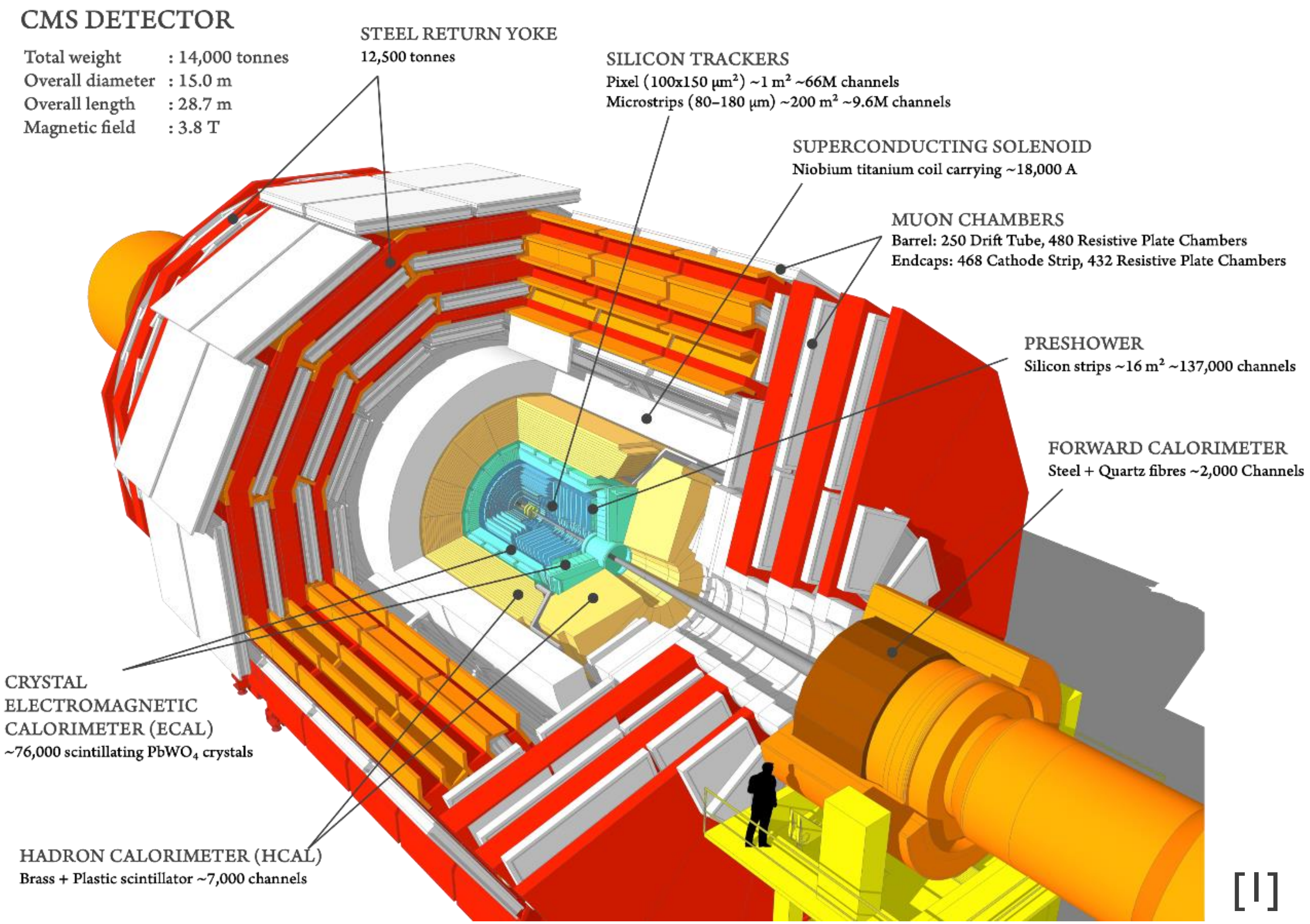
Study of ttH Kinematics with the CMS Experiment at the LHC

Student: **Kristijonas Mikas Silius**, Supervisor: Aurelijus Rinkevicius, Consultants: Darius Jurčiukonis, Nana Chychkalo

2024-06-21

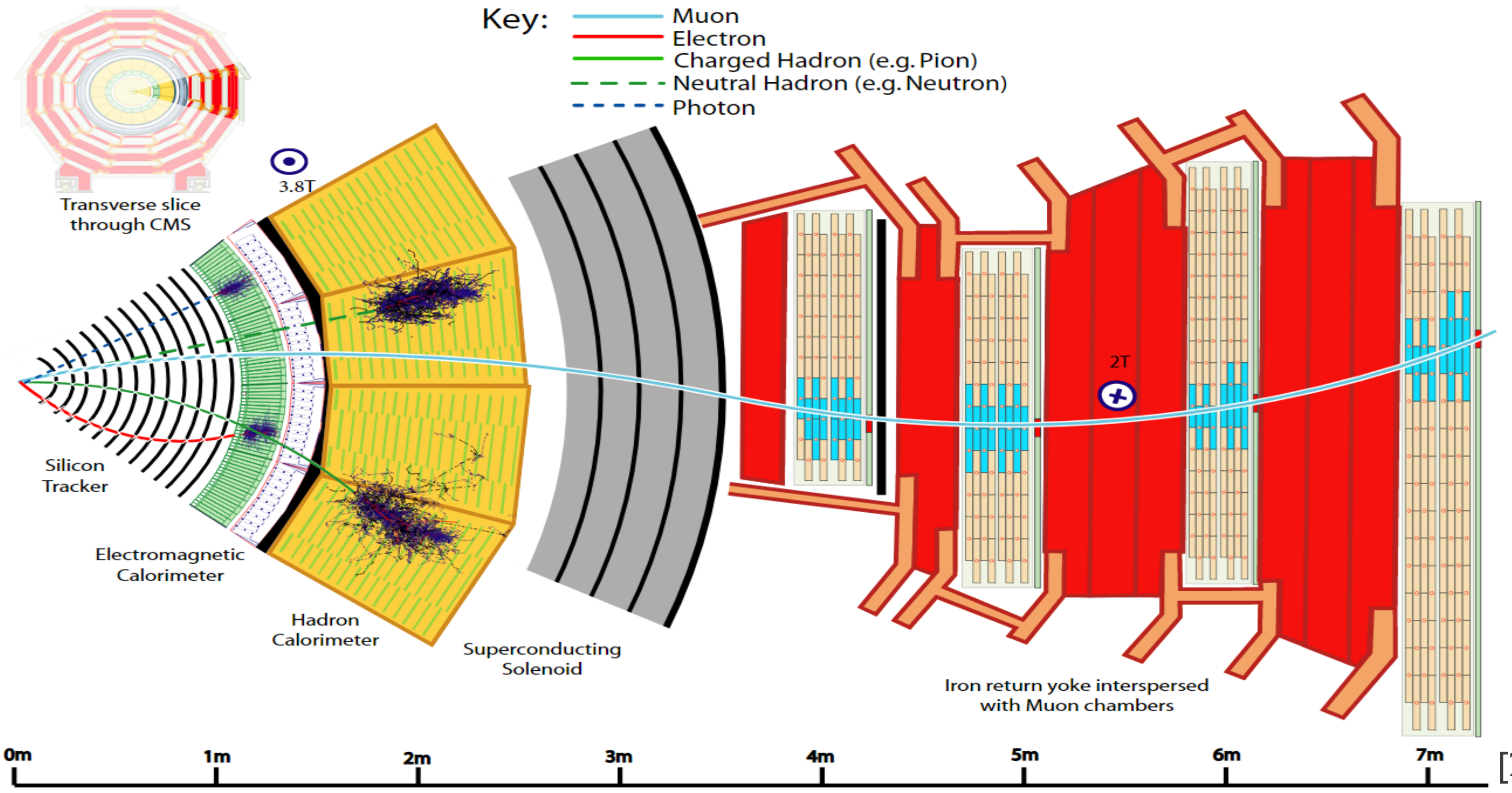
CMS EXPERIMENT

- CMS and ATLAS discover the Higgs boson in 2012.
- Particle collisions happen at the center.
- Consists of different layers to detect different particles.
- Shape of a cylinder.
- 14 000 tonnes.
- 100 m underground.



[1]

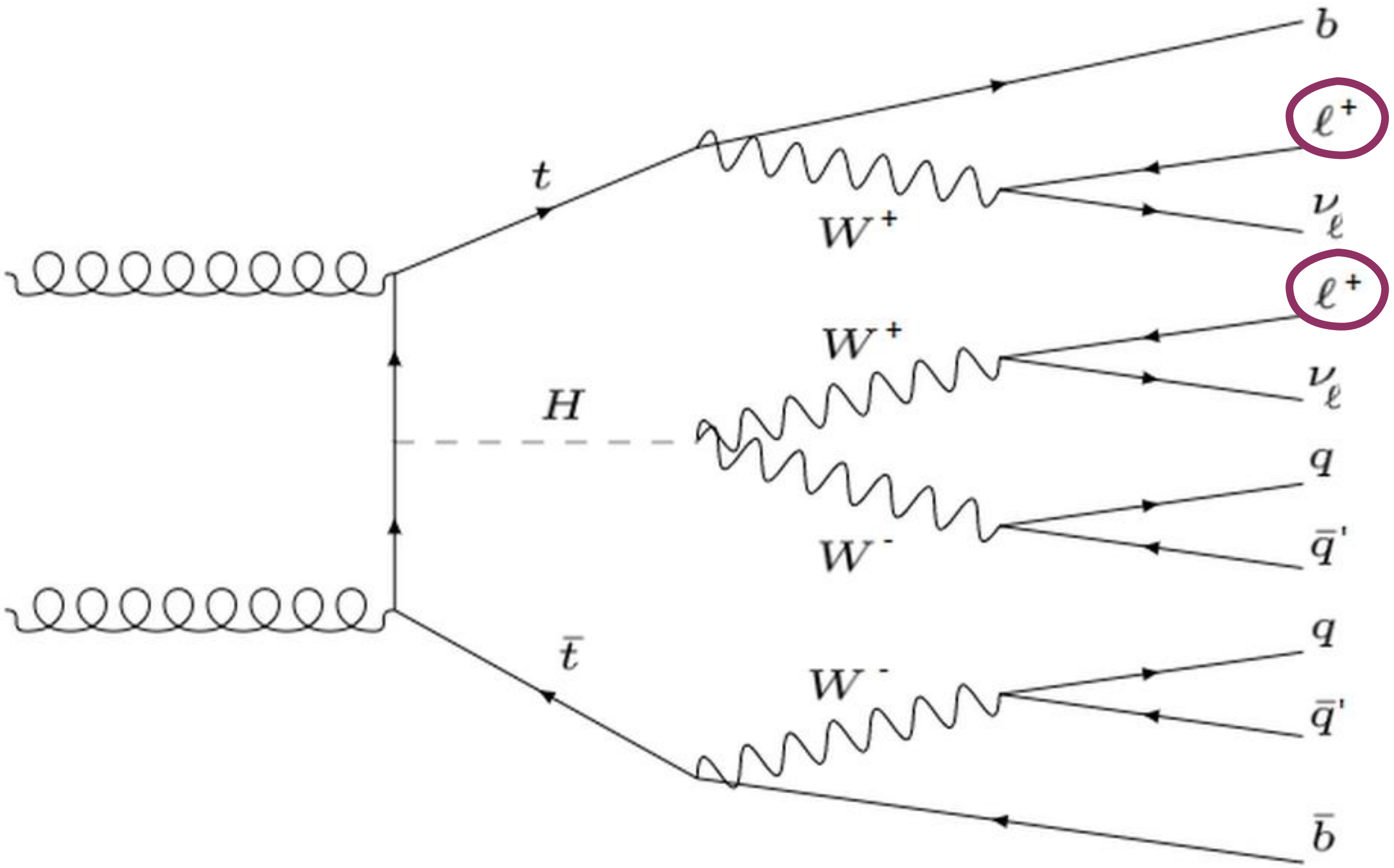
CMS SLICE


Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
Transverse slice through CMS
3.8T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0m
1m
2m
3m
4m
5m
6m
7m

[2]

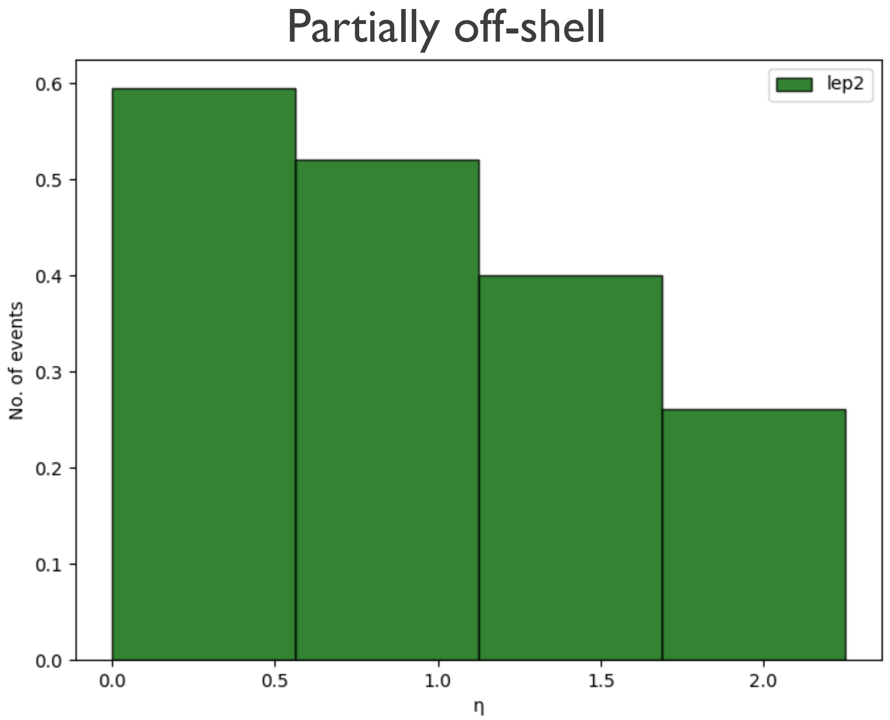
FINAL STATE ttH 2lss

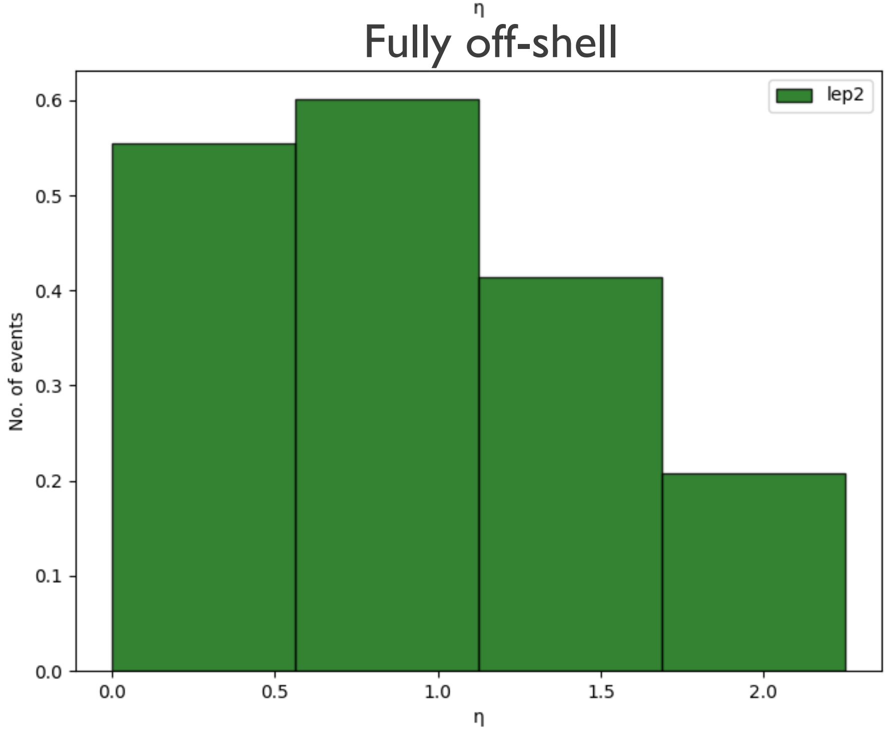
- The study used generated ttH 2lss final state simulation data. Example Feynmann diagram shown [3].
- $t, \bar{t}$ – tops, $W^{+/-}$ – W bosons, $q, \bar{q}$ – quarks, ν_l – leptonic neutrinos.
- The highlighted parts indicate the same sign leptons (positively charged), that are the products of the W^+ decays.
- This process is interesting because it allows to study the Yukawa coupling constant between Higgs and fermions.
- Any significant deviations from SM predictions might indicate new physics.



FULLY OFF-SHELL SIMULATION USED

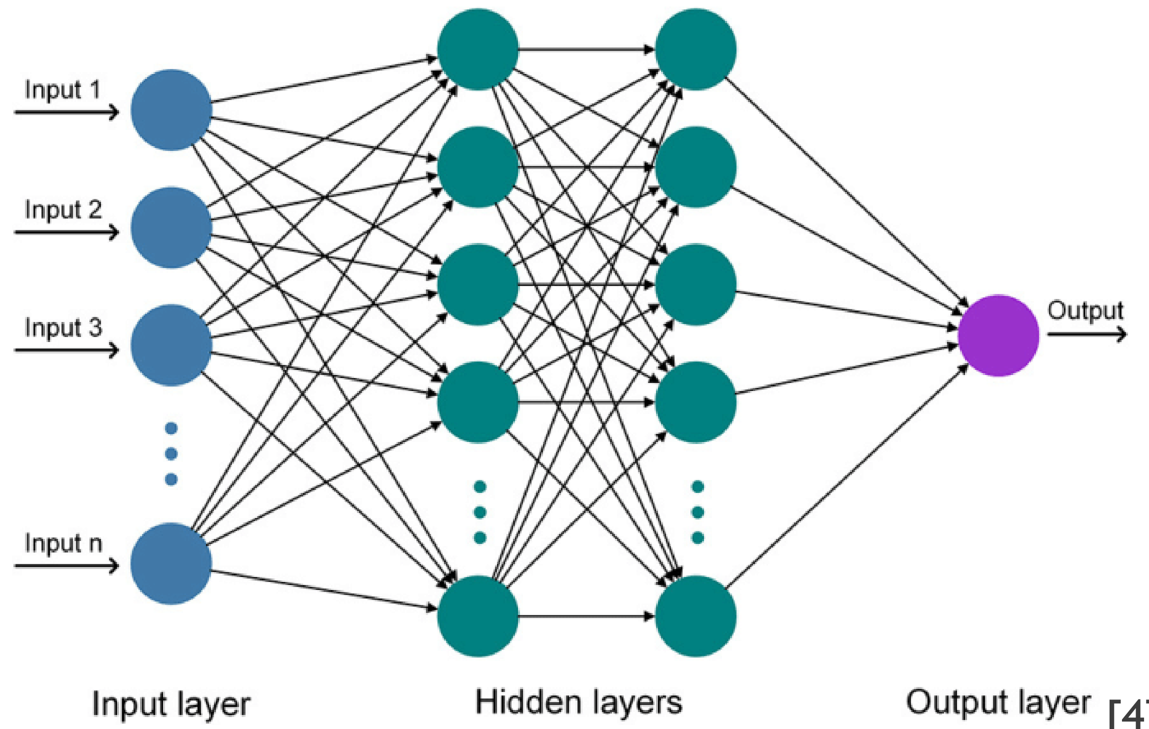
- Partially off-shell and a fully off-shell simulation of ttH 2lss final state were generated.
- Presented graphs show the distribution of the pseudorapidity (η) of the second lepton for partially and fully off-shell processes respectively.
- Fully off-shell approach better represents real particle collision conditions.
- Therefore, fully off-shell simulation was used for further studies.





REGRESSION: HIGGS pT

- **Regression:** Regression analysis is a statistical method used to model the relationship between a dependent variable and one or more independent variables.
- **Purpose:** It helps in predicting the value of the dependent variable based on the values of the independent variables.
- Two neural networks were tested:
  - Nominal deep model
  - Alternative minimal model



[4]

- Output node maps to Higgs pT

MODEL ARCHITECTURE

Deep network structure:

- 1st layer: 512 nodes, activation = ReLU;
- 2nd layer: 256 nodes, activation = SELU;
- 3rd layer: 128 nodes; activation = SELU;
- 4th layer: 64 nodes; activation = SELU;
- Final layer: 1 node, activation = ELU.

Deep network parameters:

- Training epochs = 100;
- Batch size = 128;
- Learning rate = 1e-4;
- doWeights = False;
- Custom loss function = MSEDeltaVar.

Minimal network structure:

- 1st hidden layer: 36 nodes, activations = SELU, SELU;
  - Partial increments of nodes from 33 to 36
- 2nd hidden layer: 8 nodes, activations = ReLU, SELU, ReLU, Linear;
  - Partial increments of nodes from 5 to 8
- Regression layer.

Minimal network parameters:

- Encoder epochs (hidden layer 1) = 100;
- Encoder epochs (hidden layer 2) = 100;
- Regressor epochs = 200;
- Learning rate (hidden layer 1) = 1e-3;
- Learning rate (hidden layer 2) = 1e-5;
- Learning rate (regressor) = 1e-4;
- Batch size = 264.

TRAINING INPUTS

- Example input:
- ['Hreco_Lep0_pt', 'Hreco_Lep0_eta', 'Hreco_Lep0_phi', 'Hreco_Lep1_pt', 'Hreco_Lep1_eta', 'Hreco_Lep1_phi', 'Hreco_Jet0_pt', 'Hreco_Jet0_eta', 'Hreco_Jet0_phi', 'Hreco_Jet1_pt', 'Hreco_Jet1_eta', 'Hreco_Jet1_phi', 'Hreco_Jet2_pt', 'Hreco_Jet2_eta', 'Hreco_Jet2_phi', 'Hreco_Jet3_pt', 'Hreco_Jet3_eta', 'Hreco_Jet3_phi', 'Hreco_Jet4_pt', 'Hreco_Jet4_eta', 'Hreco_Jet4_phi', 'Hreco_Jet5_pt', 'Hreco_Jet5_eta', 'Hreco_Jet5_phi', 'Hreco_HadTop_pt', 'Hreco_HadTop_eta', 'Hreco_HadTop_phi', 'Hreco_met', 'Hreco_met_phi']
- Green – base inputs used everywhere, ordered leptons and hadTop
- Yellow – individual Jets
- Light Blue – MET
- Some cases include individual neutrino information, in grey:
- ['Hreco_met1_pT', 'Hreco_met2_pT', 'Hreco_met1_eta', 'Hreco_met2_eta', 'Hreco_met1_phi', 'Hreco_met2_phi', 'Hreco_met_phi']

TRAINING INPUTS

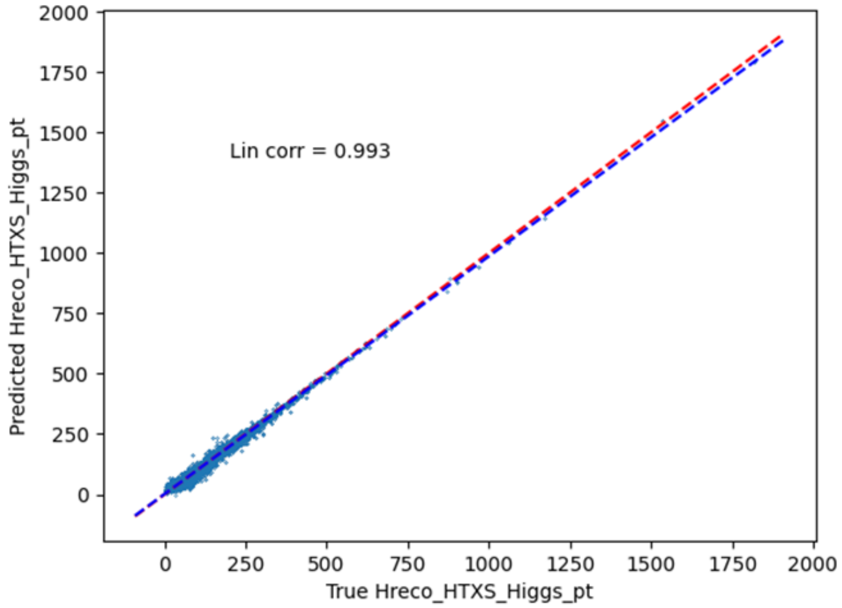
Matching training variables to ones used by ttH analysis group

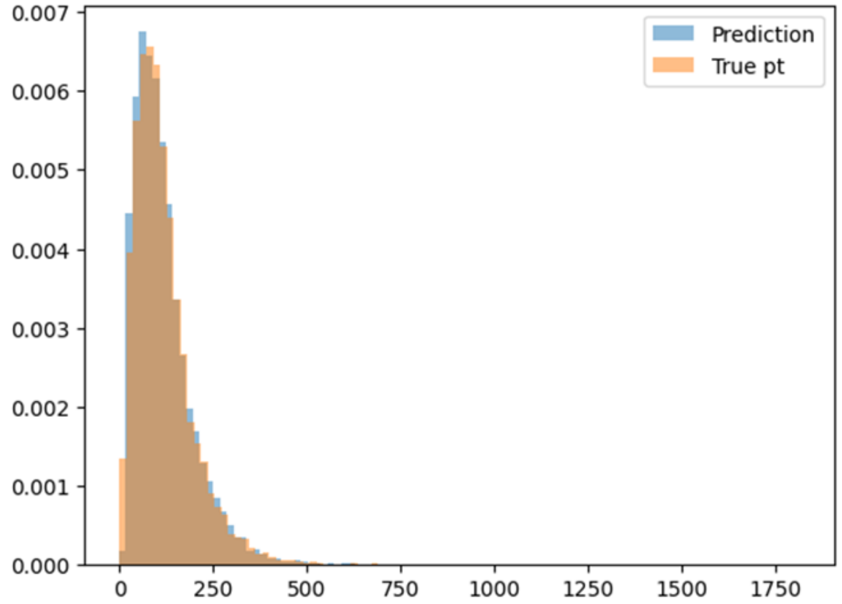
Current:

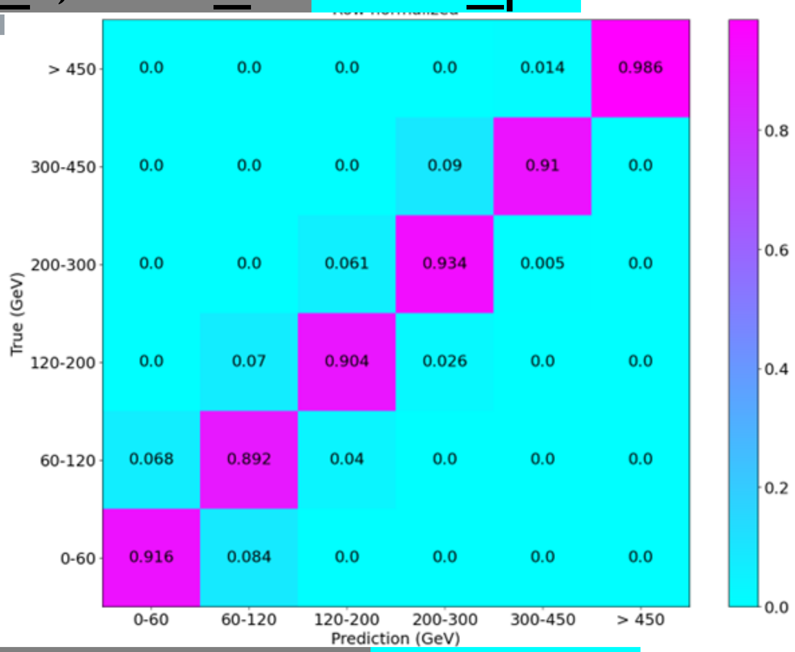
- ['Hreco_Lep0_pt', 'Hreco_Lep0_eta', 'Hreco_Lep0_phi', 'Hreco_Lep1_pt', 'Hreco_Lep1_eta', 'Hreco_Lep1_phi', 'Hreco_HadTop_pt', 'Hreco_HadTop_eta', 'Hreco_HadTop_phi', 'Hreco_met', 'Hreco_Jets_plus_Lep_pt', 'Hreco_Jets_plus_Lep_eta', 'Hreco_Jets_plus_Lep_phi', 'Hreco_More5_Jets_pt', 'Hreco_More5_Jets_eta', 'Hreco_More5_Jets_phi', 'Hreco_All5_Jets_pt', 'Hreco_All5_Jets_eta', 'Hreco_All5_Jets_phi', 'Hreco_met_phi']
- Used in multilepton analysis:
- ['Hreco_Lep0_pt', 'Hreco_Lep0_eta', 'Hreco_Lep0_phi', 'Hreco_Lep1_pt', 'Hreco_Lep1_eta', 'Hreco_Lep1_phi', 'Hreco_HadTop_pt', 'Hreco_HadTop_eta', 'Hreco_HadTop_phi', 'Hreco_TopScore', 'Hreco_met', 'Hreco_Jets_plus_Lep_pt', 'Hreco_Jets_plus_Lep_eta', 'Hreco_Jets_plus_Lep_phi', 'Hreco_More5_Jets_pt', 'Hreco_More5_Jets_eta', 'Hreco_More5_Jets_phi', 'Hreco_All5_Jets_pt', 'Hreco_All5_Jets_eta', 'Hreco_All5_Jets_phi', 'Hreco_met_phi']
- Purple – grouped jets or jets + lep
- TopScore – did not replicate during current training

Comparing best deep model and best minimal model results

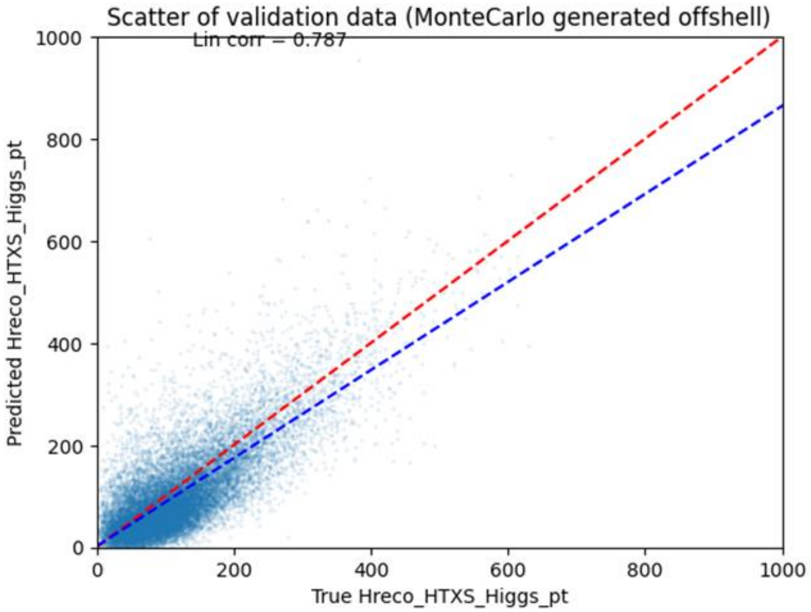
Deep model 10k events: Lep_*, HadTop_*, Jet0_*, Jet1_*, …, met1_*, met2_ * + met_phi

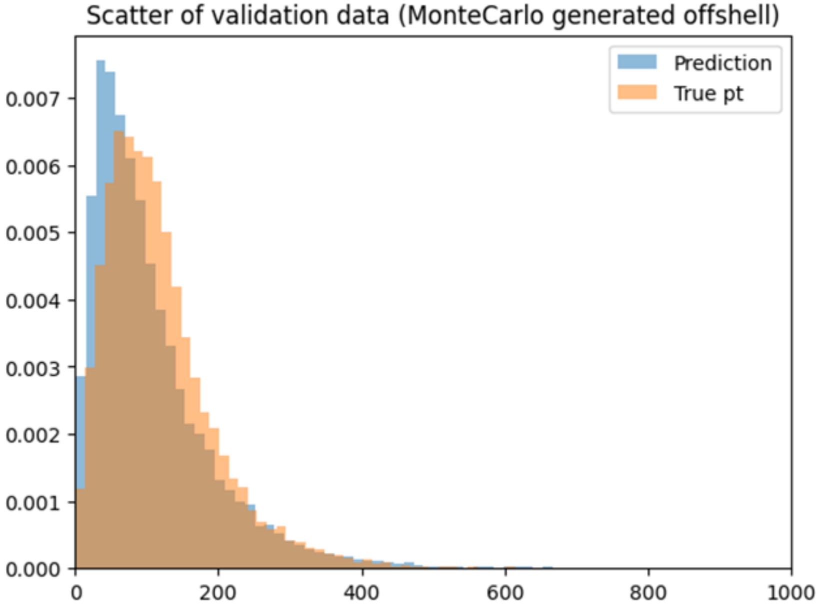


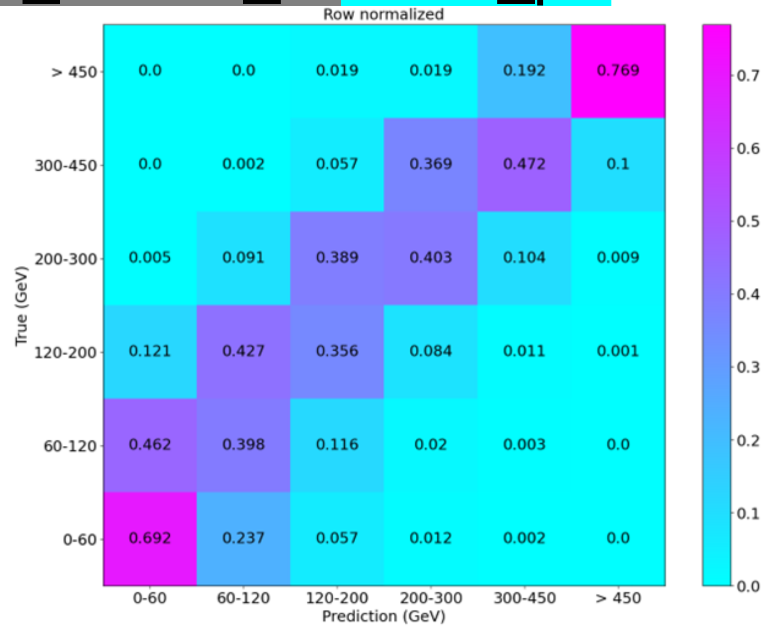




Minimal model 65k events: Lep_*, HadTop_*, Jet0_*, Jet1_*, …, met1_*, met2_ * + met_phi

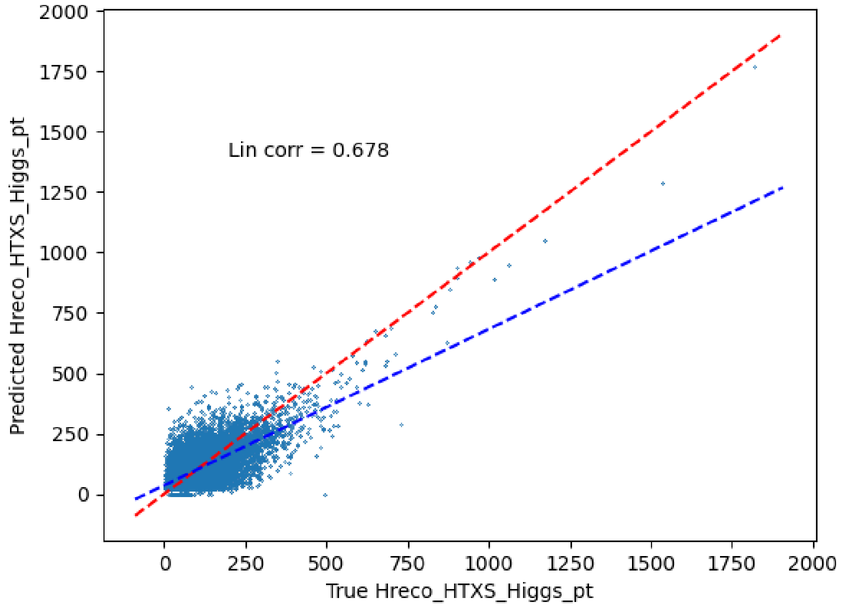


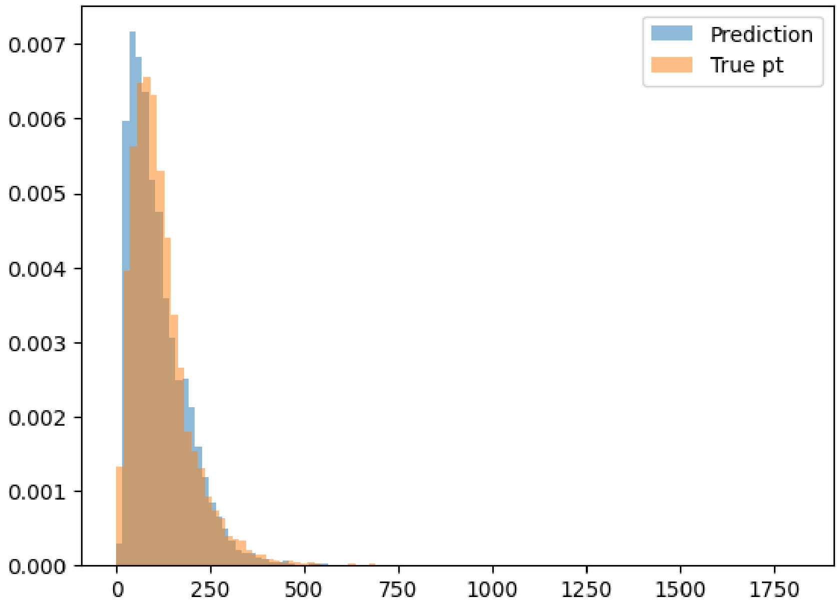


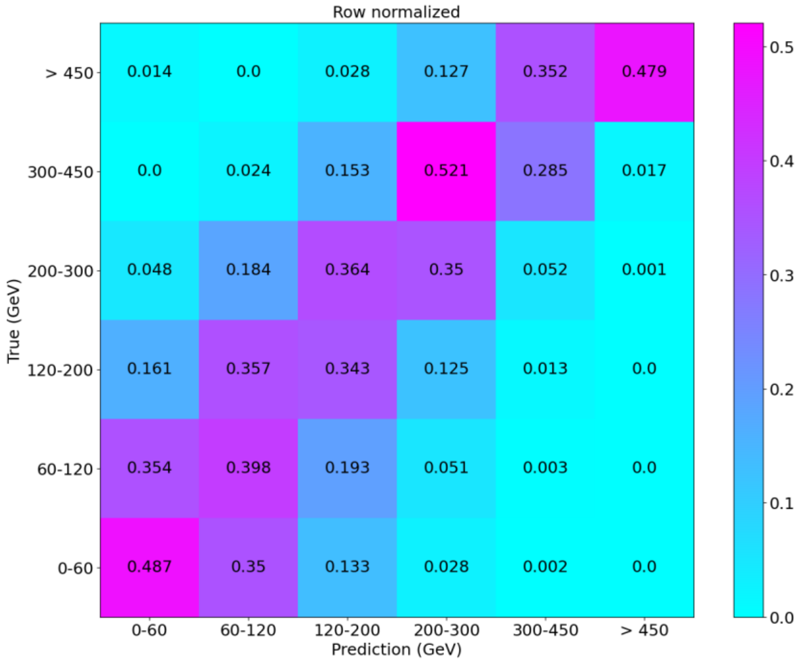


Deep model input variable testing

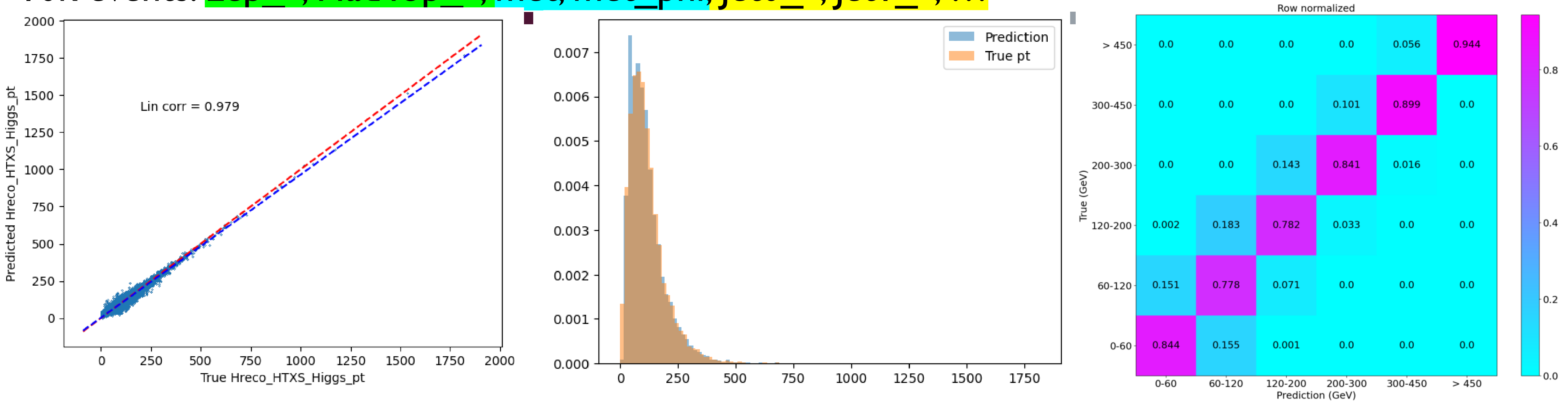
10k events: Lep_*, HadTop_*, met, met_phi



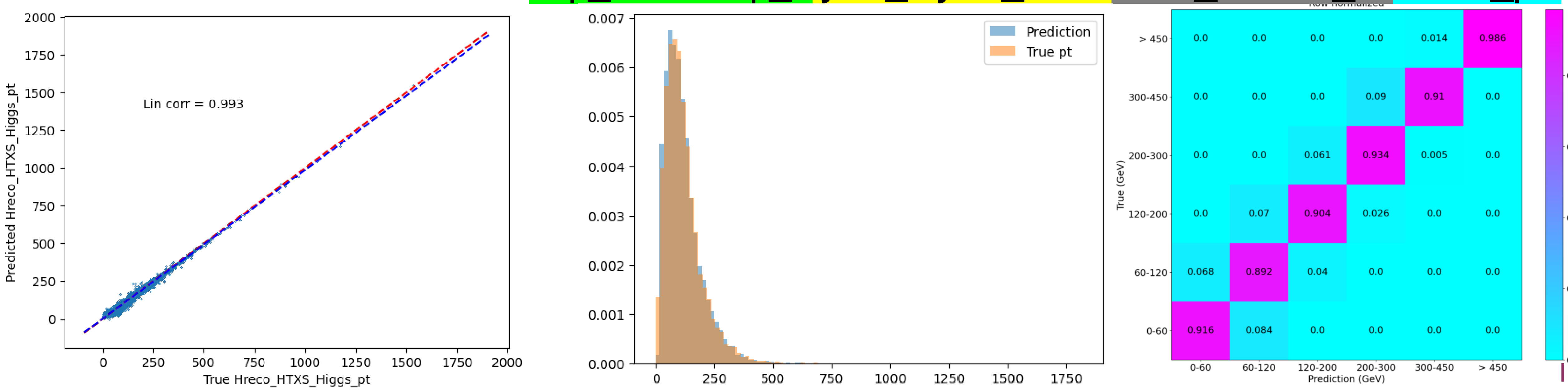




10k events: Lep_*, HadTop_*, met, met_phi, Jet0_*, Jet1_*, …



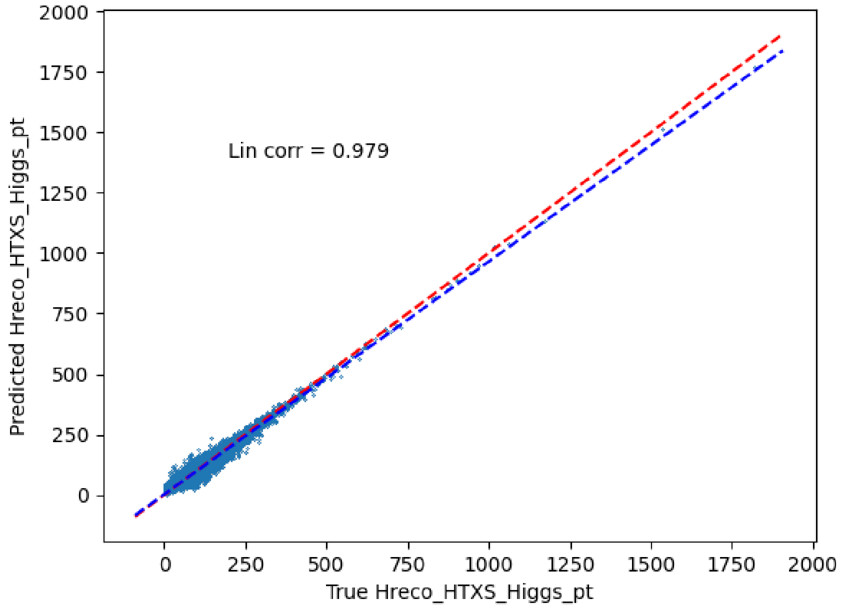
10k events, individual neutrinos: Lep_*, HadTop_*, Jet0_*, Jet1_*, …, met1_*, met2_* + met_phi

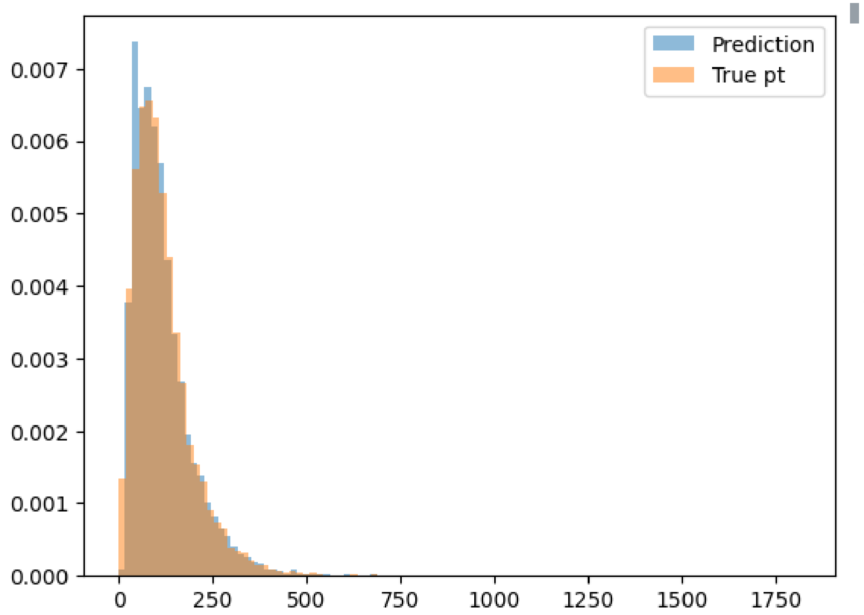


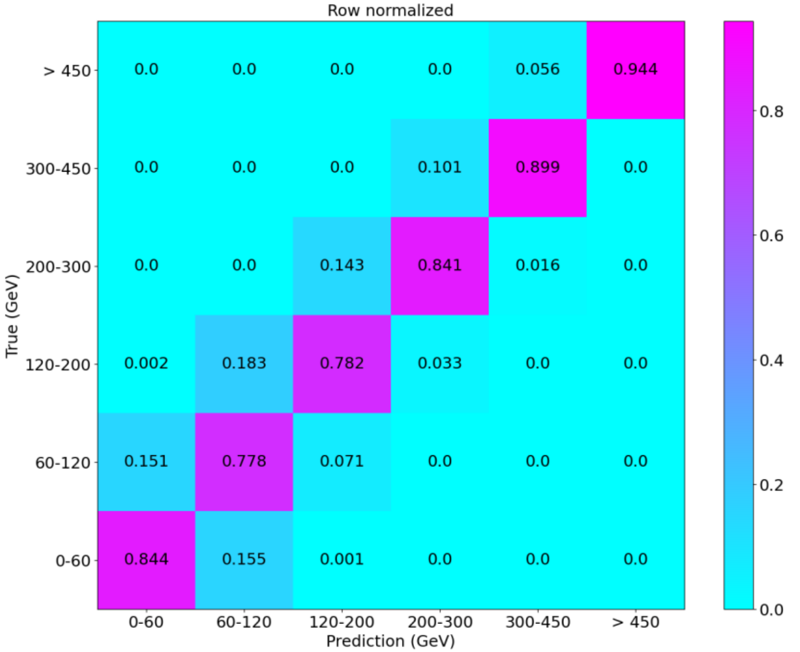
IMPORTANCE OF COMBINED met_phi

- Removing combined met_phi made loss function diverge (NaN)
- Combined met_phi was kept in studies
- Applies when **both** neutrinos kinematic data is used

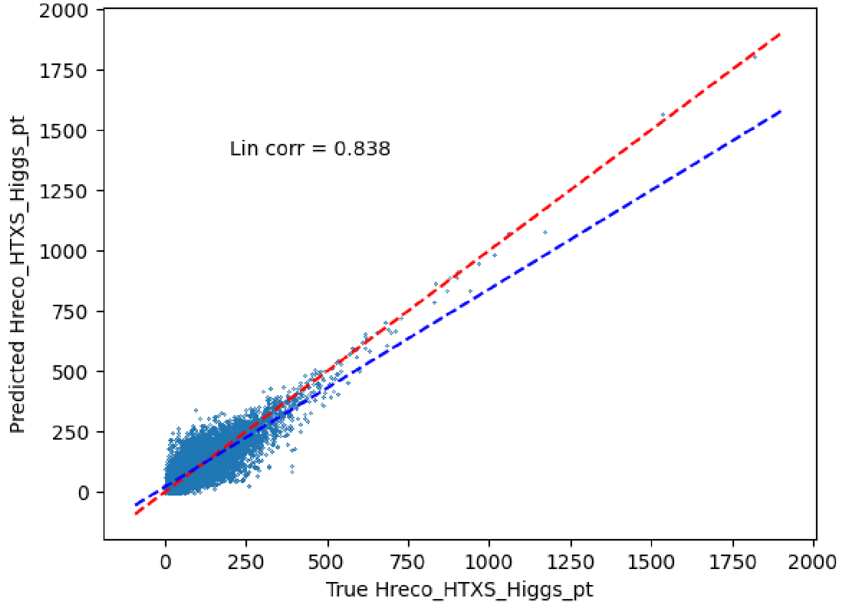
10k events: Lep_*, HadTop_*, met, met_phi, Jet0_*, Jet1_*, …

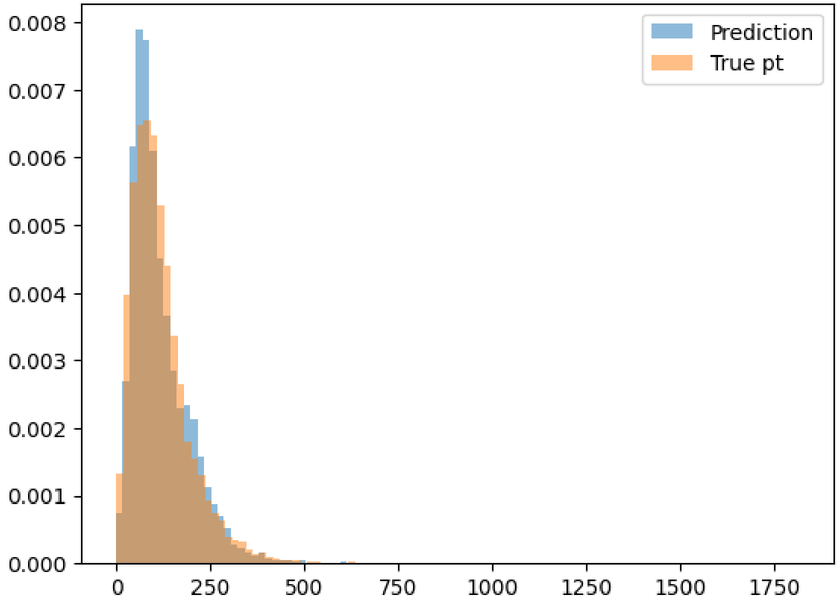


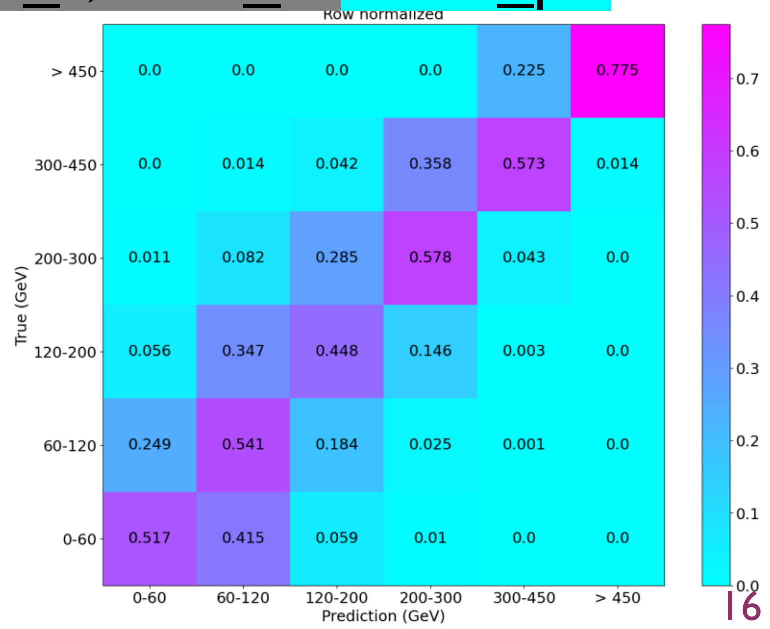




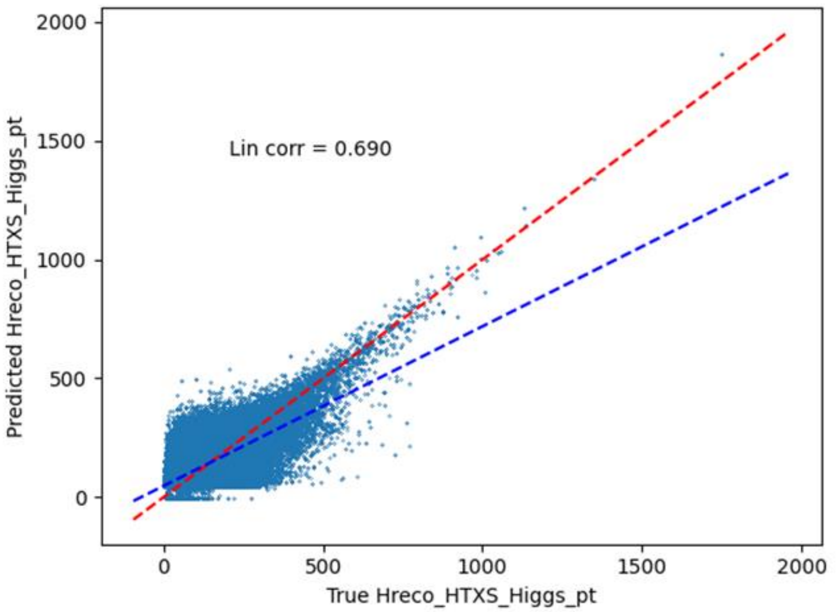
10k events, individual neutrinos without Jets: Lep_*, HadTop_*, met1_*, met2_* + met_phi

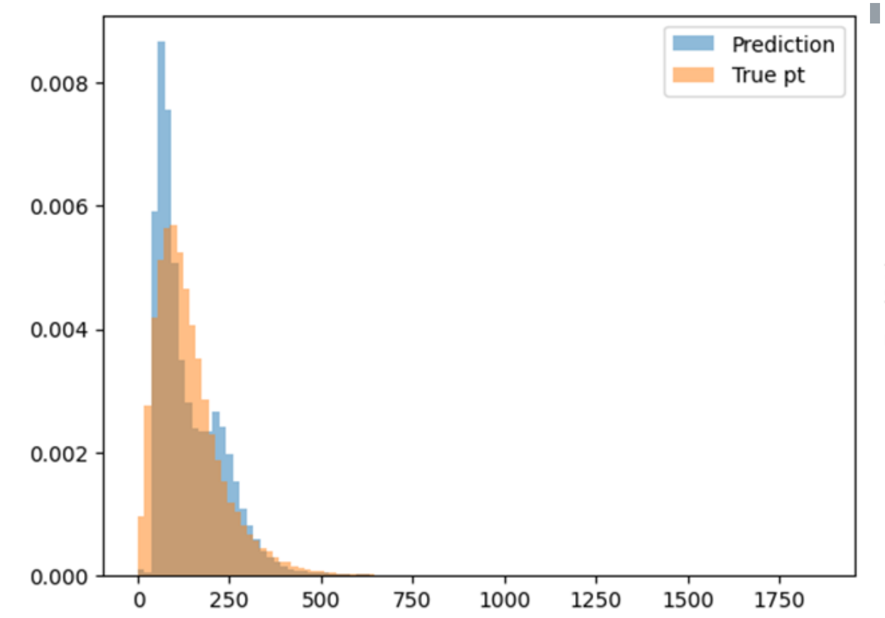


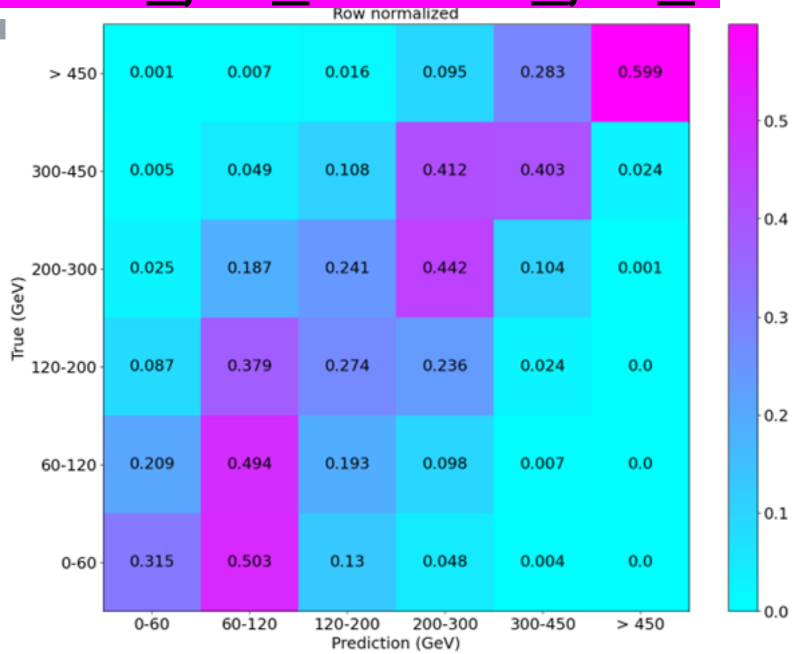




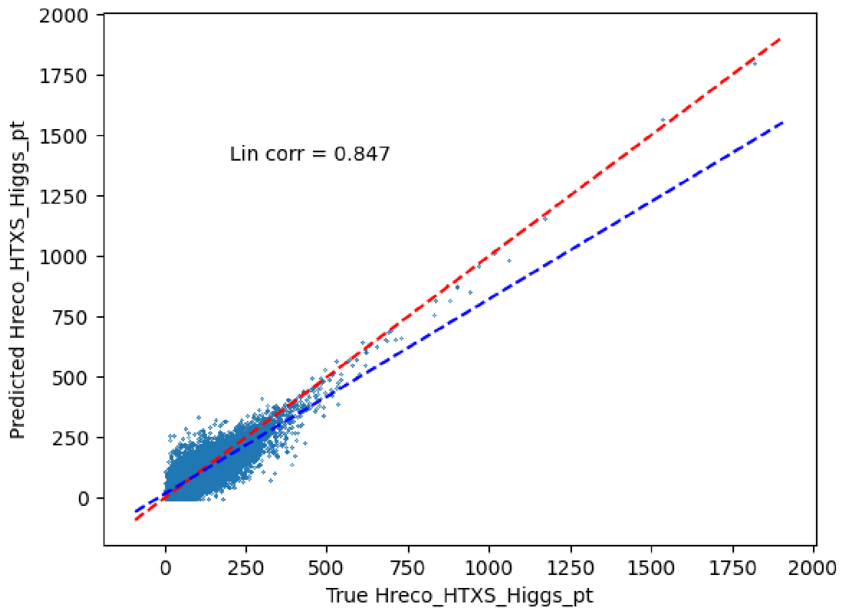
Full simulation: Lep_*, HadTop_*, met, met_phi, Jets_plus_lep_*, All5_jets_*, More5_jets_*

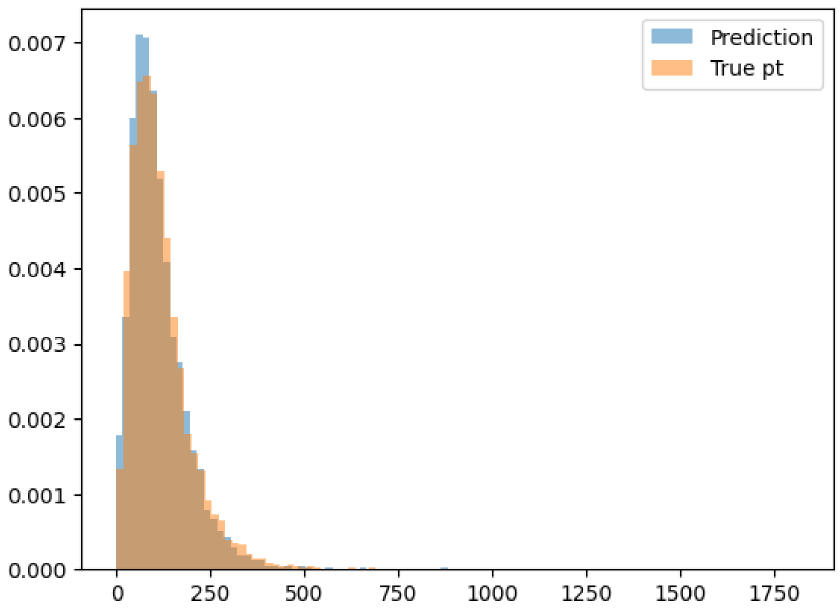


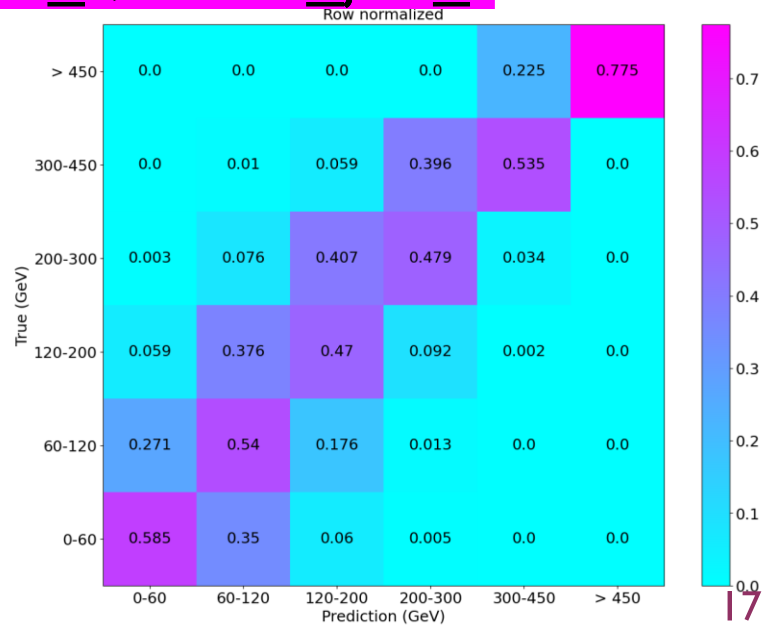




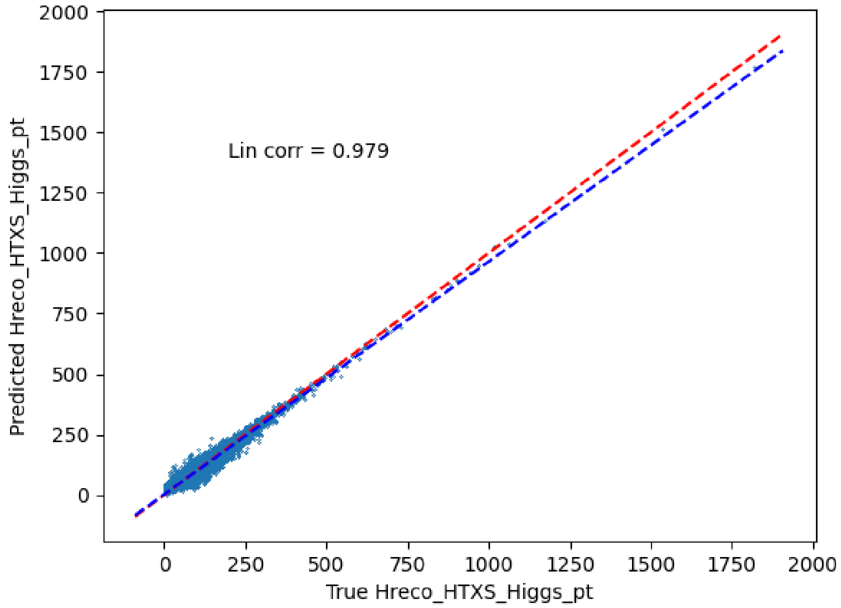
10k events: Lep_*, HadTop_*, met, met_phi, Jets_plus_lep_*, All5_jets_*, More5_jets_*

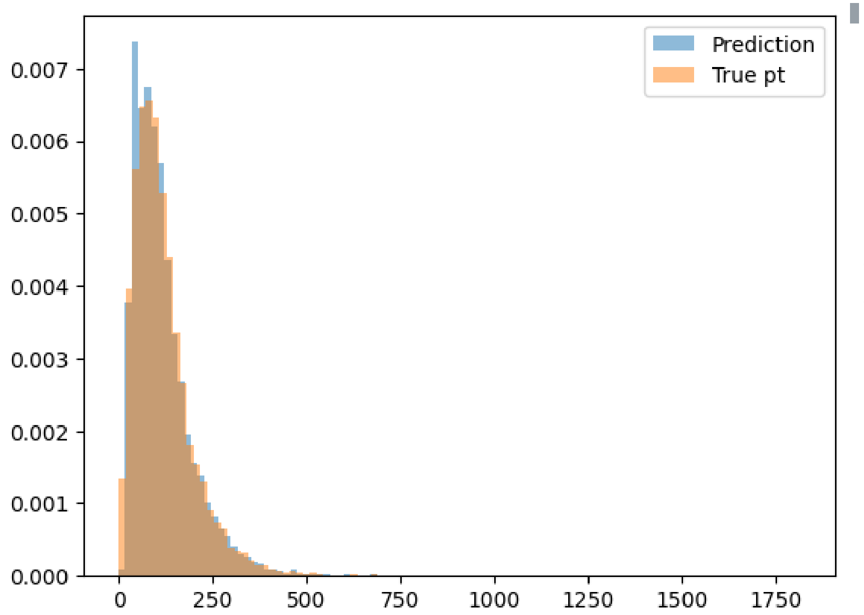


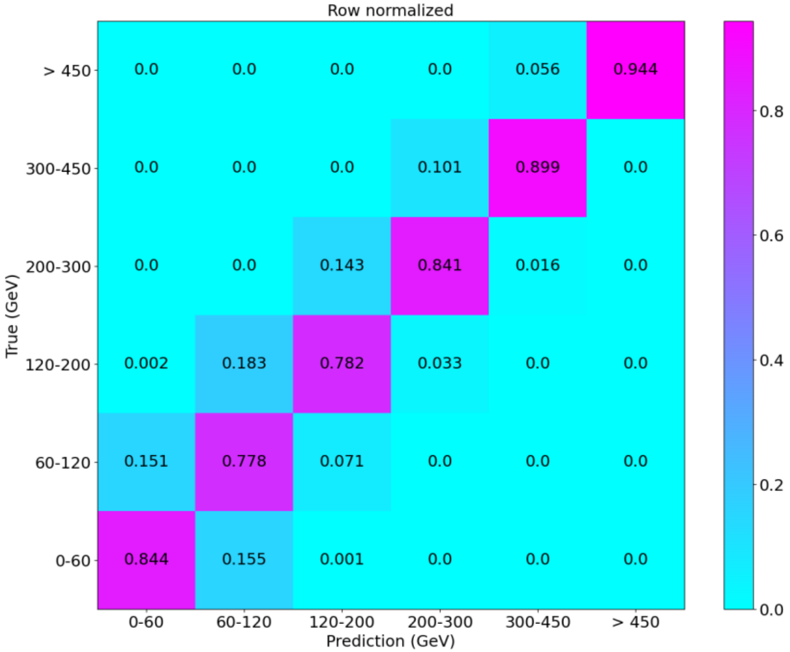




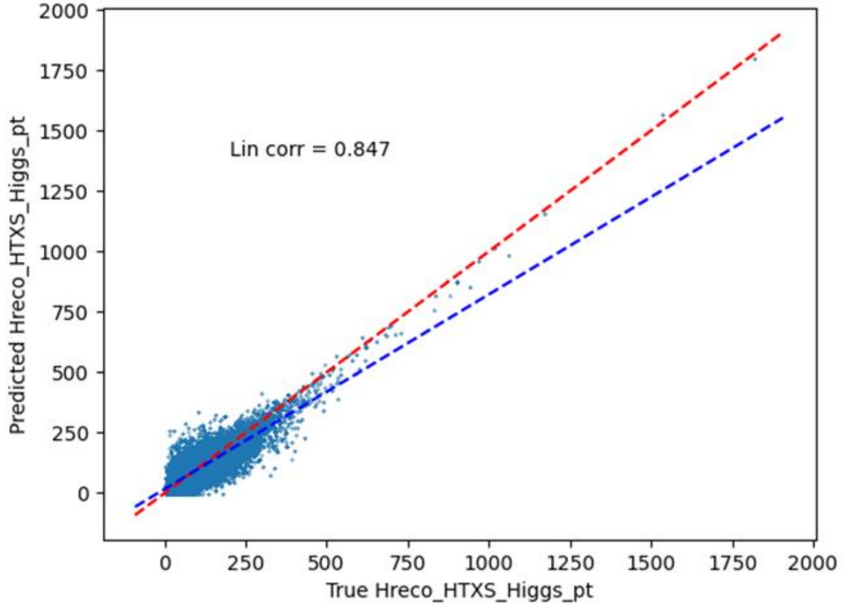
10k events: Lep_*, HadTop_*, met, met_phi, Jet0_*, Jet1_*, …

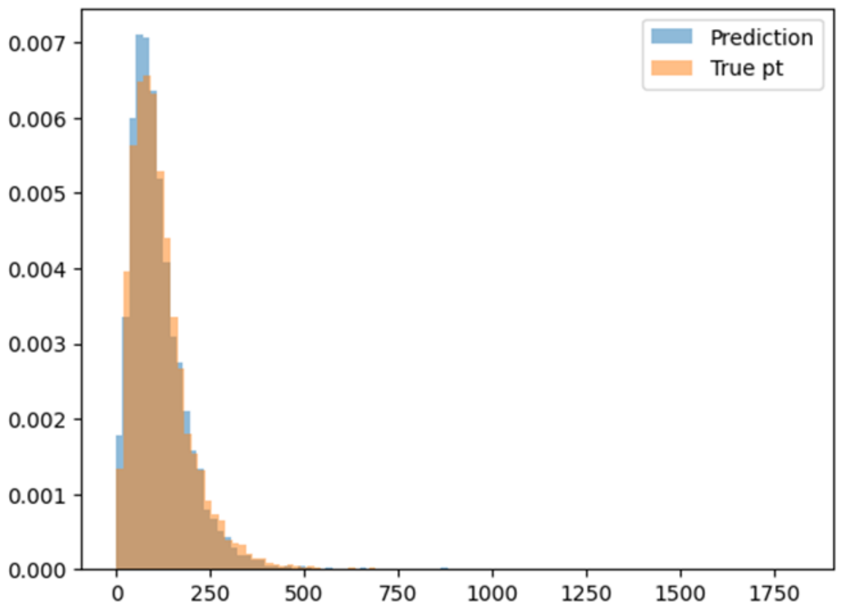


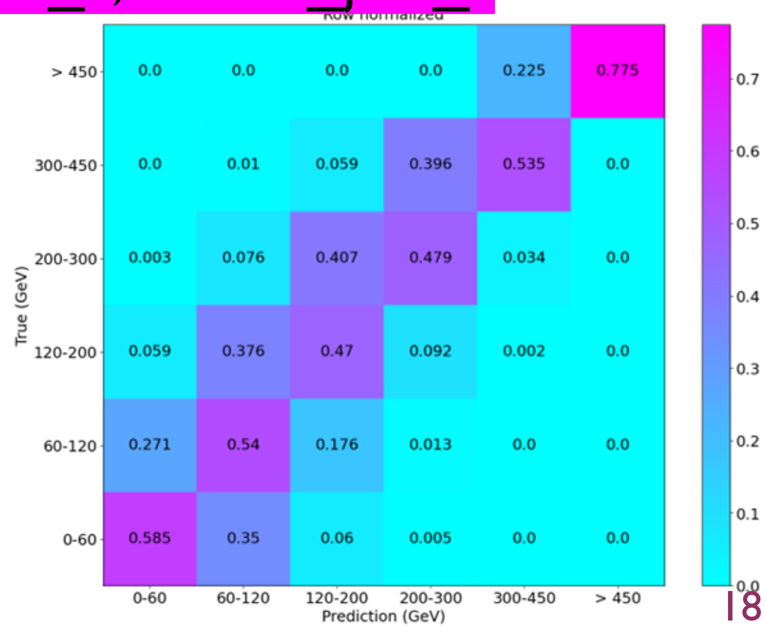




10k events: Lep_*, HadTop_*, met, met_phi, Jets_plus_lep_*, All5_jets_*, More5_jets_*







CONCLUSIONS

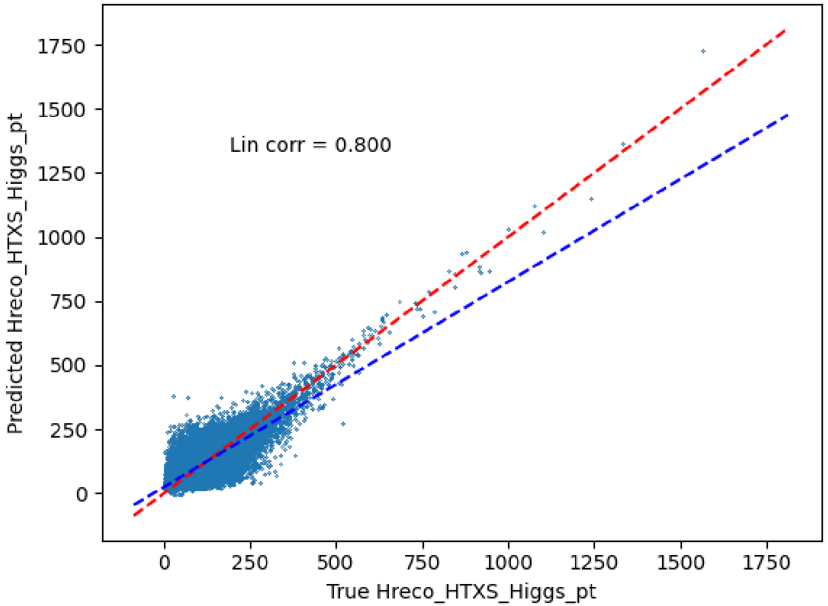
- Minimal NN is designed for the full MC simulation.
- Nominal deep model showed a better performance than the minimal network while using gen MC:
  - Individual neutrino kinematics improves model performance, but individual jet kinematics is more significant
  - Having full and separate individual jet kinematic data improves the regression.
- Near-perfect regression is possible using the full generator data.

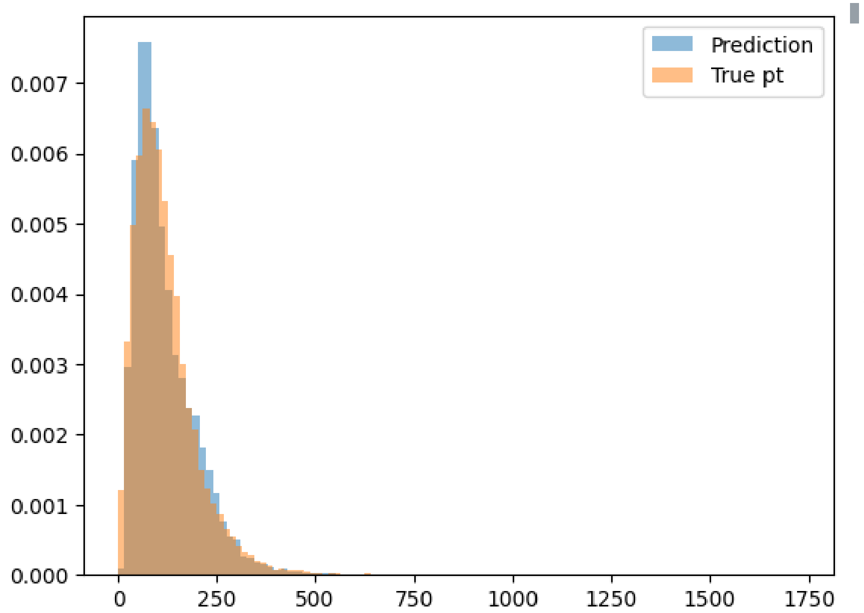
REFERENCES

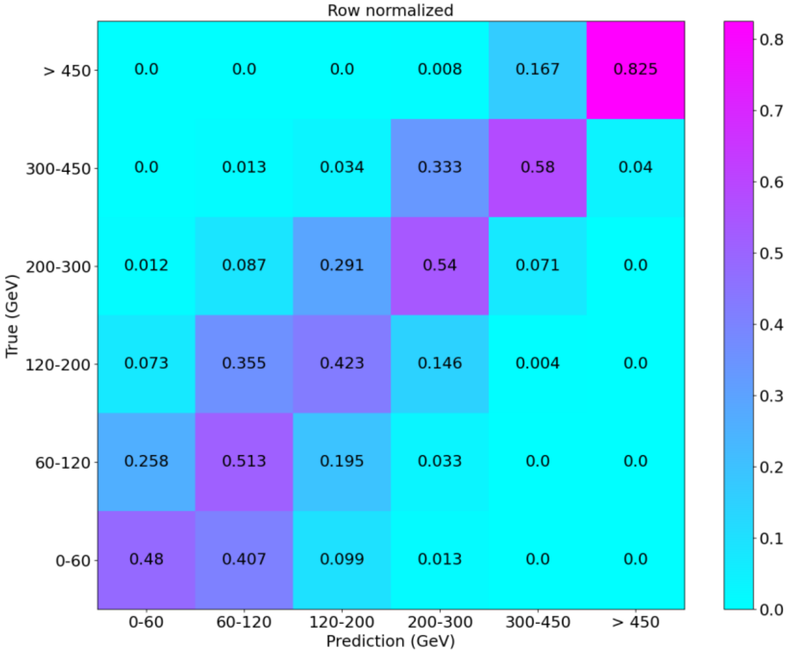
- [1] T. Sakuma, 3D SketchUp images of the CMS detector, CMS, 2016. https://cds.cern.ch/record/2628527
- [2] D. Barney, CMS Slice, 2015.
- [3] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, and Aben, Search for the associated production of the higgs boson with a top quark pair in multilepton final states with the atlas detector, Physics Letters B, Oct. 2015, 749, 519–541.
- [4] Sahraei, Amir & Chamorro, Alejandro & Kraft, Philipp & Breuer, Lutz. (2021). Application of Machine Learning Models to Predict Maximum Event Water Fractions in Streamflow. Frontiers in Water. 3. 652100. 10.3389/frwa.2021.652100.

Backup

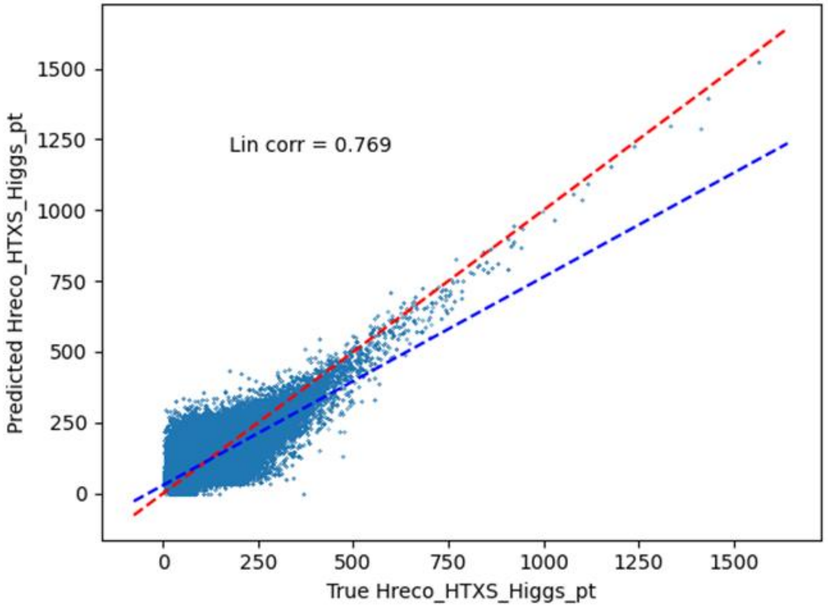
20k events: Lep_*, HadTop_*, met, met_phi, Jets_plus_lep_*, All5_jets_*, More5_jets_*

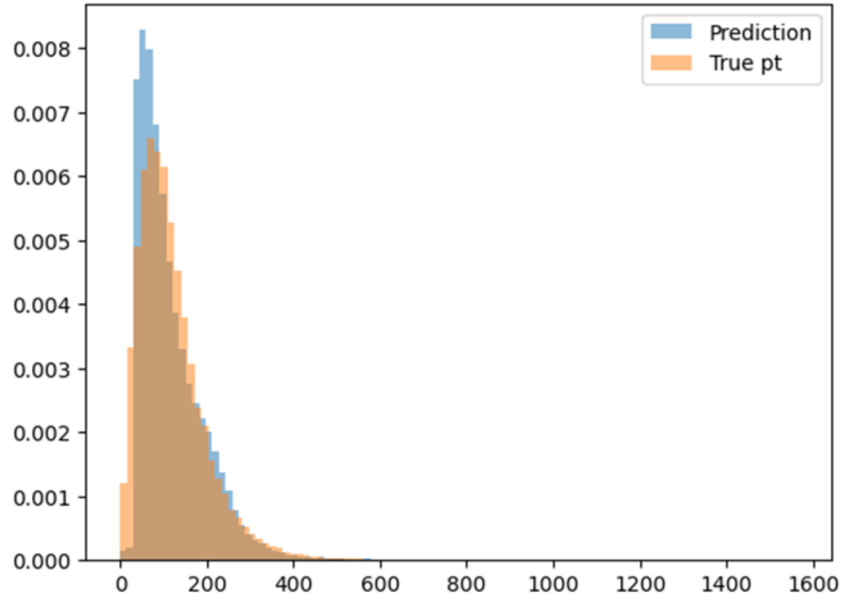


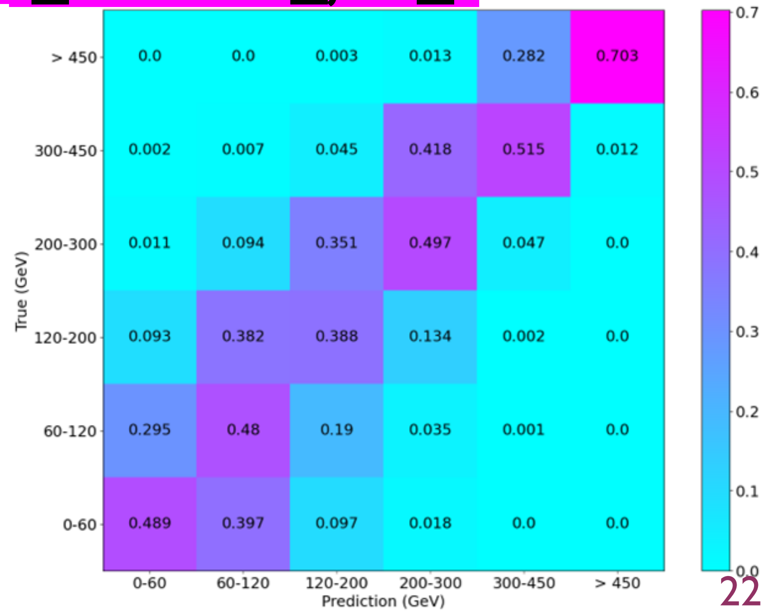




65k events: Lep_*, HadTop_*, met, met_phi, Jets_plus_lep_*, All5_jets_*, More5_jets_*

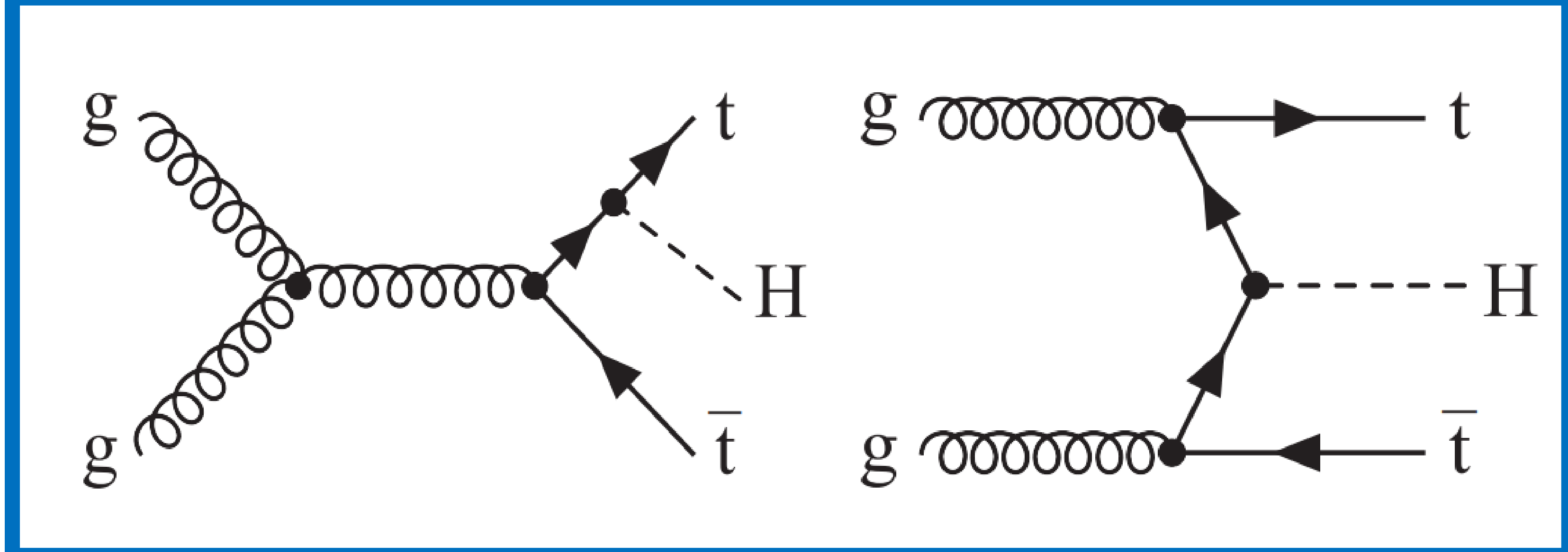


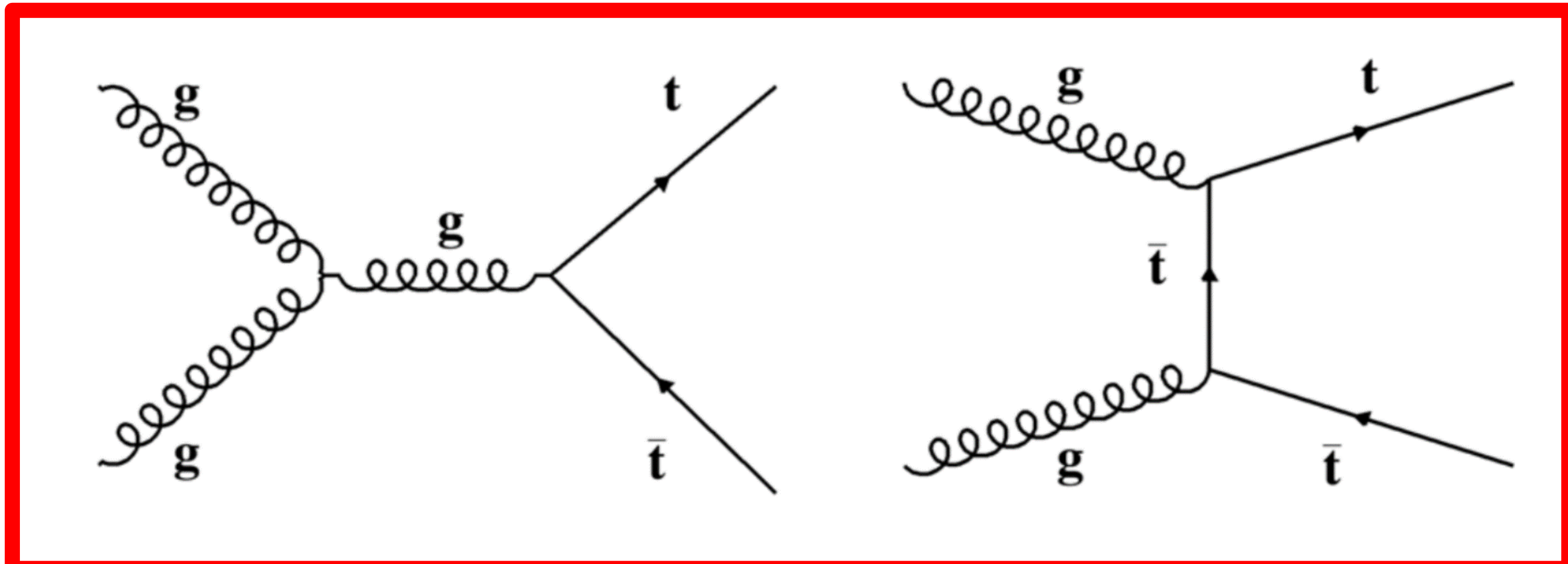




Backup 1: tth procesas

- Atliktame tyrime simuliuojamas $t\bar{t}H$ procesas, dešinėje pateiktos jo esminės Feynmano diagramos.
- top – t, anti-top – $\bar{t}$.
- Viršuje matomos $t\bar{t}H$ susidarymo proceso diagramos.
- Žemiau tik $t\bar{t}$ susidarymo diagramos.
- Šis procesas svarbus Higss‘o bozono tyrimuose, nes leidžia matuoti Yukawa sąryšio konstatą tarp Higgs ir fermionų.
- Vertės nuokrypis gali indikuoti naują fiziką neapibrėžtą standartinio modelio.





Backup 2: Pagrindiniai MC generacijos parametrai

- Monte Carlo įvykių generatorius - MadGraph5_aMC@NLO:
  - MG versija - v3.5.3;
  - Įvykių skaičius n_{events} = 10 000;
  - Masės centro energija $\sqrt{s}$ = 13.6 TeV;
  - ebeam1 = 6800 GeV;
  - ebeam2 = 6800 GeV.