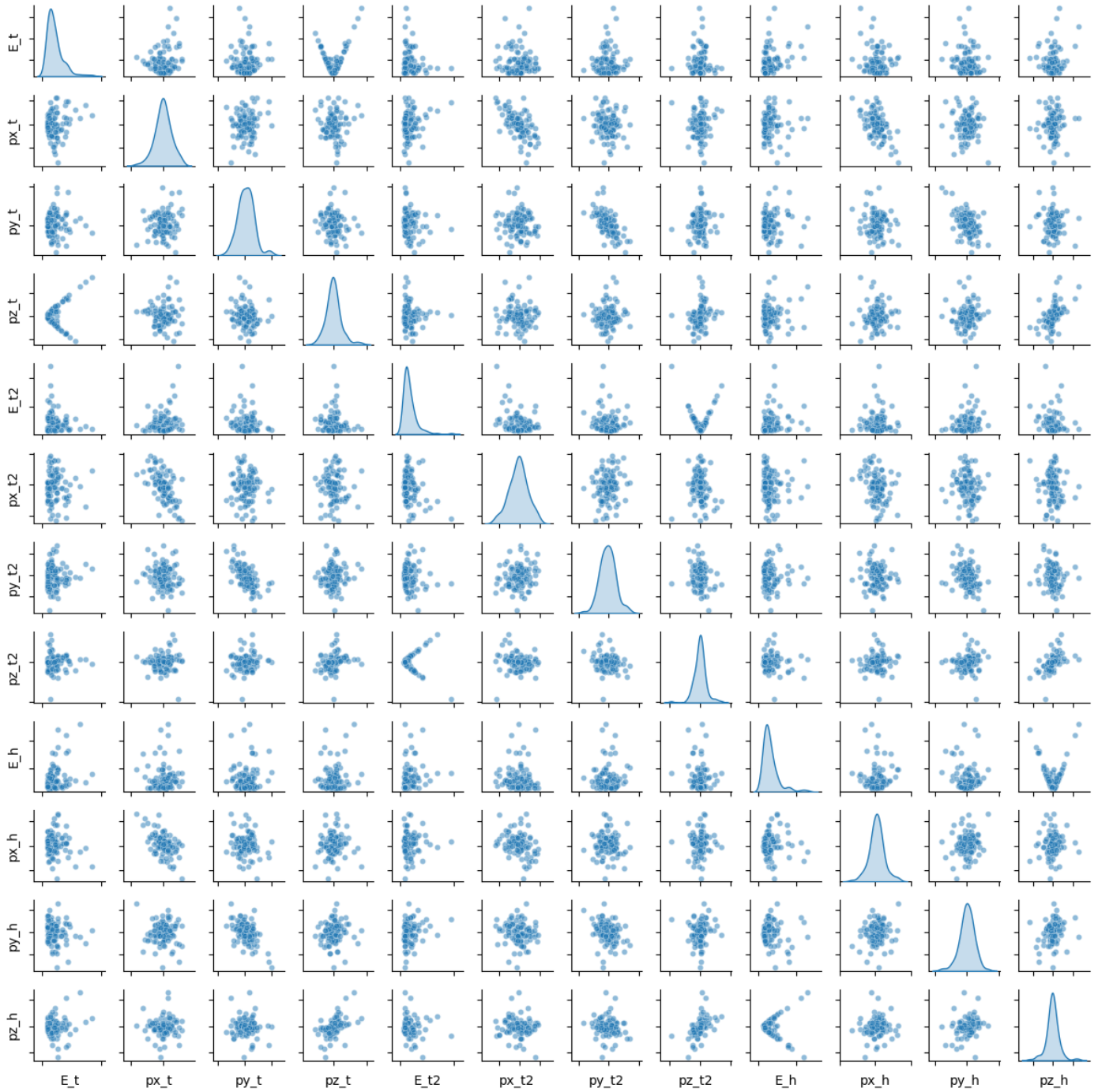
pid_t	E_t	px_t	py_t	pz_t	pT_t	eta_t	pid_t2	E_t2	px_t2	py_t2	pz_t2	pT_t2	eta_t2	pid_h	E_h	px_h	py_h	pz_h	pT_h	eta_h	M_inv
6	463.2705	176.3492	103.7824	377.9163	204.6211	1.372997	-6	387.9539	-162.335	-58.4856	-301.34	172.5489	-1.32411	25	134.5495	-14.0145	-45.2968	15.17787	47.41522	0.314876	979.6296
6	807.4687	258.7377	3.865722	-745.061	258.7665	-1.77956	-6	236.3007	-132.787	-30.7236	85.63138	136.2948	0.592924	25	200.1999	-125.951	26.85792	88.71311	128.7827	0.643516	1103.592
6	549.3163	-343.672	-166.184	-355.094	381.7431	-0.83114	-6	932.1703	459.4385	139.7581	-779.998	480.225	-1.26176	25	196.9337	-115.766	26.42626	95.1719	118.7442	0.733829	1315.65
6	338.0329	-128.756	201.6179	-164.649	239.2235	-0.64303	-6	373.5744	95.73716	-264.733	174.297	281.5127	0.585172	25	240.9867	33.01869	63.11547	-193.328	71.23058	-1.72394	898.8512
6	524.7242	206.2773	154.4485	-423.086	257.691	-1.27095	-6	327.4506	-18.9251	-18.1589	-276.78	26.22794	-3.05178	25	405.7319	-187.352	-136.29	-308.736	231.6802	-1.09817	731.6523

top

anti top

Higgs

All

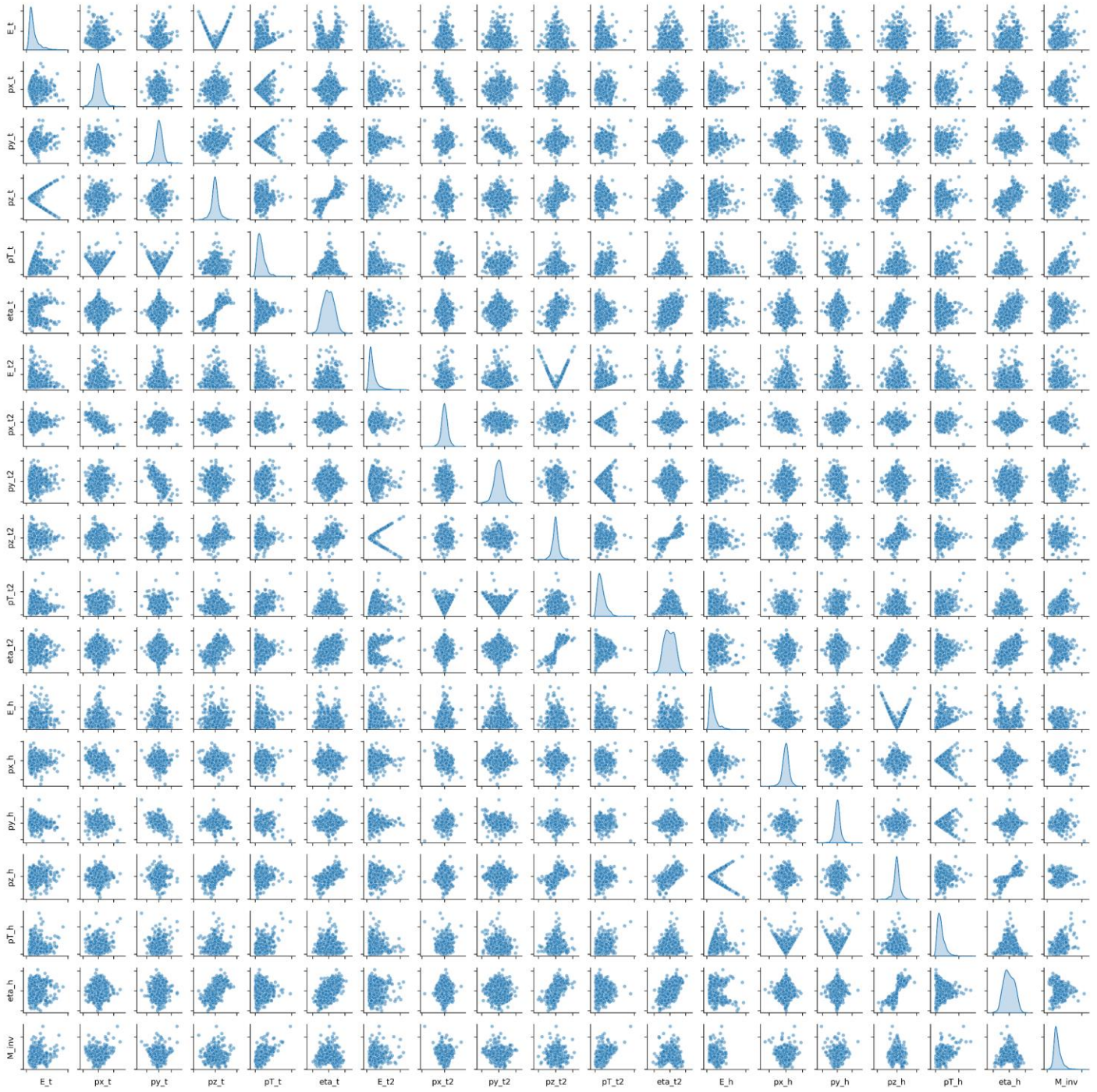


For some reason corner.corner() did not work for me, therefore:

Used seaborn.pairplot() function;

Not even all parameters

Plotting only 100 rows (for speed)



1000 rows plotted

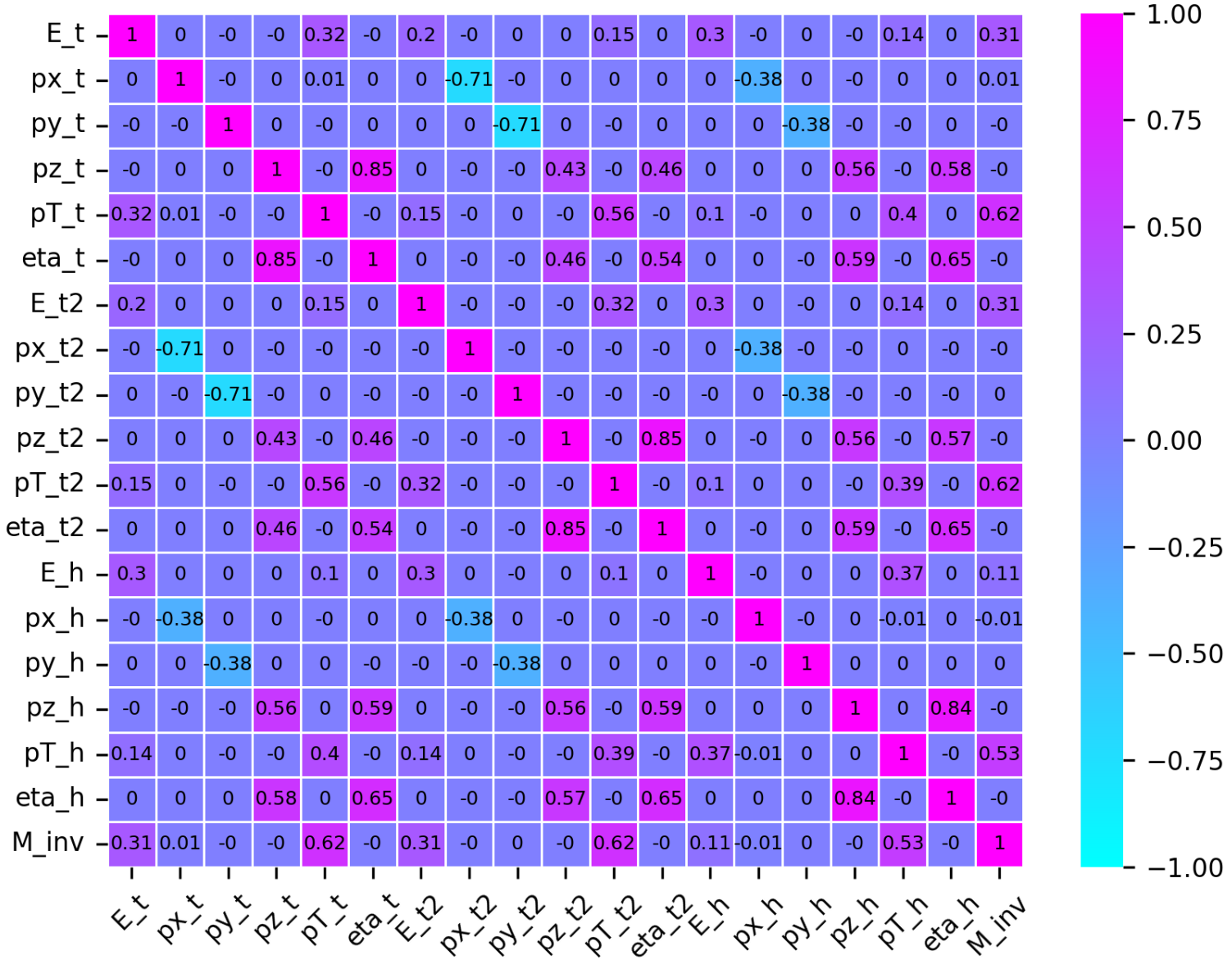
Not very readable, need different approach

Correlation coefficient

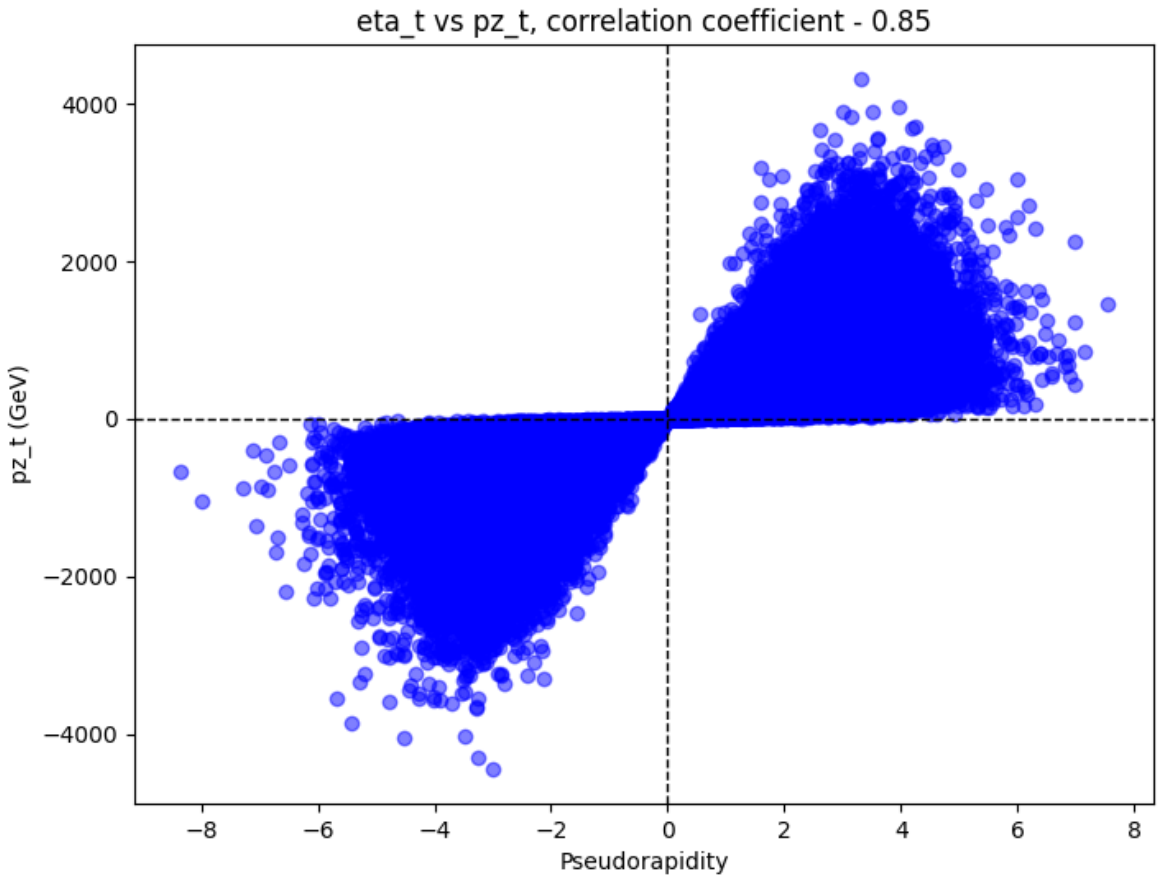
$$\rho_{X,Y} = \operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

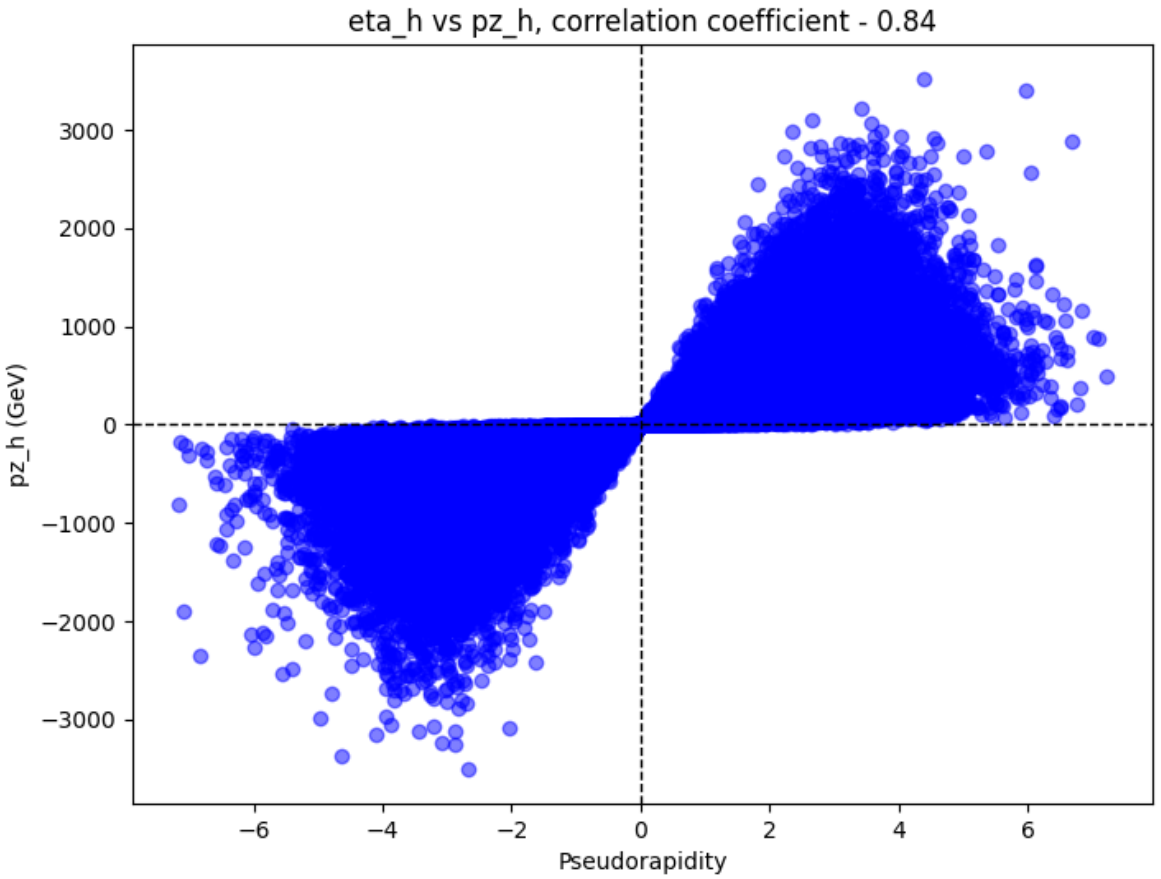
$$\operatorname{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)).$$

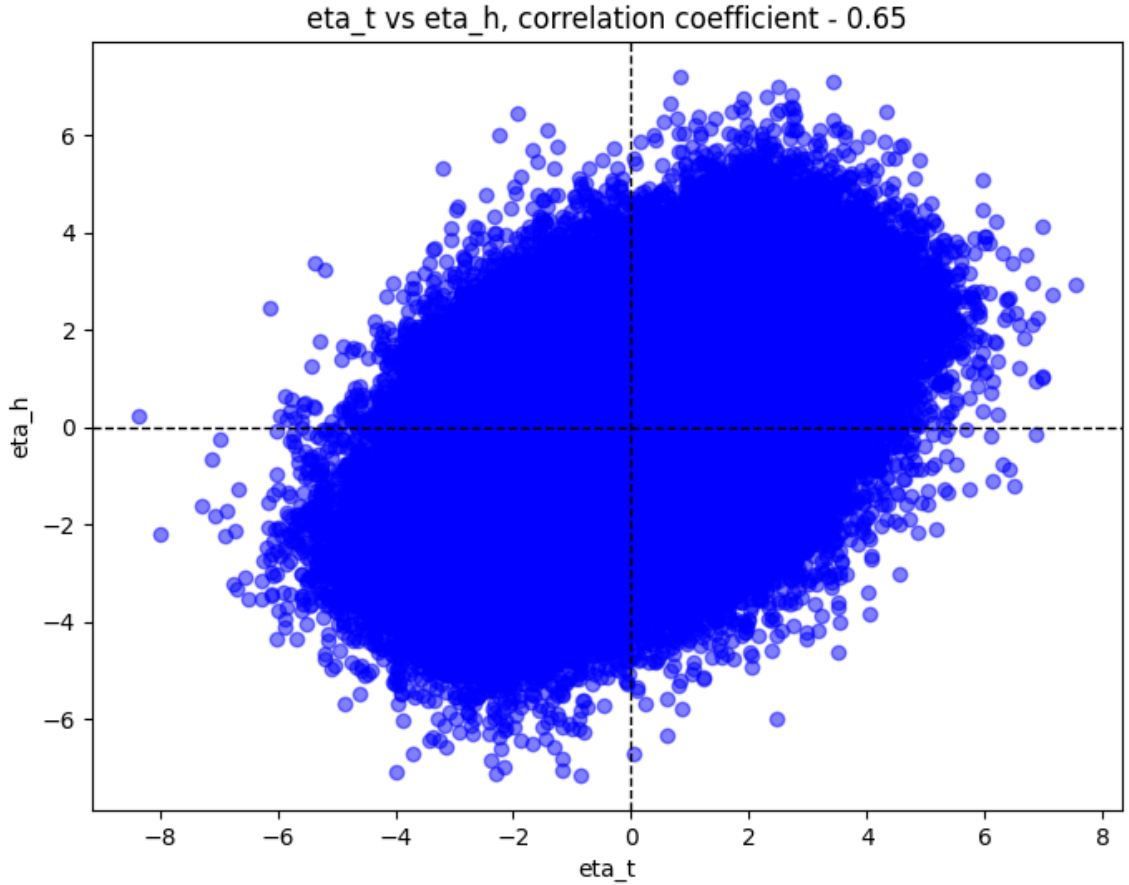
Plotting a correlation matrix

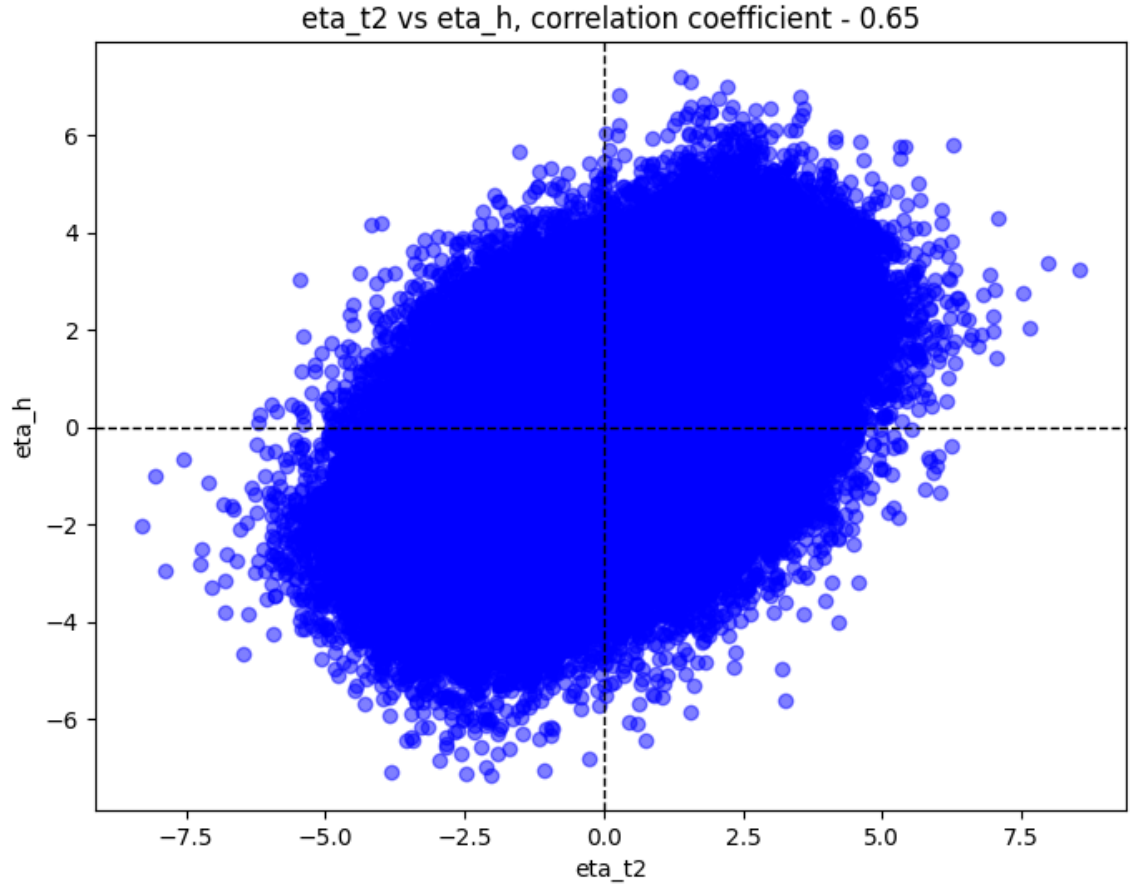


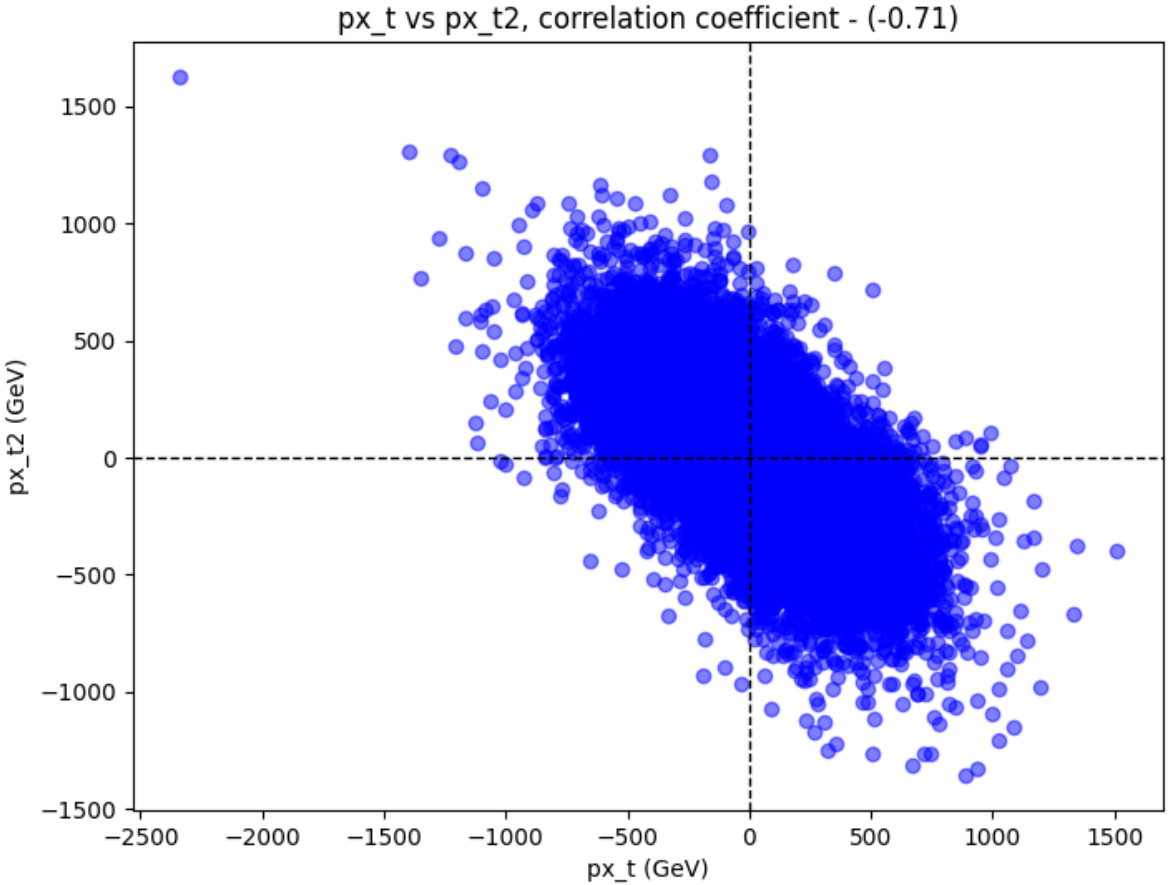
Plotting the higher correlation coefficient plots

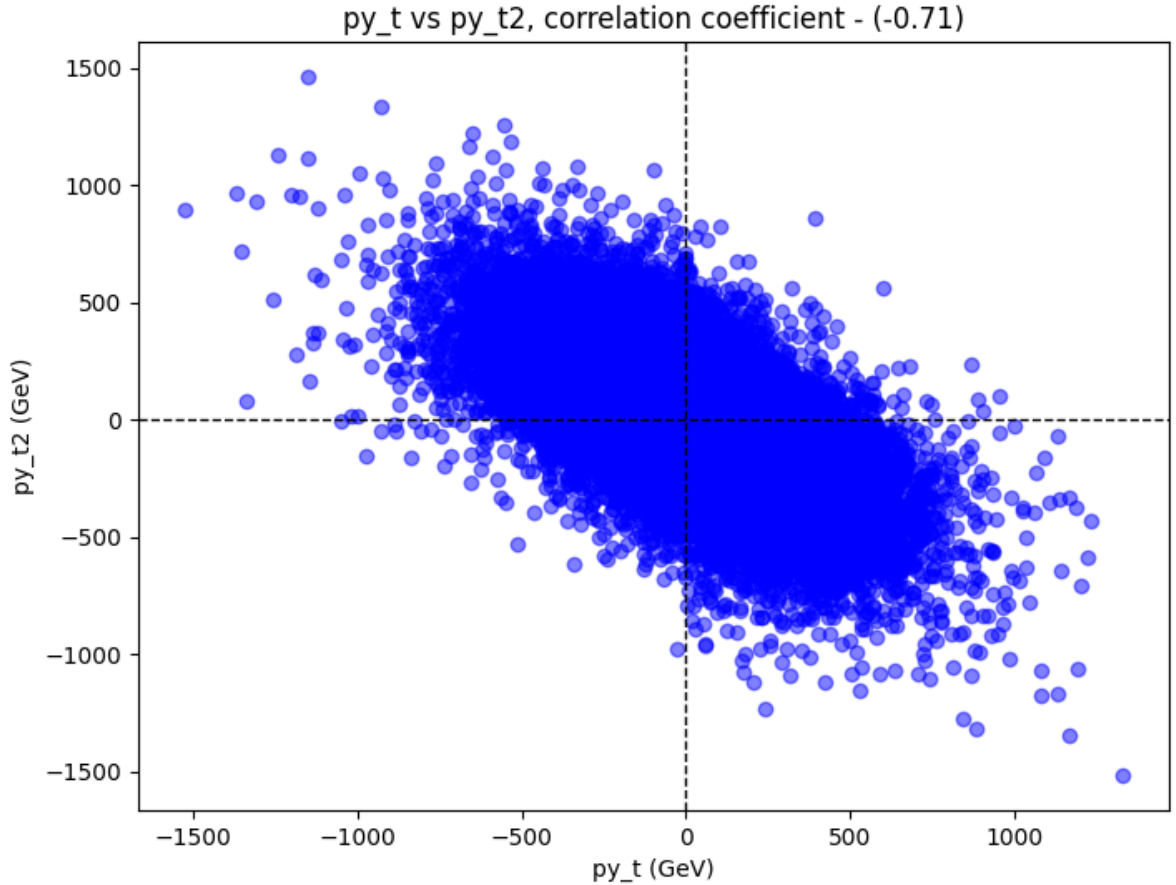


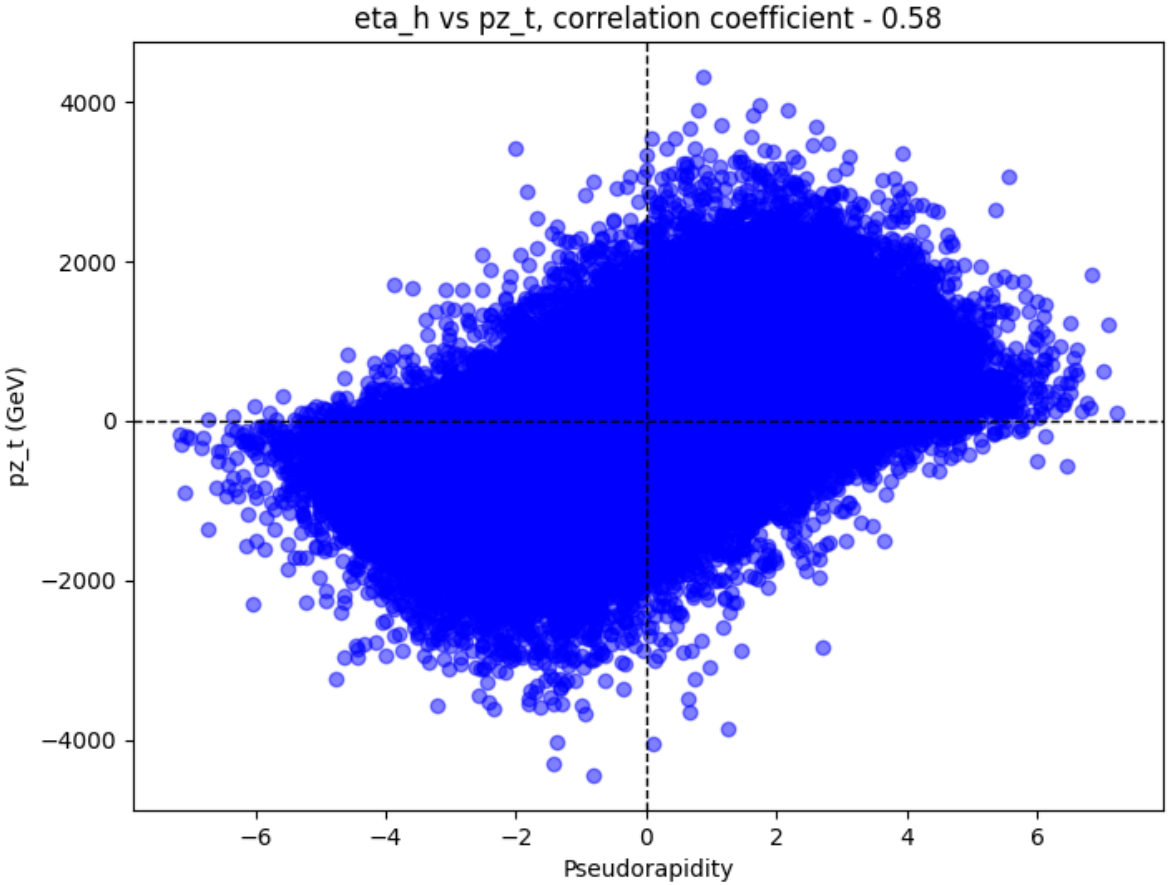


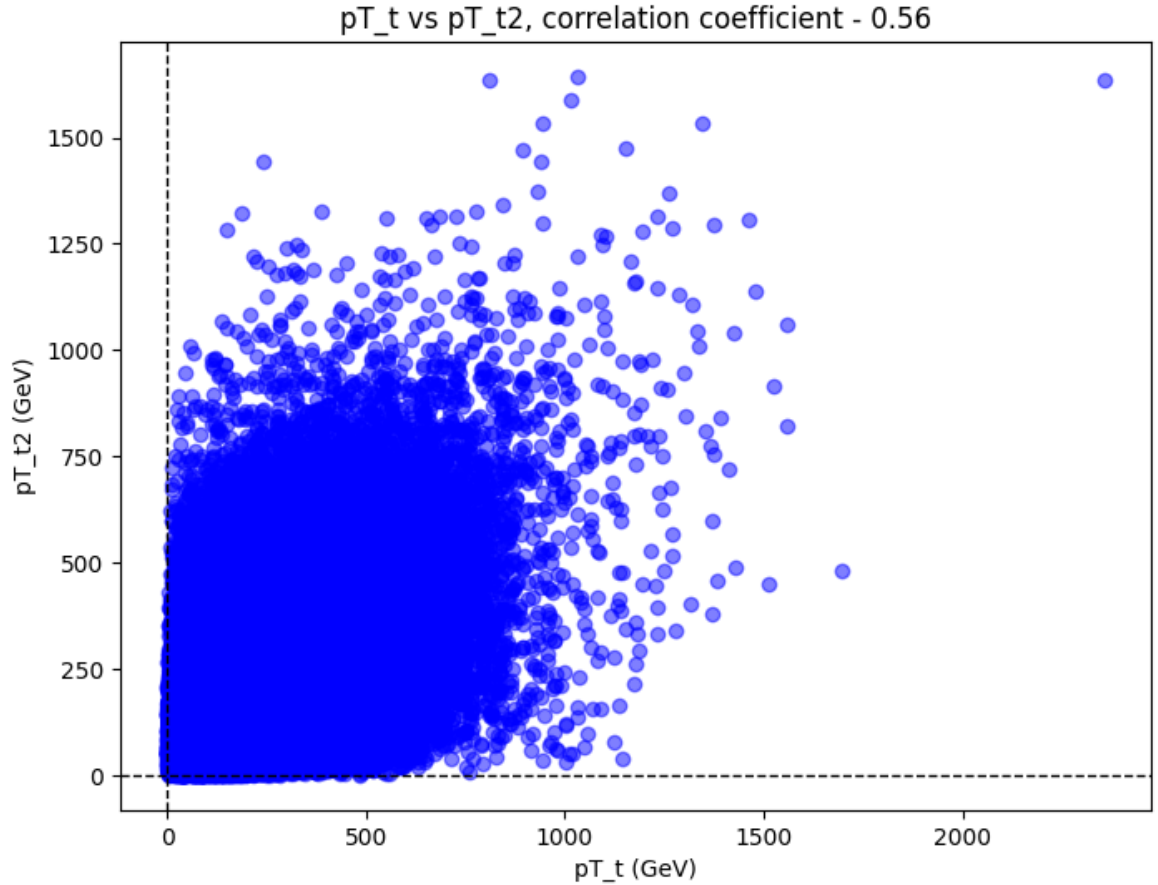

eta_t vs eta_h, correlation coefficient - 0.65
eta_h
eta_t
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6
8

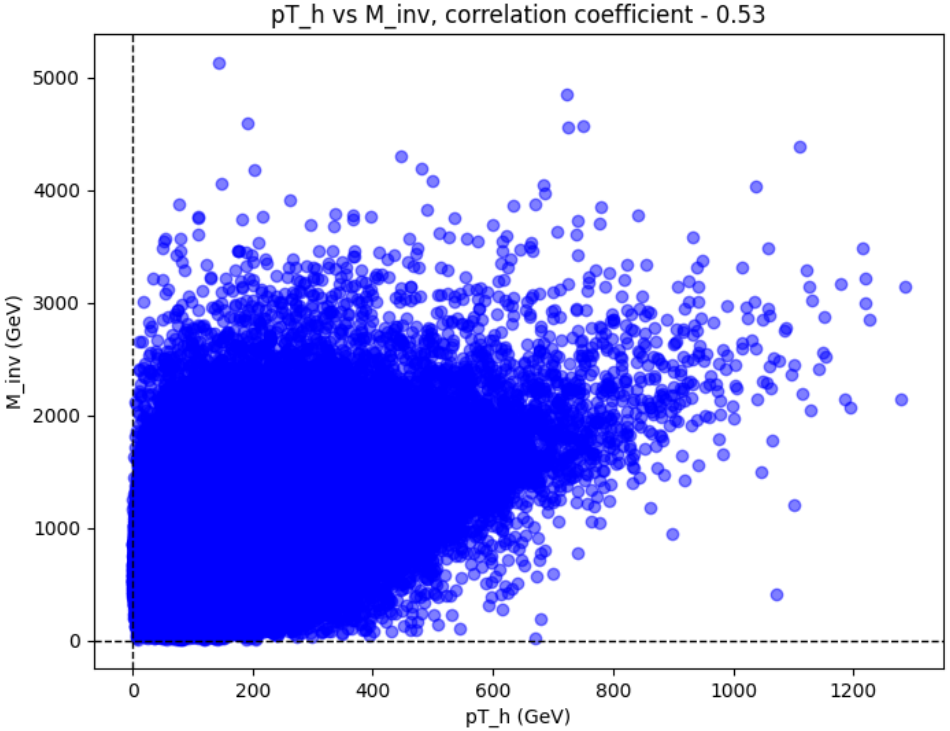

eta_t2 vs eta_h, correlation coefficient - 0.65
eta_h
eta_t2
6
4
2
0
−2
−4
−6
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5

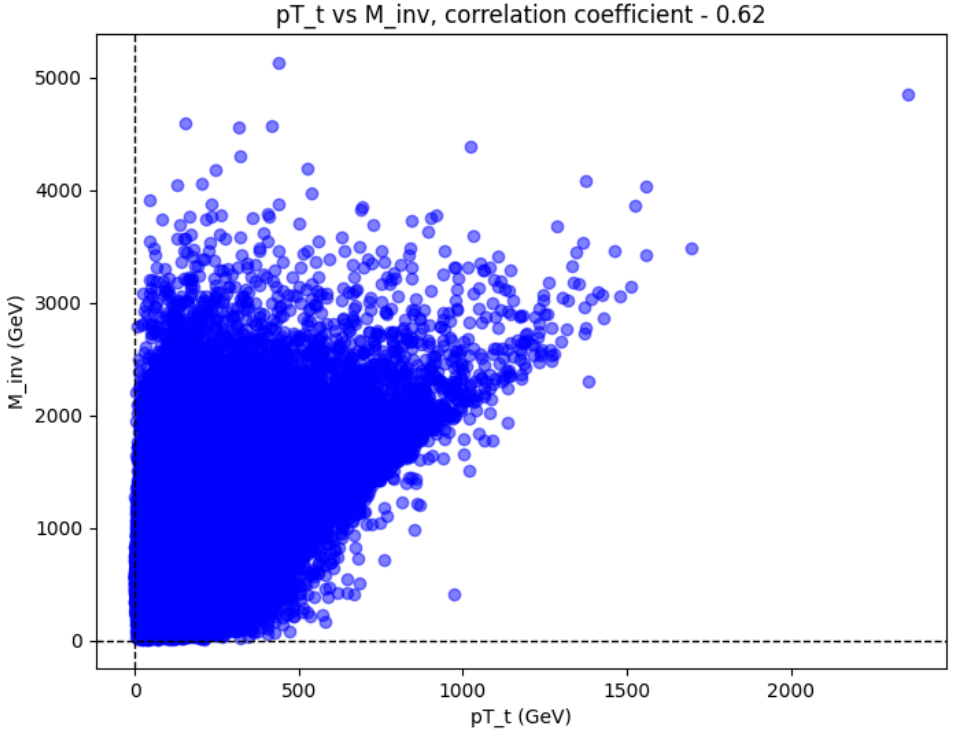

px_t vs px_t2, correlation coefficient - (-0.71)
px_t (GeV)
px_t2 (GeV)
1500
1000
500
0
−500
−1000
−1500
−2500
−2000
−1500
−1000
−500
0
500
1000
1500

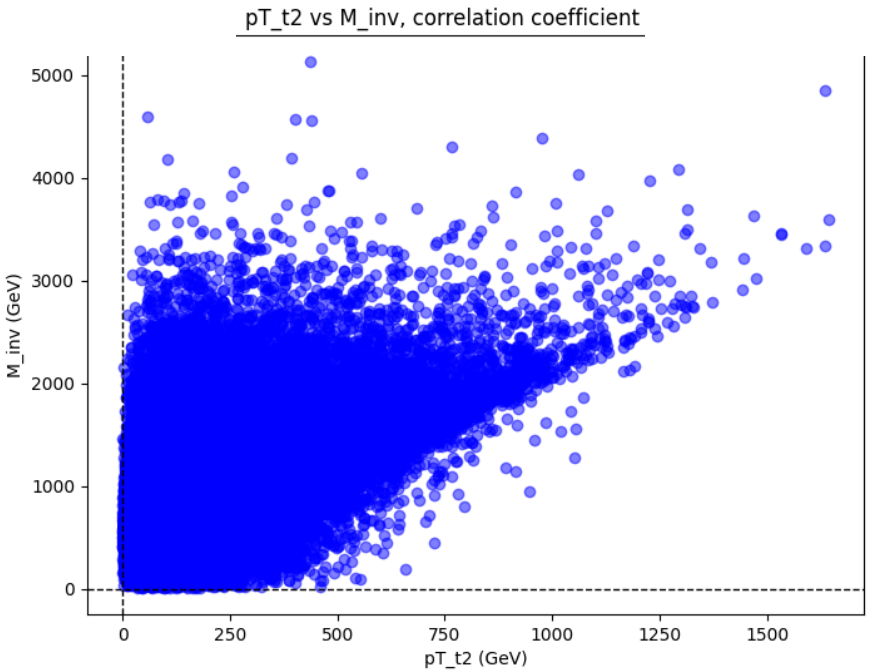

py_t vs py_t2, correlation coefficient - (-0.71)
py_t (GeV)
py_t2 (GeV)
1500
1000
500
0
−500
−1000
−1500
−1500
−1000
−500
0
500
1000

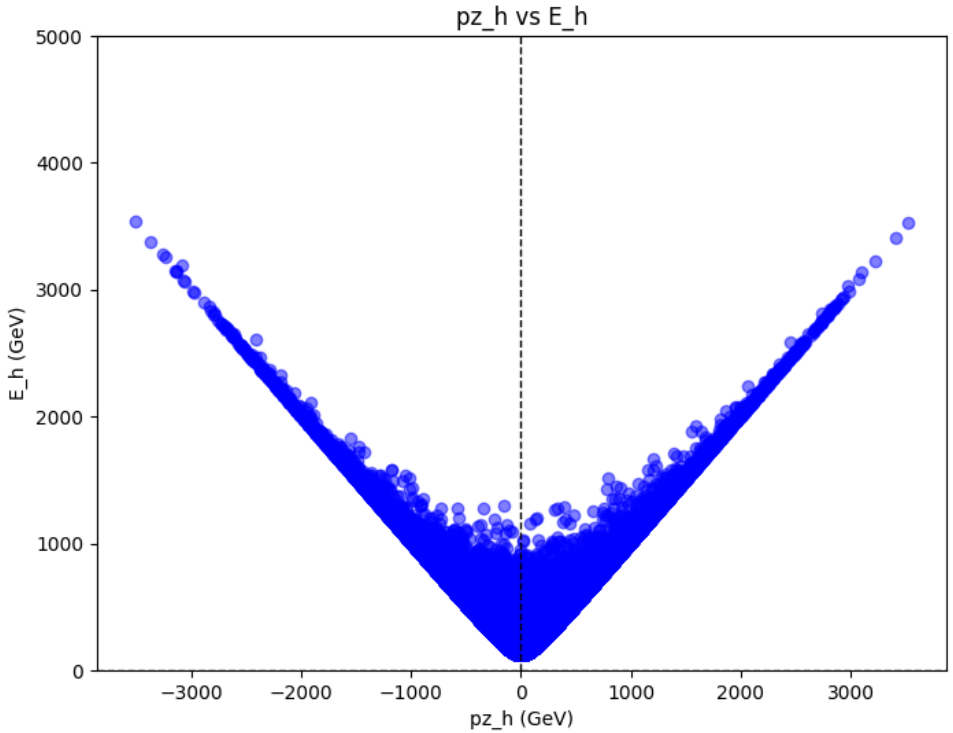

eta_h vs pz_t, correlation coefficient - 0.58
4000
2000
0
−2000
−4000
pz_t (GeV)
−6
−4
−2
0
2
4
6
Pseudorapidity

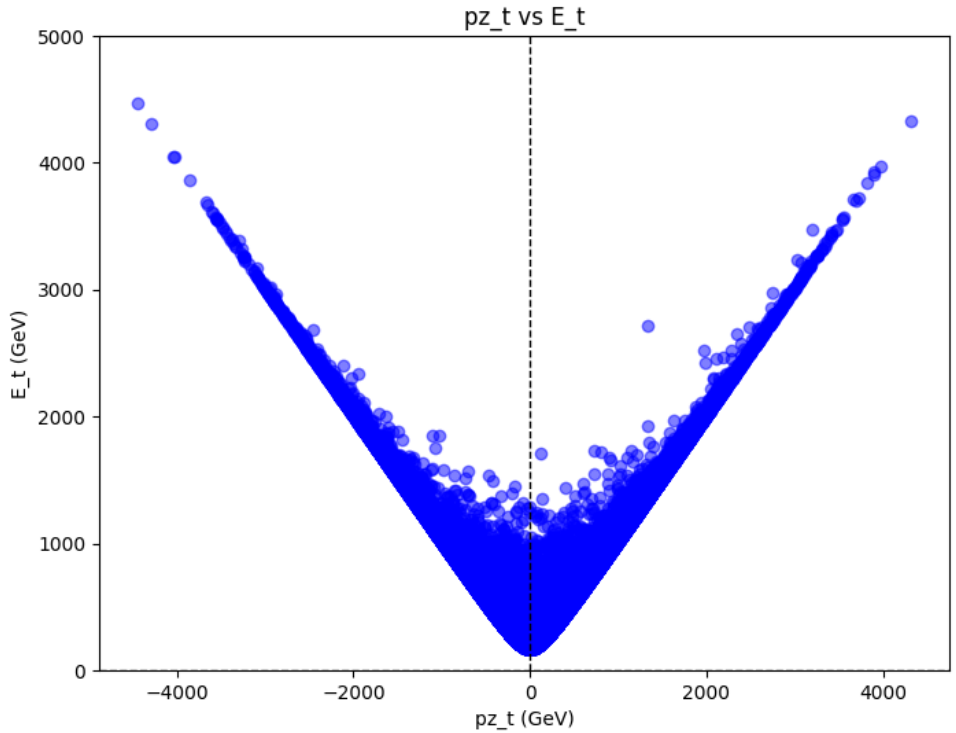

pT_t vs pT_t2, correlation coefficient - 0.56
1500
1250
1000
750
500
250
0
pT_t2 (GeV)
0
500
1000
1500
2000
pT_t (GeV)

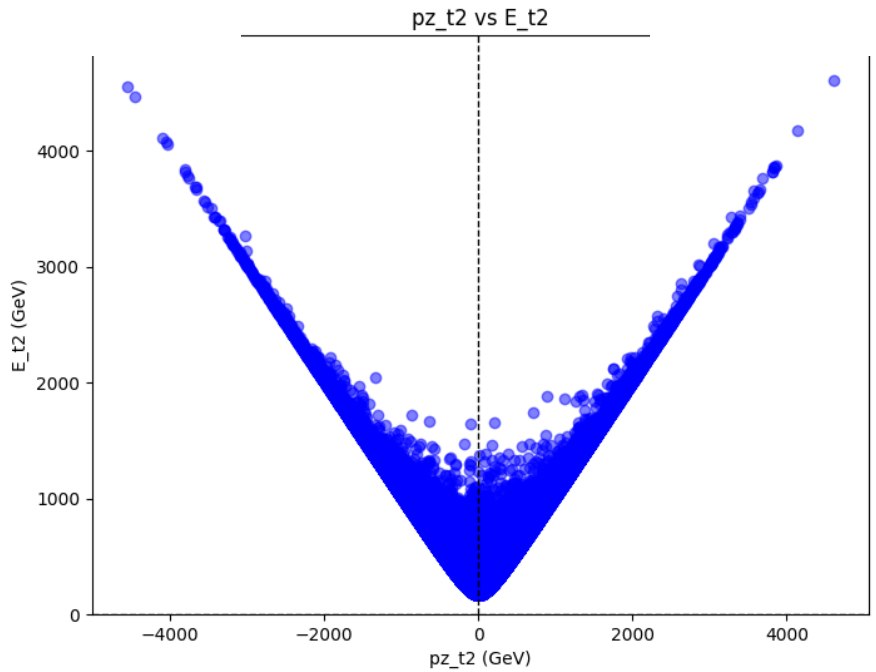

pT_h vs M_inv, correlation coefficient - 0.53
M_inv (GeV)
pT_h (GeV)
0
1000
2000
3000
4000
5000
0
200
400
600
800
1000
1200

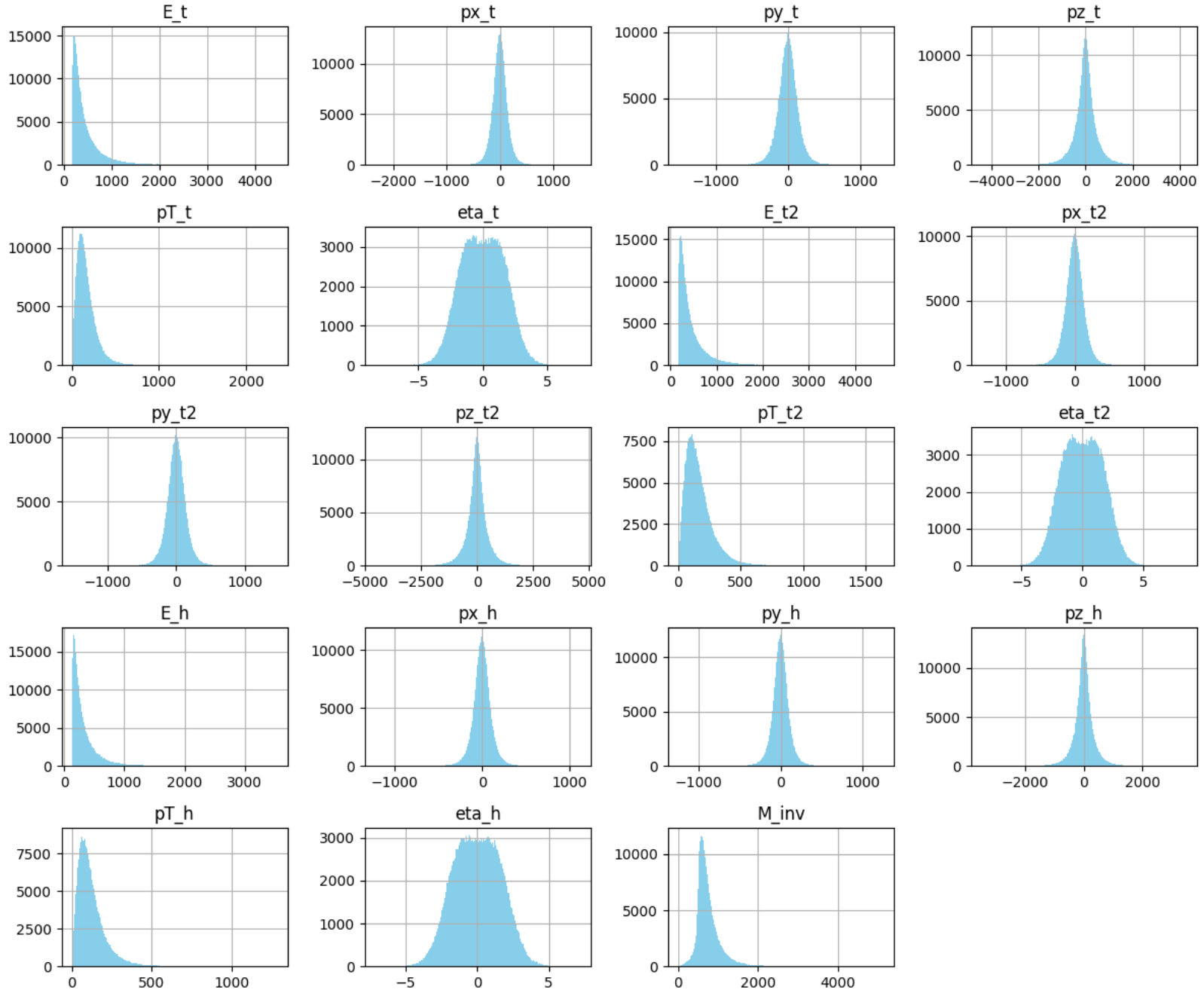

pT_t vs M_inv, correlation coefficient - 0.62
M_inv (GeV)
pT_t (GeV)
0
1000
2000
3000
4000
5000
0
500
1000
1500
2000


pT_t2 vs M_inv, correlation coefficient
M_inv (GeV)
pT_t2 (GeV)
0
1000
2000
3000
4000
5000
0
250
500
750
1000
1250
1500

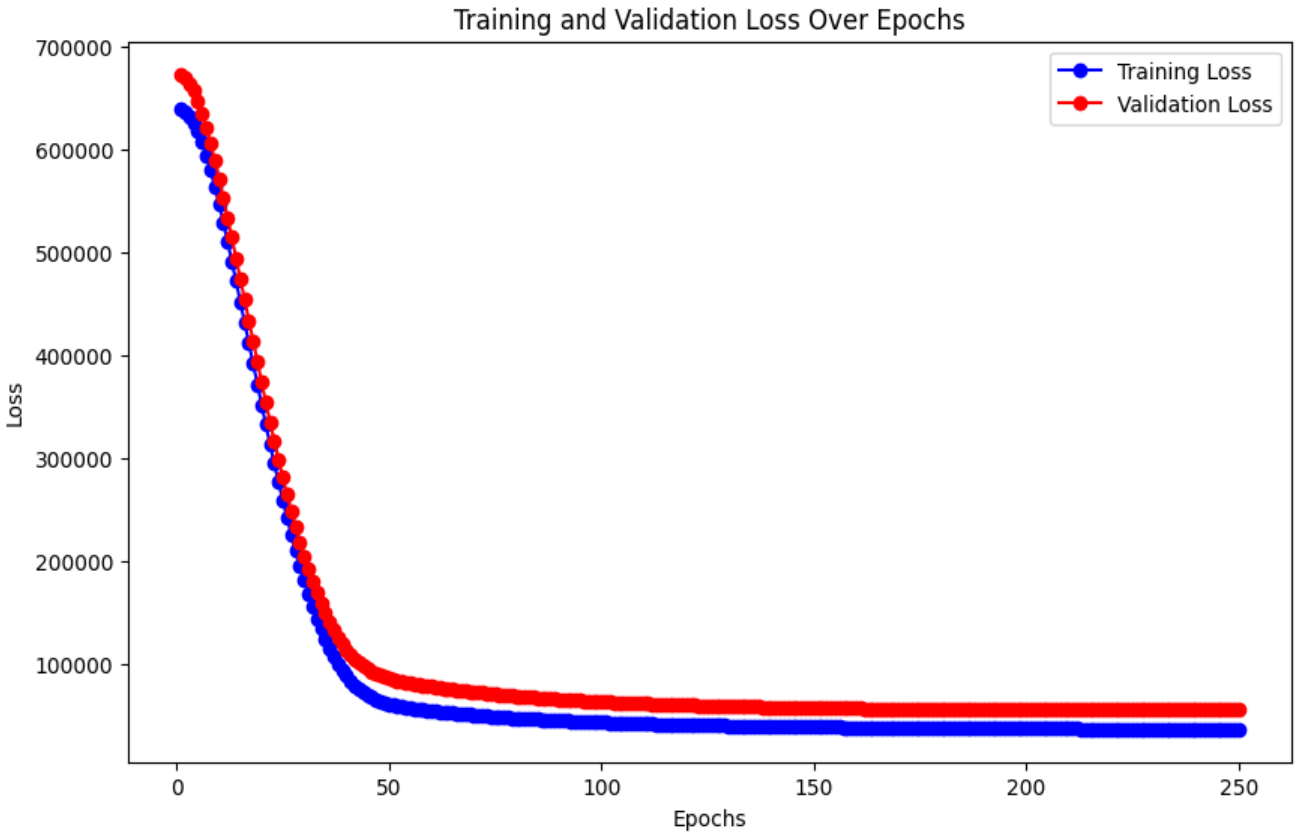

pz_h vs E_h
E_h (GeV)
pz_h (GeV)
5000
4000
3000
2000
1000
0
−3000
−2000
−1000
0
1000
2000
3000

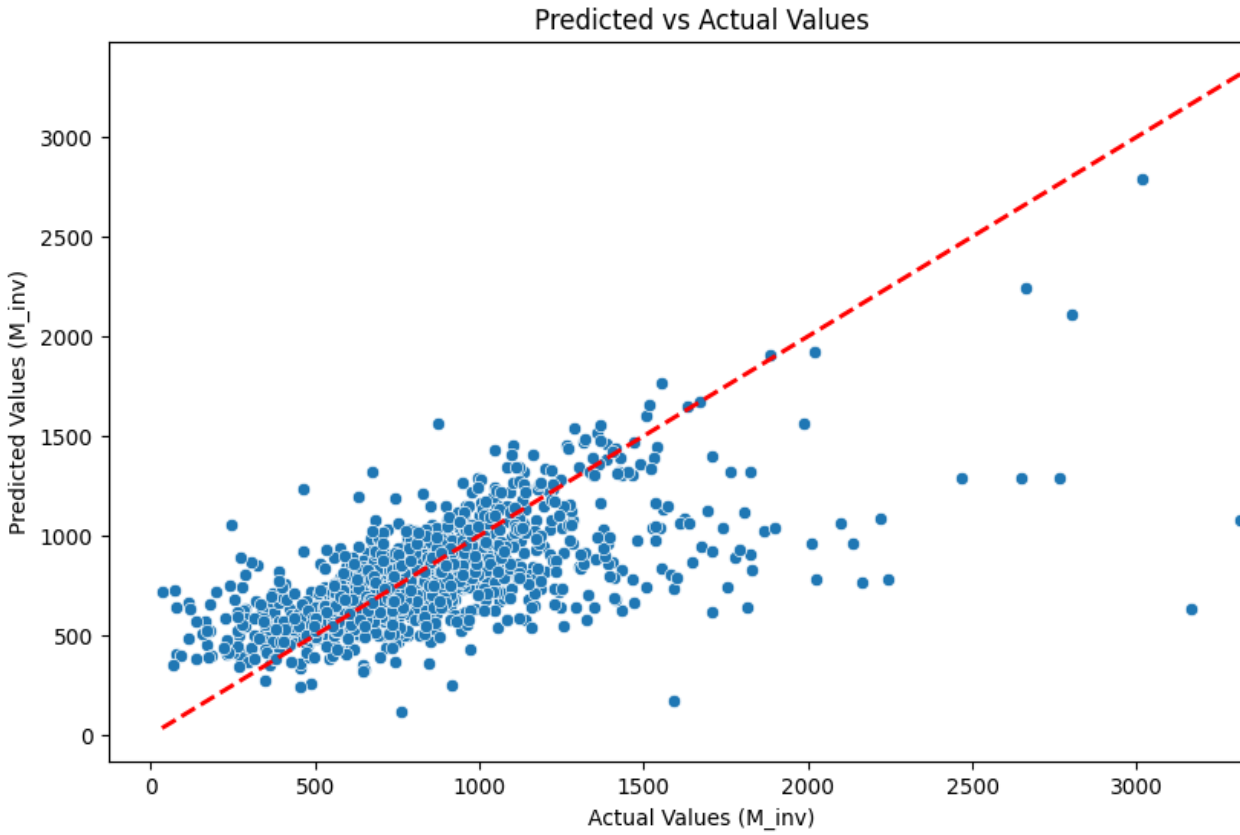

pz_t vs E_t
E_t (GeV)
pz_t (GeV)
5000
4000
3000
2000
1000
0
−4000
−2000
0
2000
4000


pz_t2 vs E_t2
E_t2 (GeV)
pz_t2 (GeV)
4000
3000
2000
1000
0
−4000
−2000
0
2000
4000

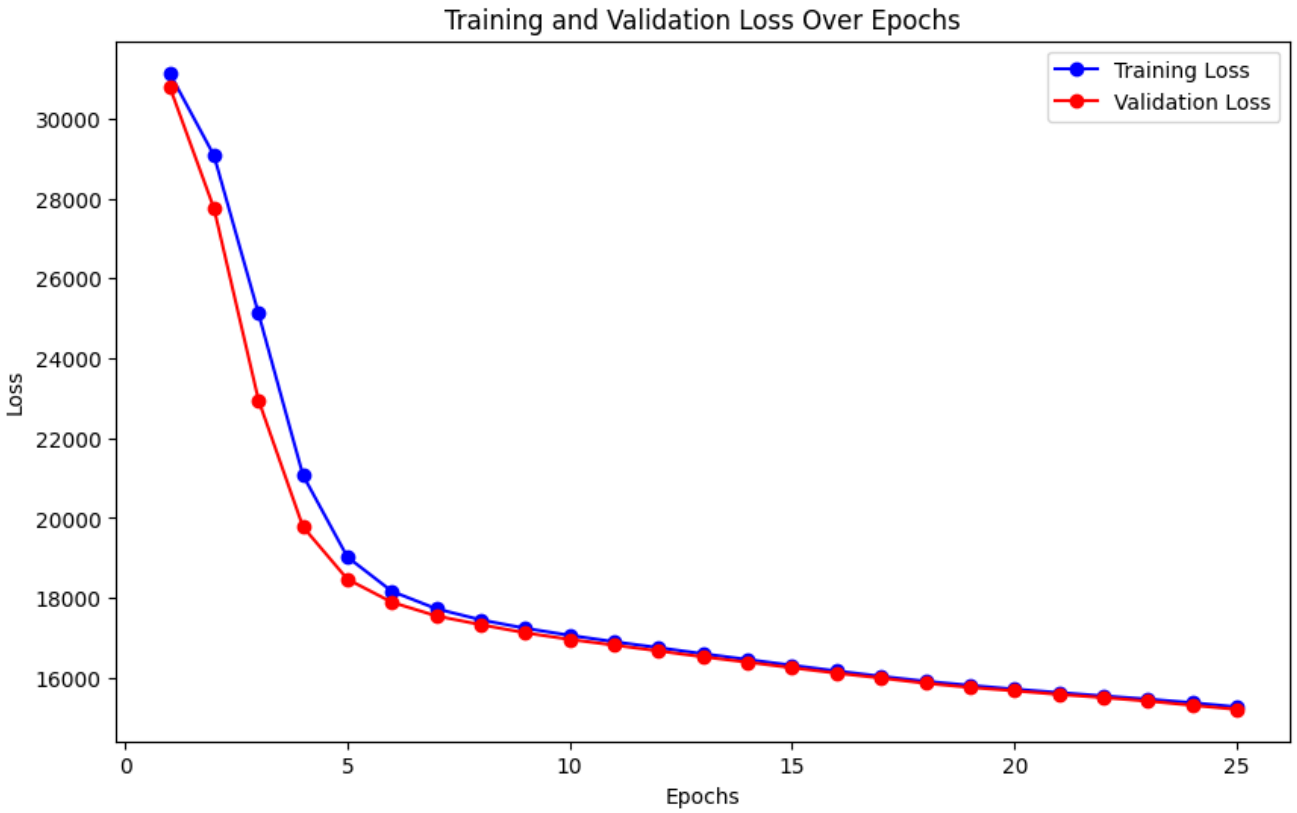

E_t
px_t
py_t
pz_t
pT_t
eta_t
E_t2
px_t2
py_t2
pz_t2
pT_t2
eta_t2
E_h
px_h
py_h
pz_h
pT_h
eta_h
M_inv

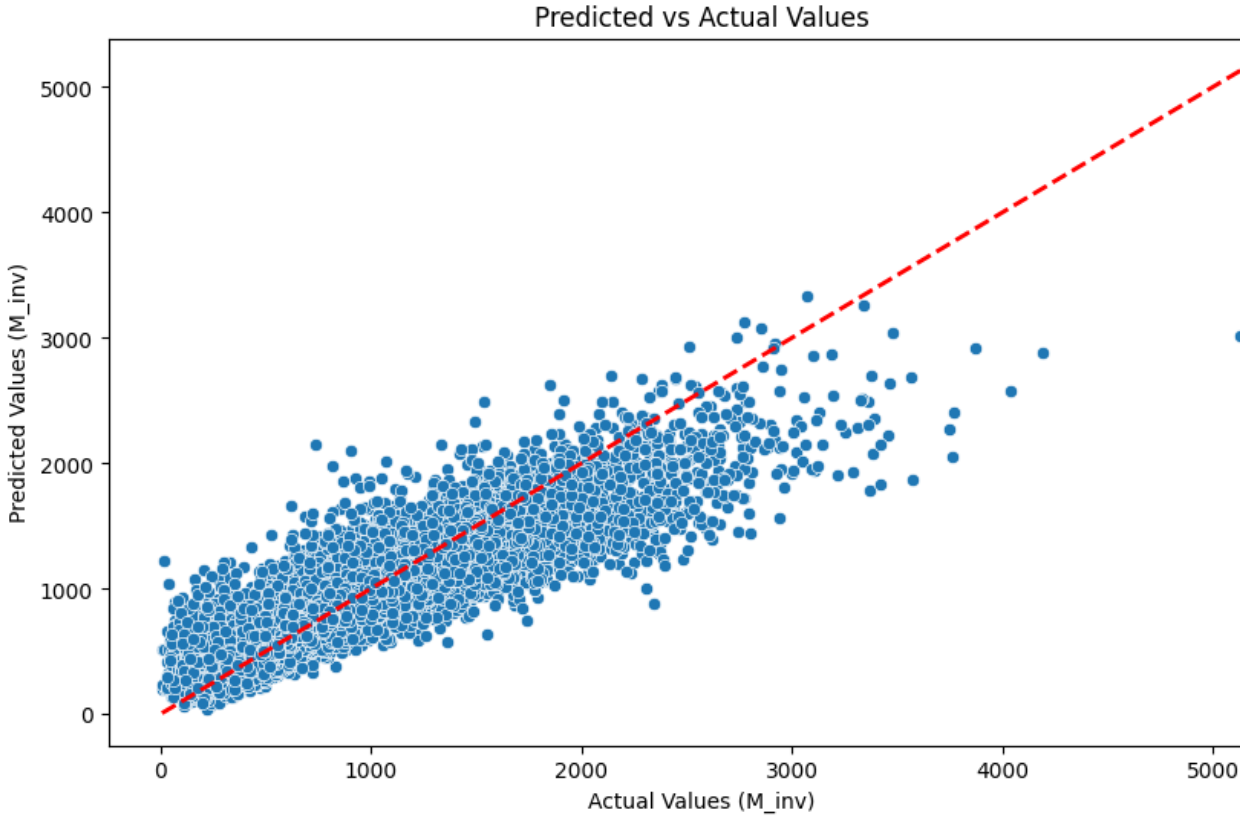
Training basic regression model to predict M_inv values
Starting with 1000 point sample over 250 epochs



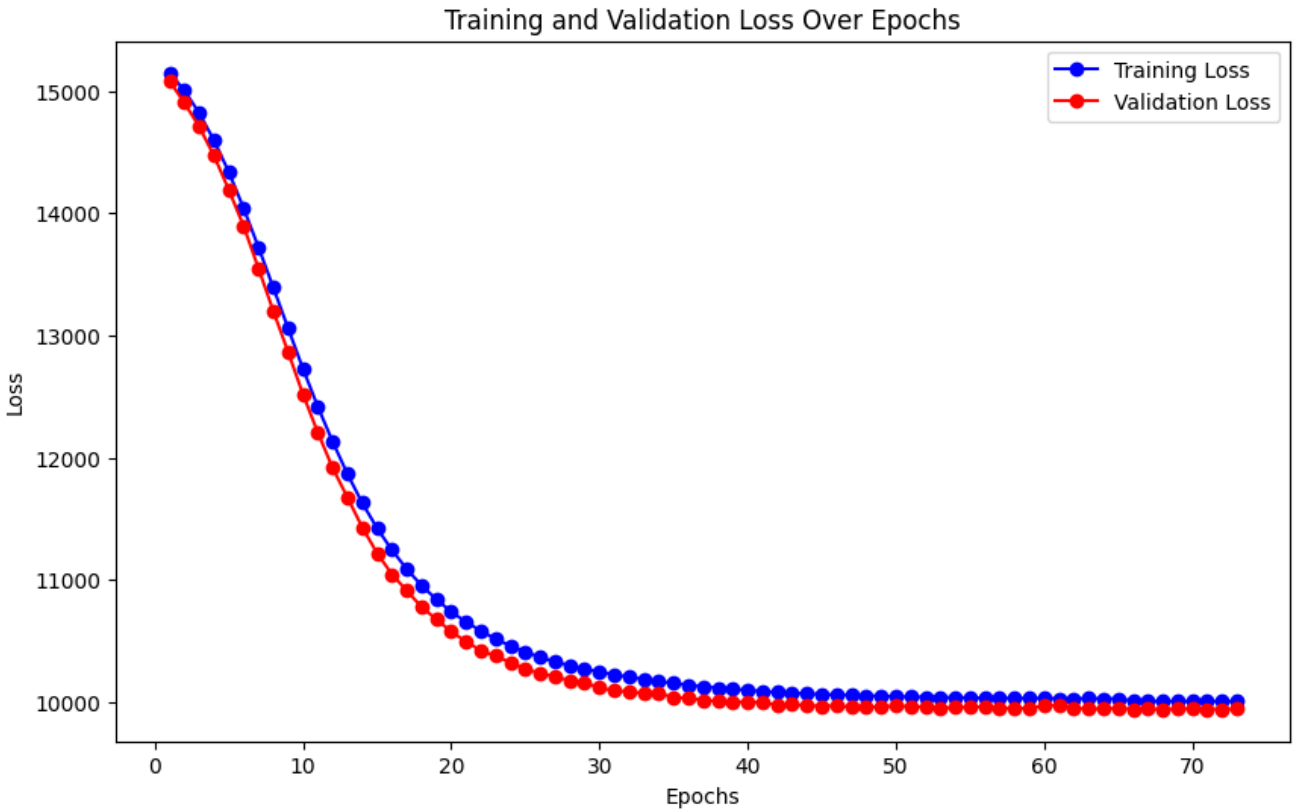


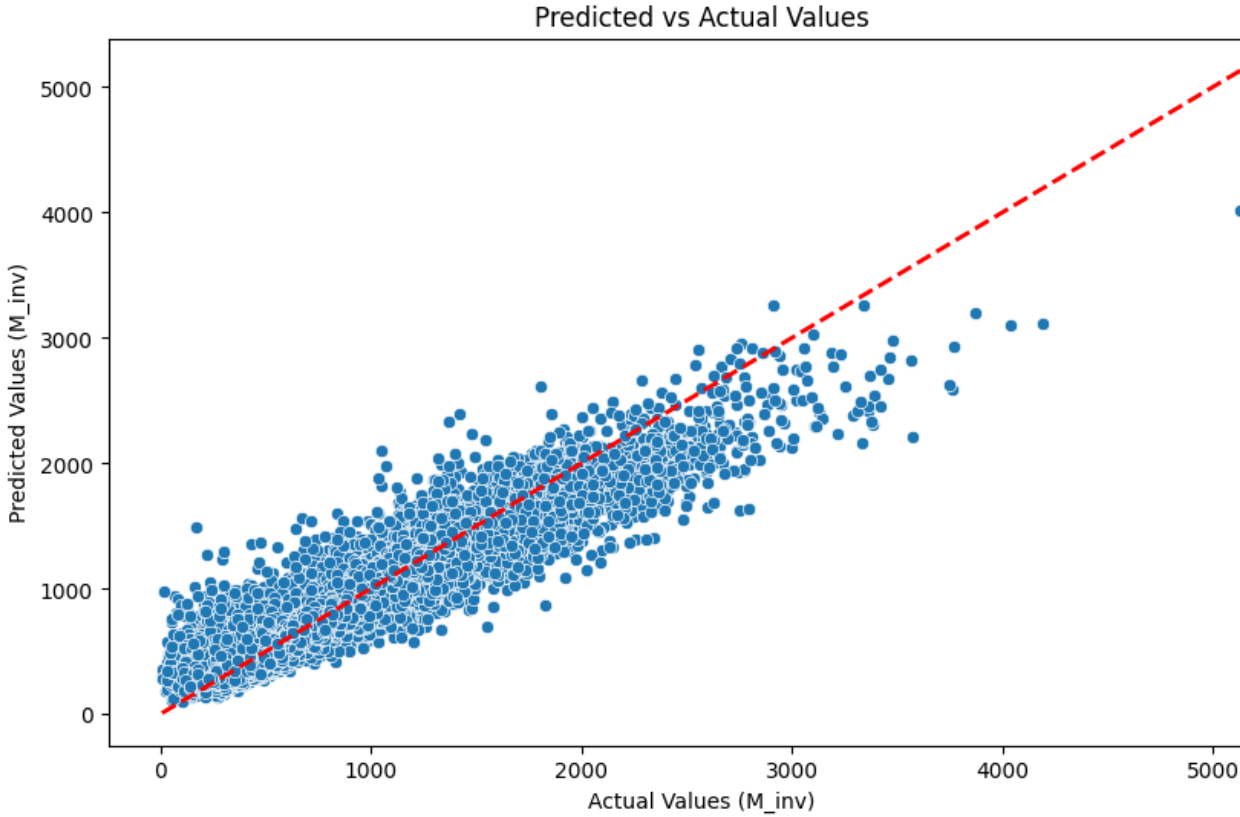
Using 500 000 points, 25 epochs





500 000 points, 100 epochs, early stopping and variable learning rate





Conclusions

- The data set was generation, parsing and all the calculations were successful
- Learned new methods of data visualization, using new libraries (such as seaborn)
- Found the correlation matrix to be a very useful tool
- Did not use pandas much before, but now I find it very useful and convenient
- The model training was completed fully, did not try out everything I could to make it better, but it was nice practice, with better understanding of what is going on, instead of tensorflow just being a “blackbox” that gives me results