QG Tag latest version

Florencia

Vars

4.4 High-level Jet Variables

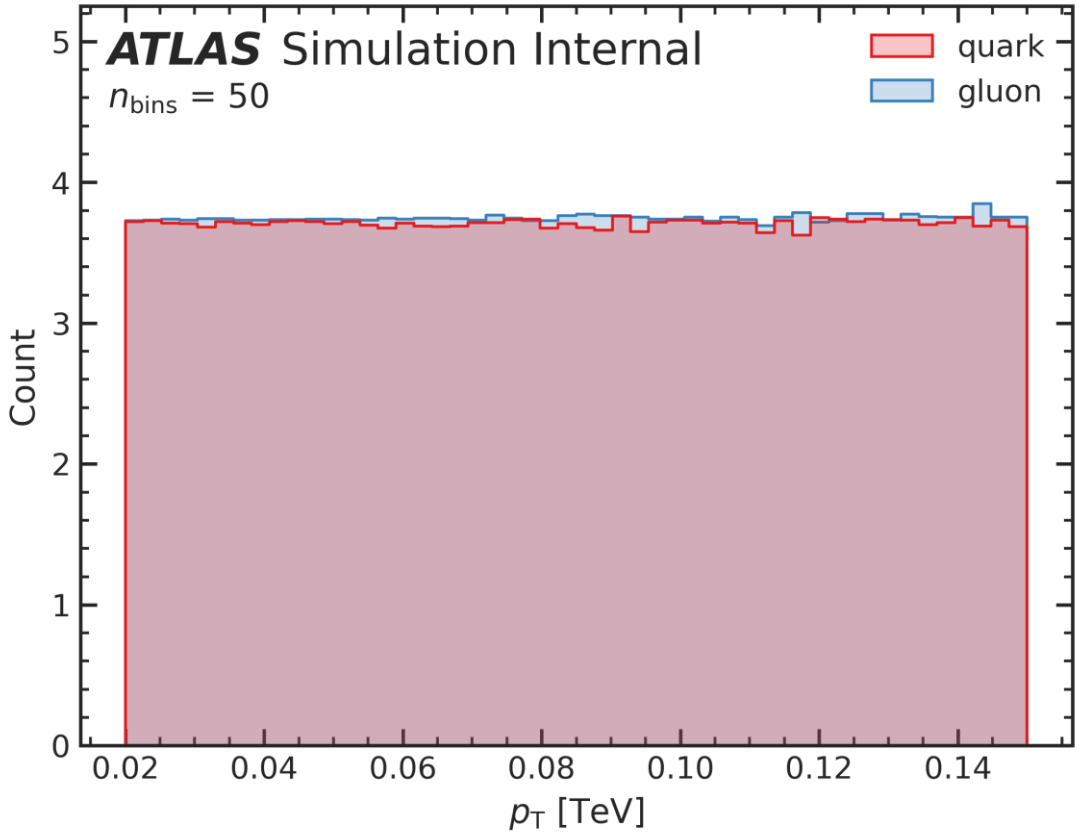
High-level jet variables are calculated from the jet constituents and describe the jet as a whole. They are used as input to the FC and Highway networks. There are 10 of these variables, forming a 10-dimensional vector as the input to the neural network. They are:

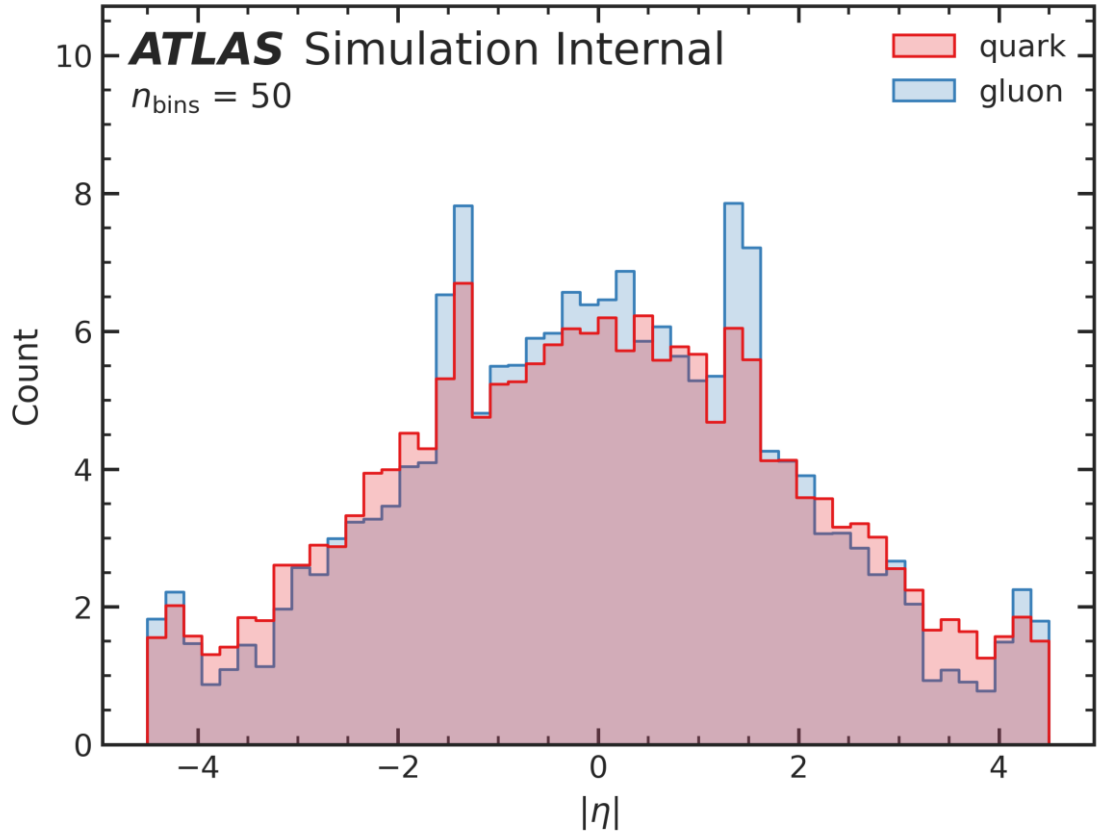
http://cds.cern.ch/record/2878932/files/ATL-PHYS-PUB-2023-032.pdf

- Calibrated jet four-momentum.
- Number of PFOs.
- Numbers of charged PFOs passing two different pT thresholds (500 and 1000 MeV).
- Jet width [15] calculated using PFOs with $p_{\mathrm{T}} > 1000$ MeV.
- Fraction of the jet energy deposited in the EM calorimeter. Fraction of the jet energy carried by charged particles.

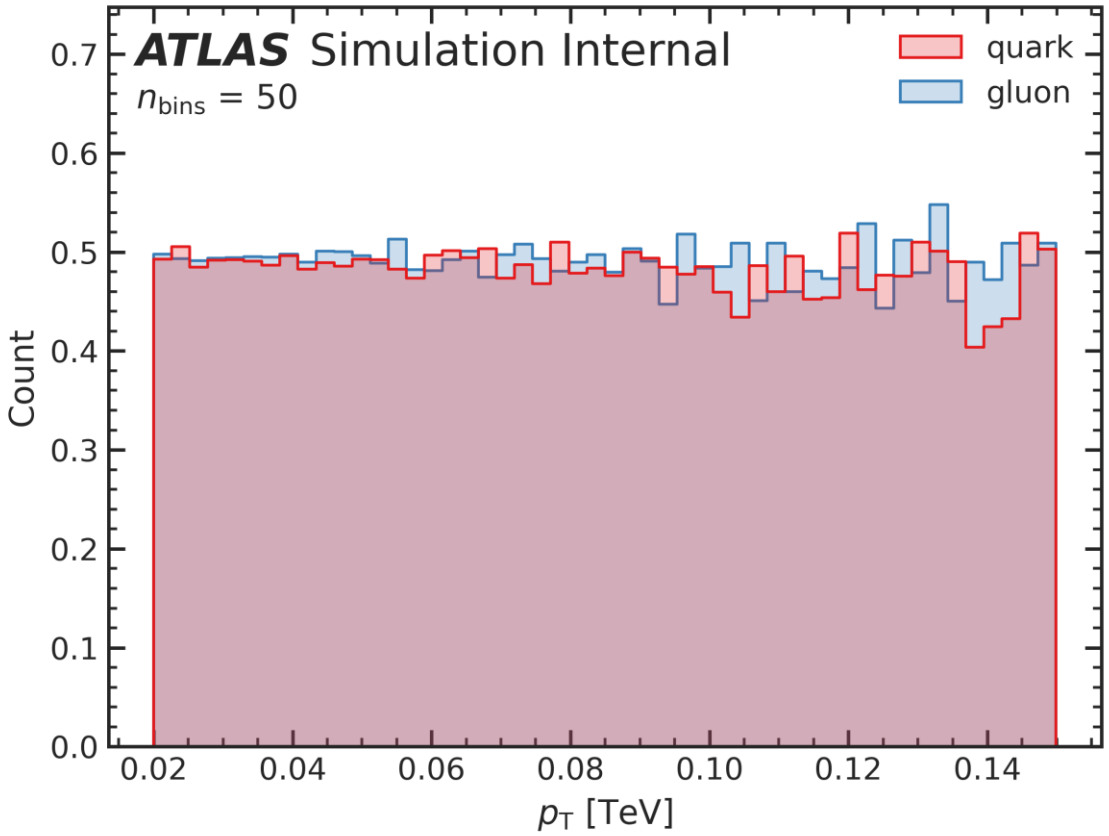
In addition to these variables, a second set of 5 high-level jet variables is employed. These variables correspond to the ones utilized in the prior ATLAS q/g tagger, as detailed in the reference [15]. Concretely, these encompass jet p_{T} and η and the count of PFOs within a jet. The last two variables are the jet width, and the two-point correlation function [13, 16], both calculated using the PFOs.

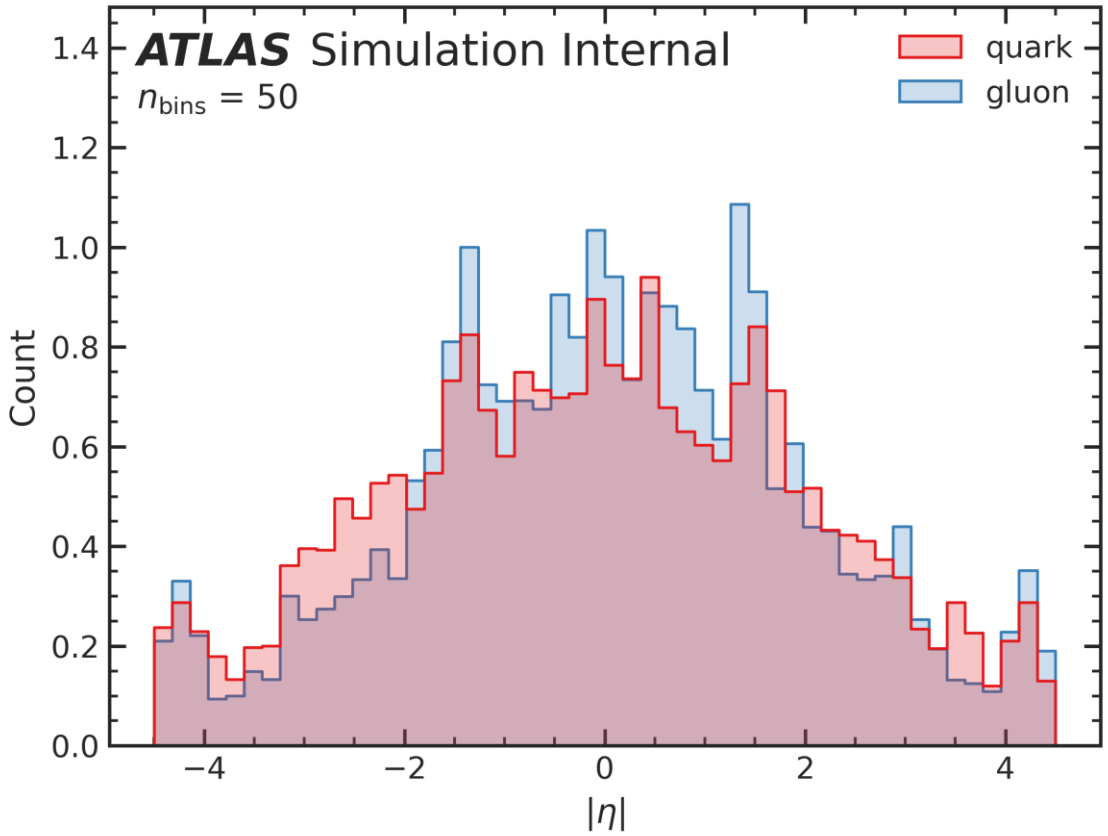
Some plots: flat training (weights for HL-LHC)



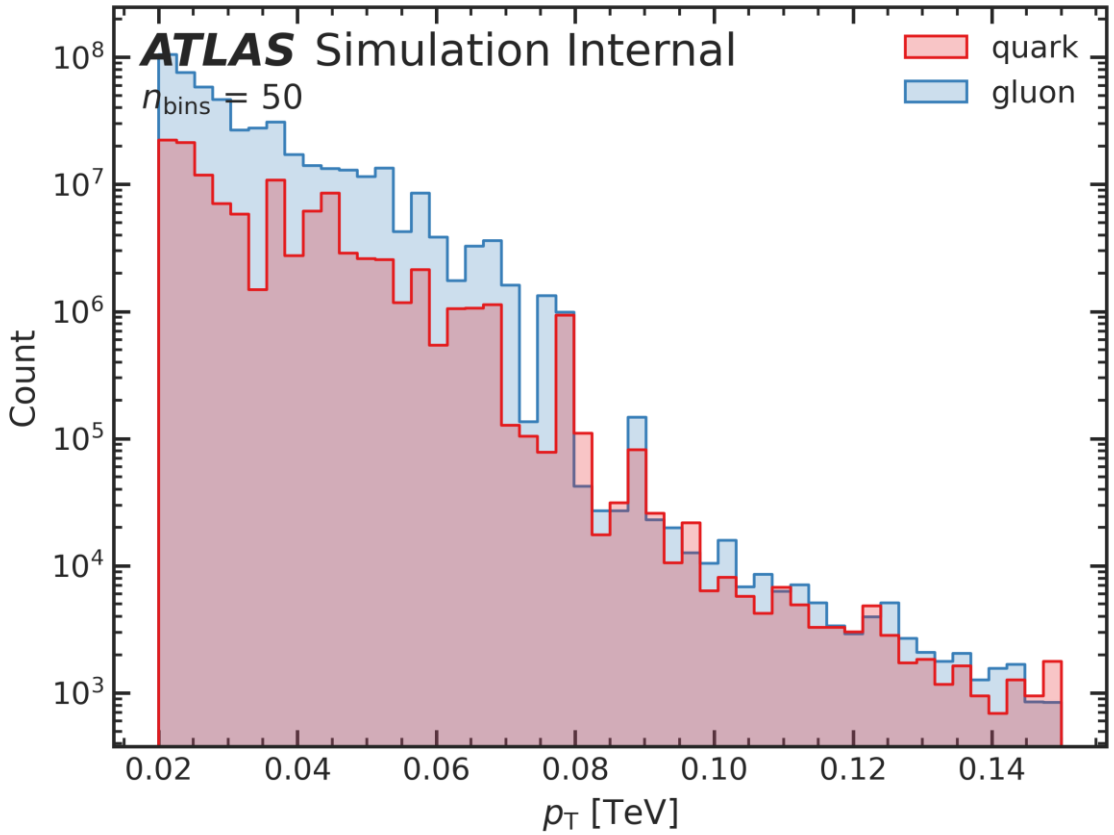


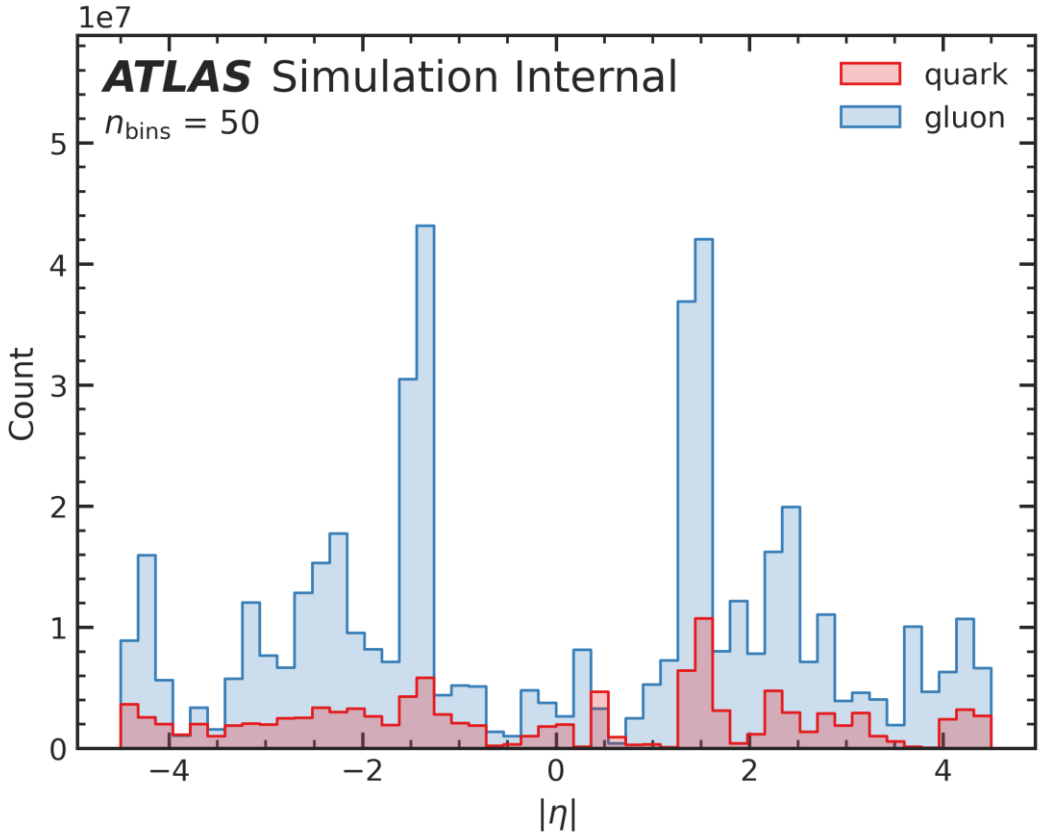
Validation data





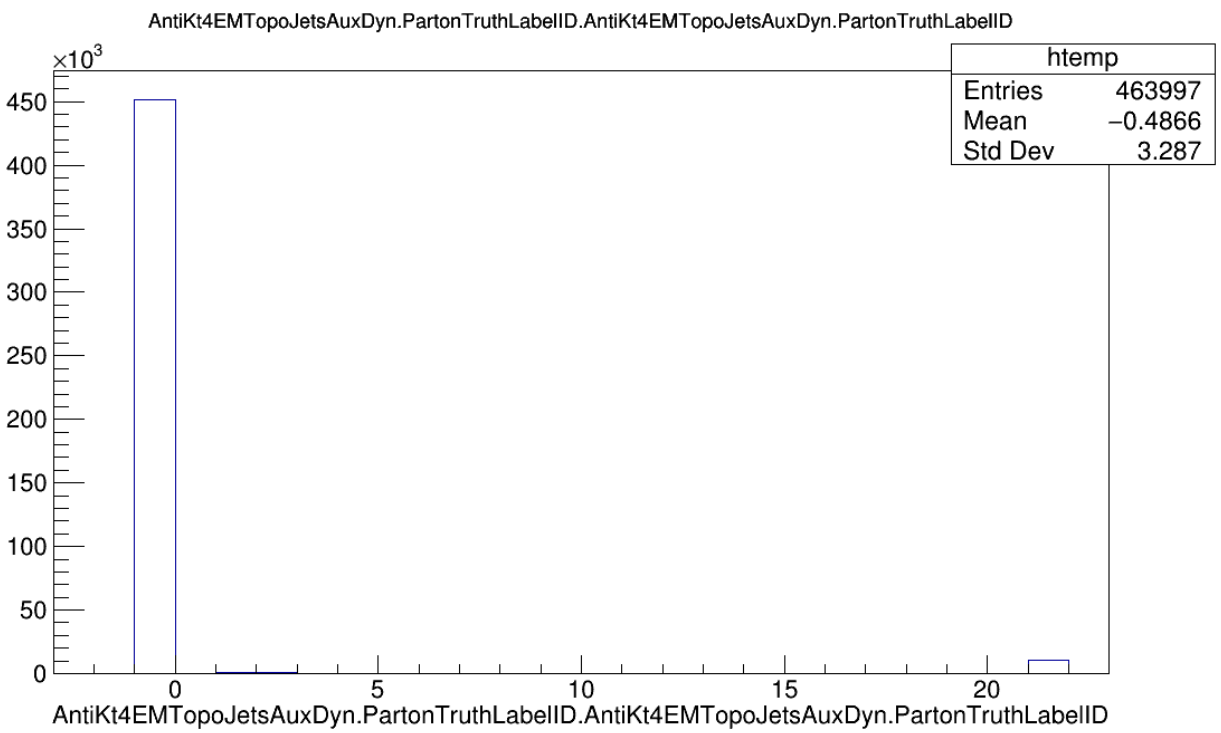
Test data (physics weights)

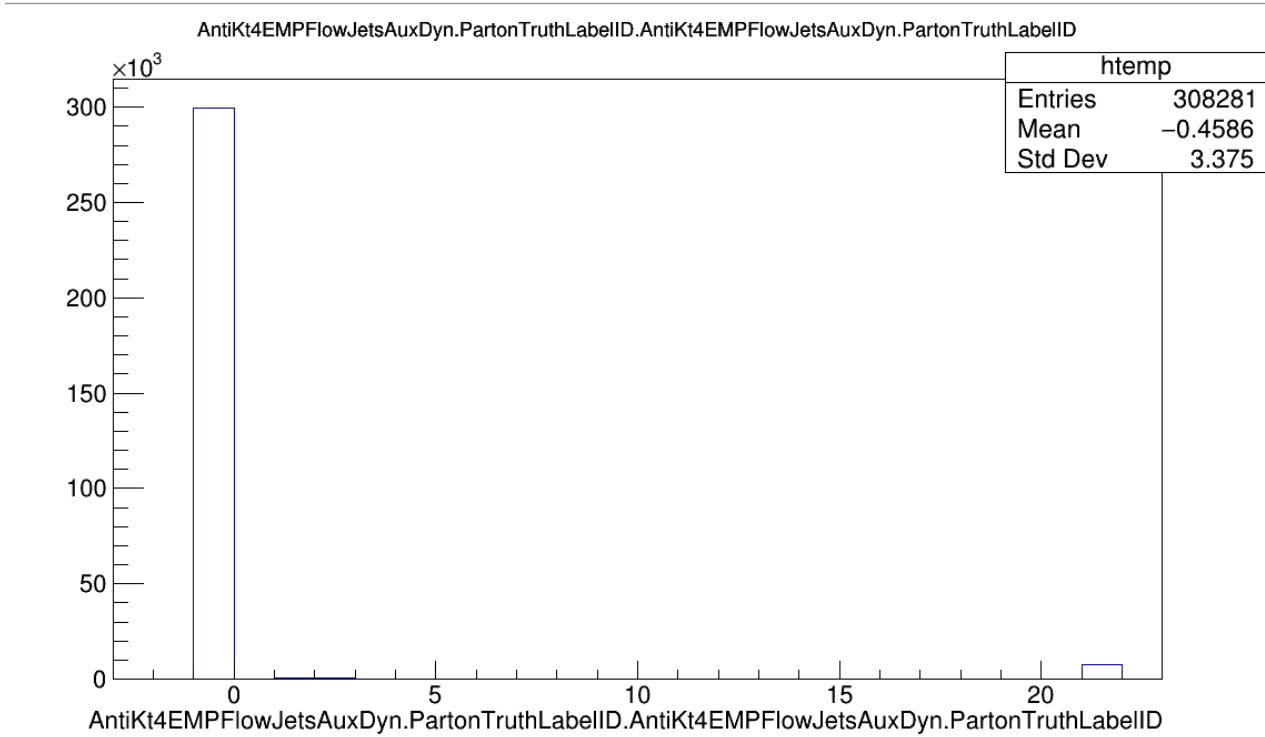




PartonTruthLabel ID

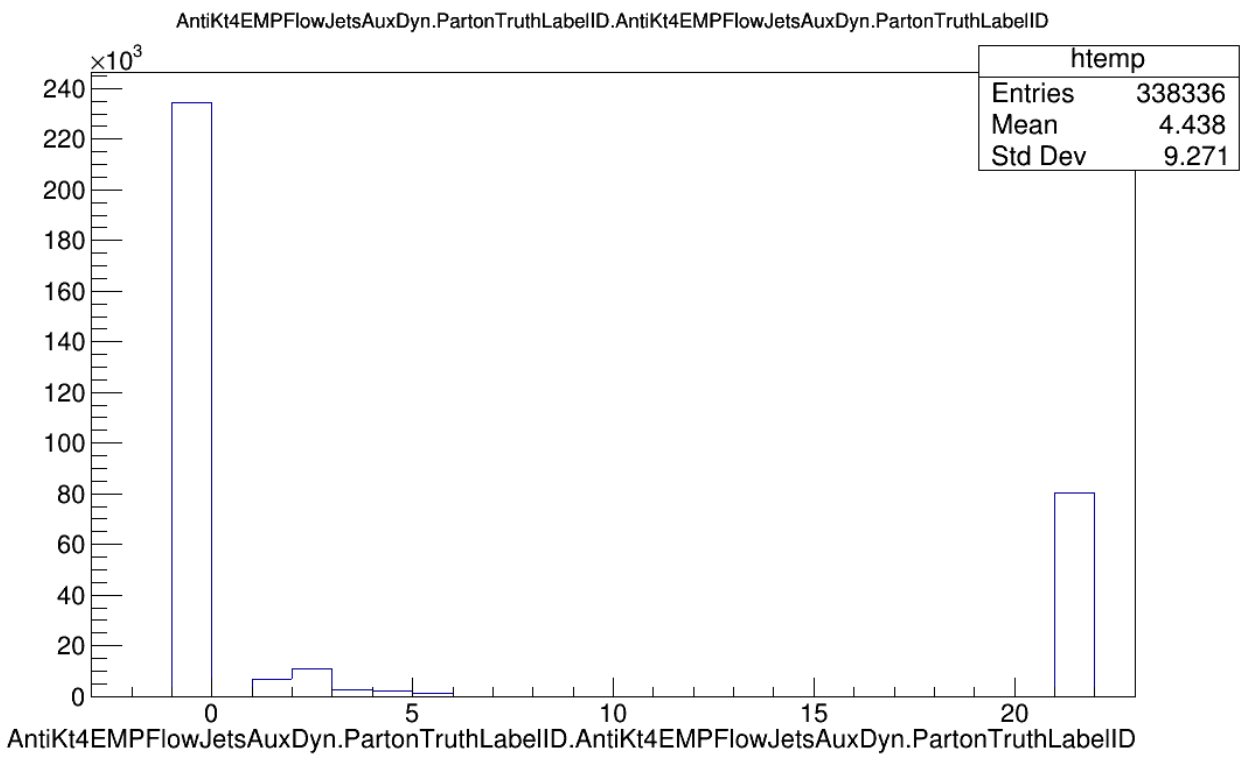
- mc21_14TeV.801165.Py8EG_A14NNPDF23LO_jj_JZ0.deriv.DAOD_PHYSVAL.e8481_s4038_r14365_p5608/DAOD_PHYSVAL.32919409._000001.pool.root.1

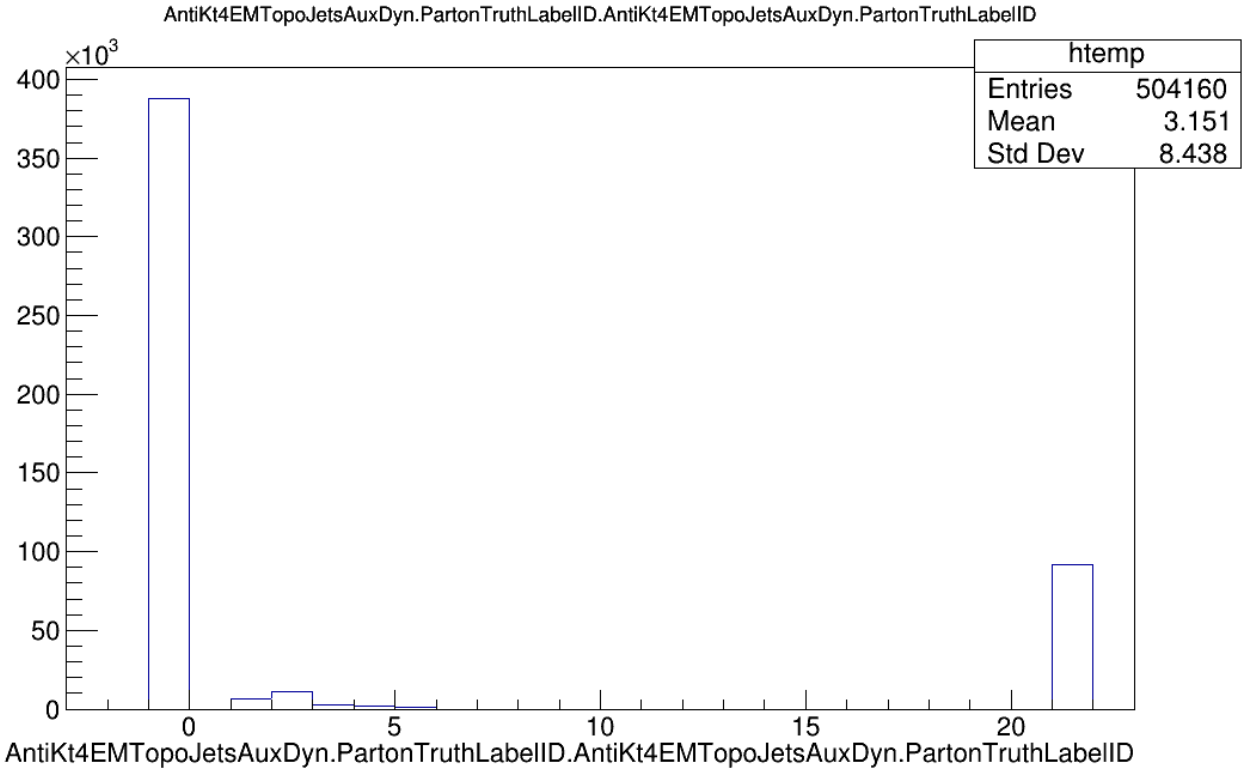




PartonTruthLabel ID

- mc21_14TeV.801170.Py8EG_A14NNPDF23LO_jj_JZ5.deriv.DAOD_PHYSVAL.e8481_s4038_r14365_p5608/DAOD_PHYSVAL.32919419._000001.pool.root.1





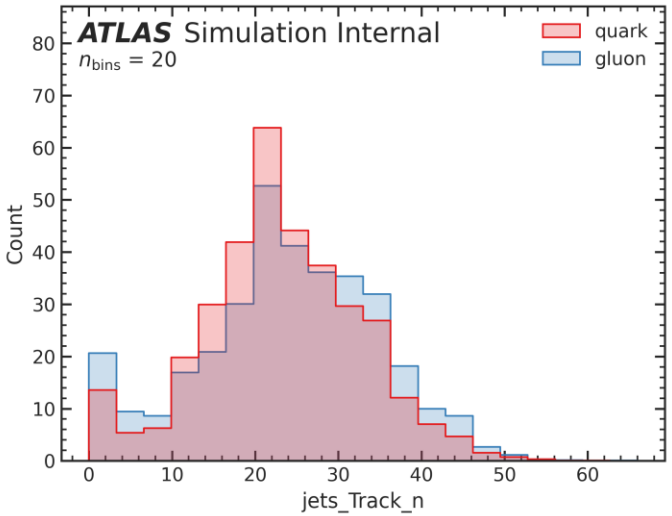
Samples used for the following studies

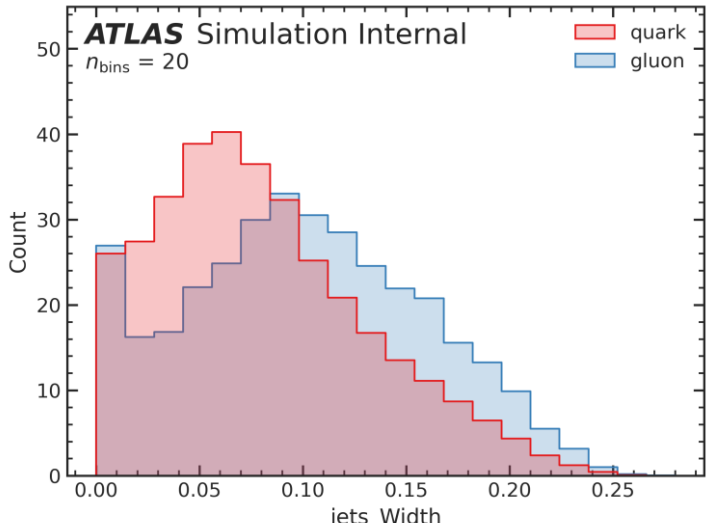
- mc21_14TeV.600026.PhH7EG_NNPDF3_AZNLO_VBFH125_ZZ4nu_MET75.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.601230.PhPy8EG_A14_ttbar_hdamp258p75_dil.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.601237.PhPy8EG_A14_ttbar_hdamp258p75_allhad.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.801165.Py8EG_A14NNPDF23LO_jj_JZ0.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.801166.Py8EG_A14NNPDF23LO_jj_JZ1.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.801167.Py8EG_A14NNPDF23LO_jj_JZ2.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.801168.Py8EG_A14NNPDF23LO_jj_JZ3.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.801169.Py8EG_A14NNPDF23LO_jj_JZ4.recon.AOD.e8481_s4038_r14365
- mc21_14TeV.801170.Py8EG_A14NNPDF23LO_jj_JZ5.recon.AOD.e8481_s4038_r14365

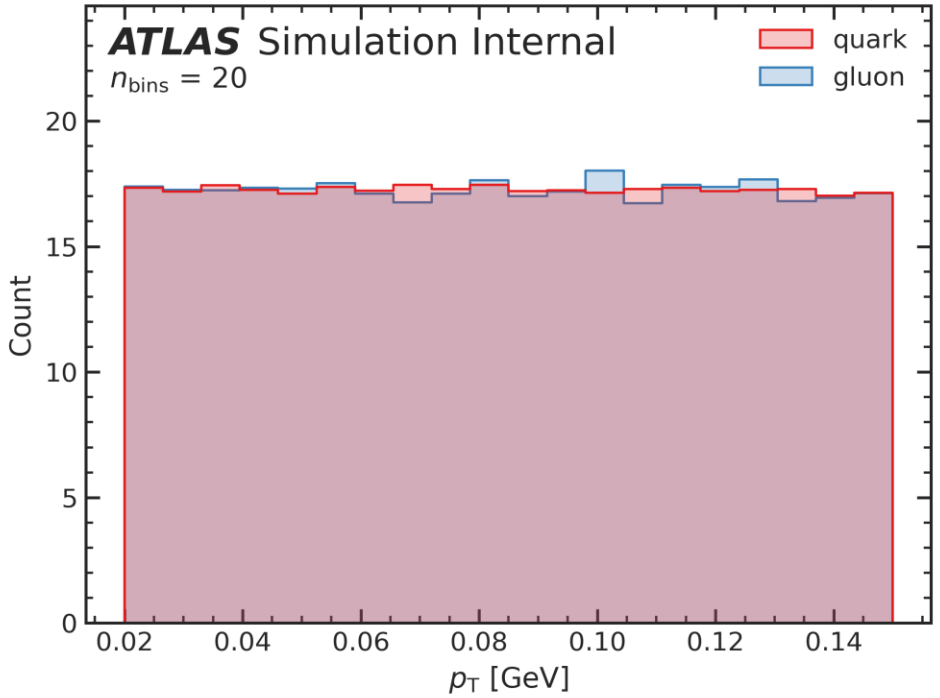
Hight stast samples still pending:
https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/JetEtmissHLLHC#Low_statistics_first_pass_latest

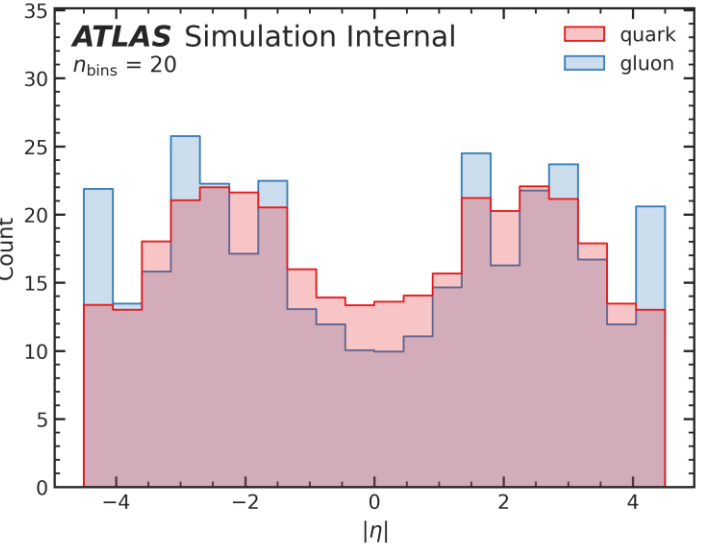
Exploring input variables

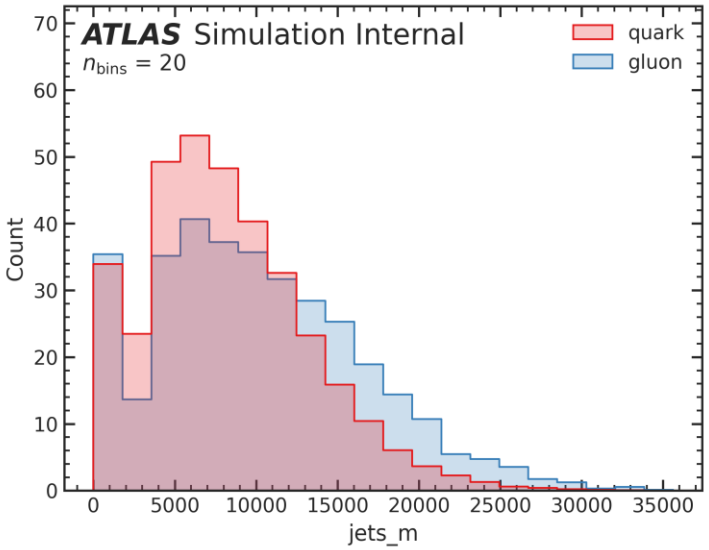
- mc21_14TeV.600026.PhH7EG_NNPDF3_AZNLO_VBFH125_ZZ4nu_MET75.





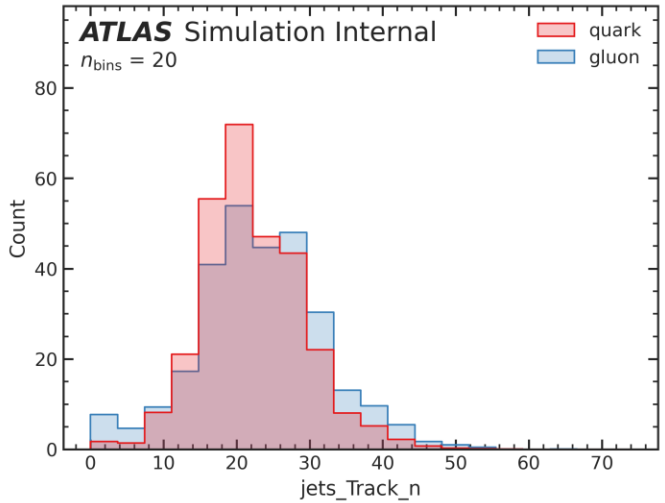


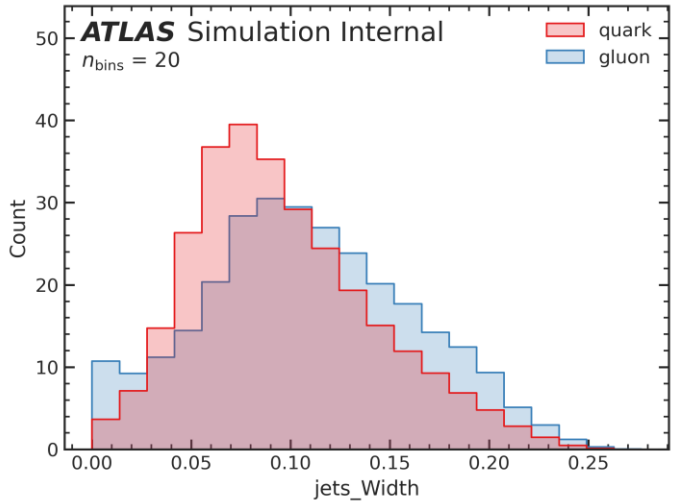


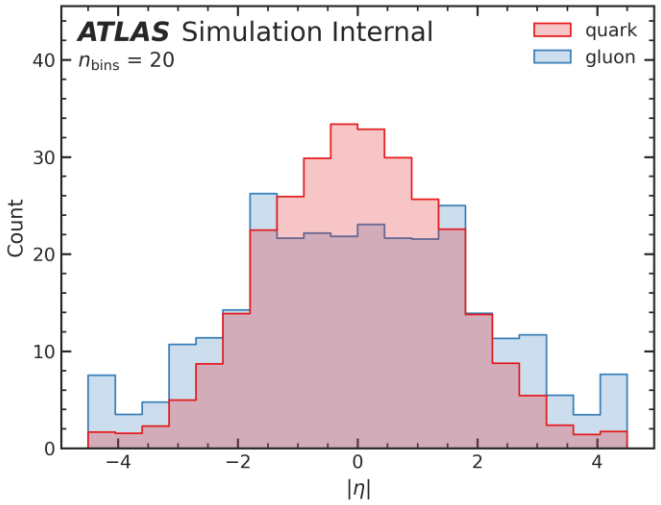


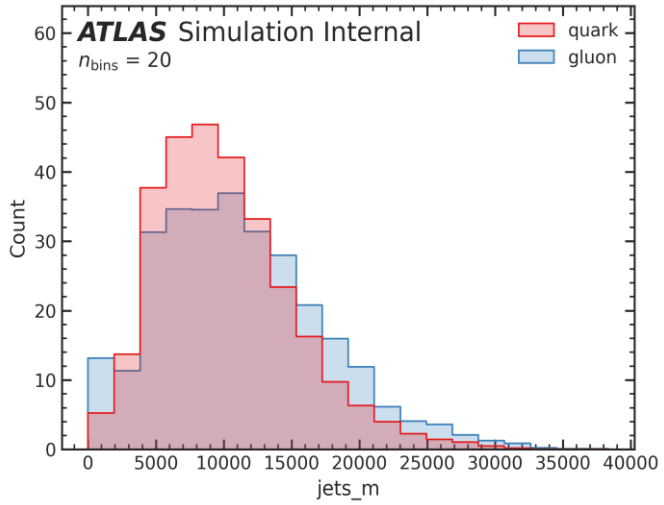
Exploring input variables

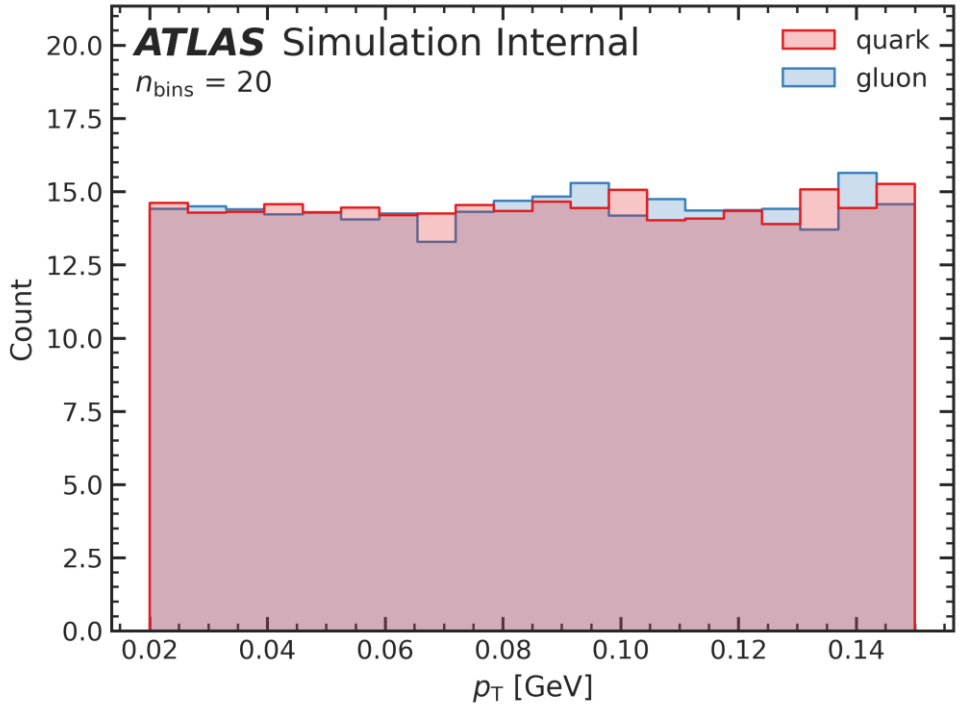
- mc21_14TeV.601229.PhPy8EG_A14_ttbar_hdamp258p75_SingleLep.





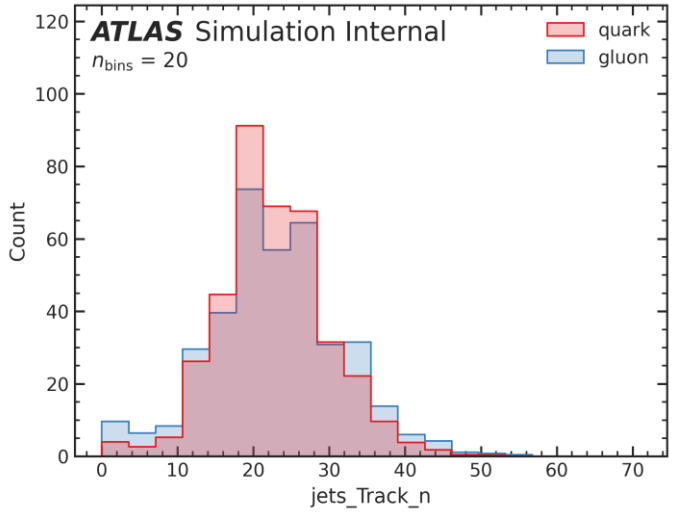


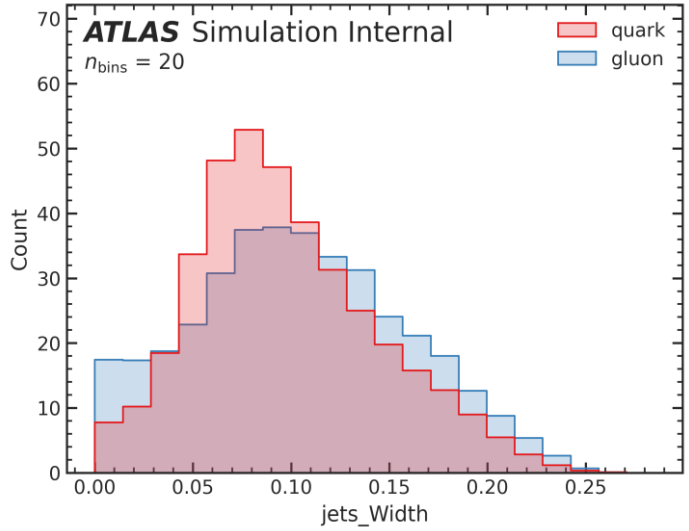


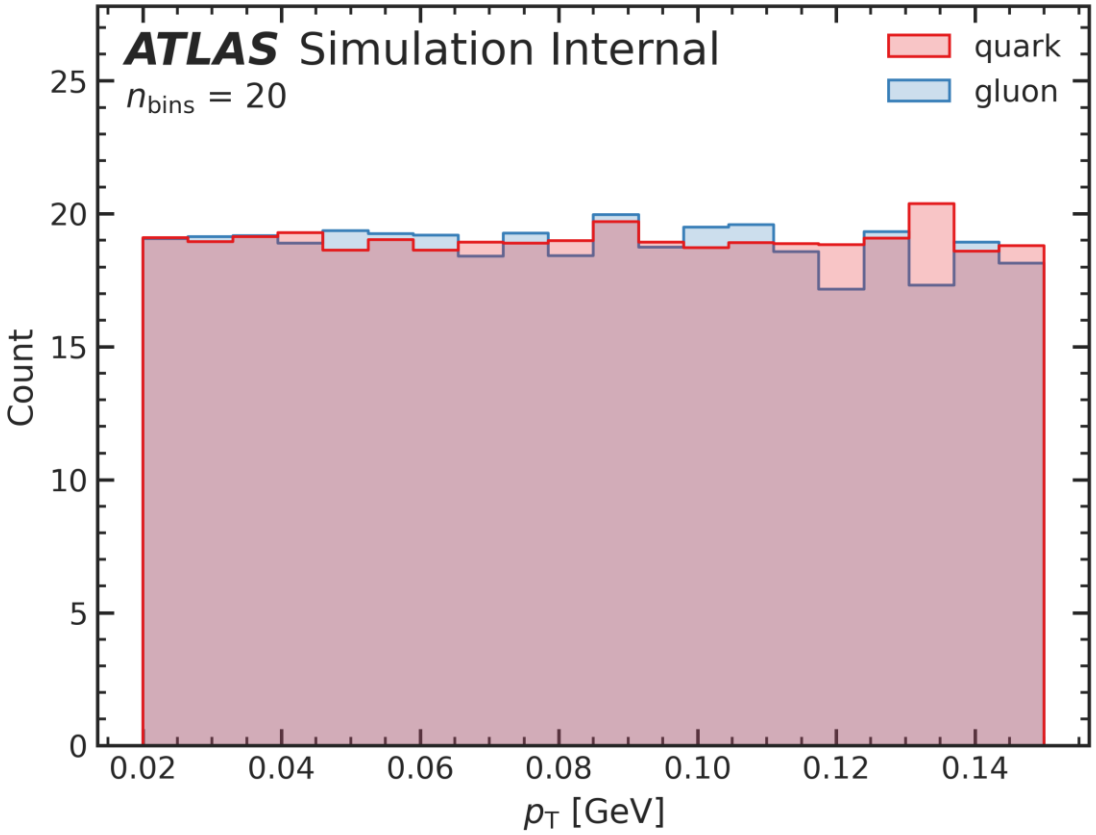


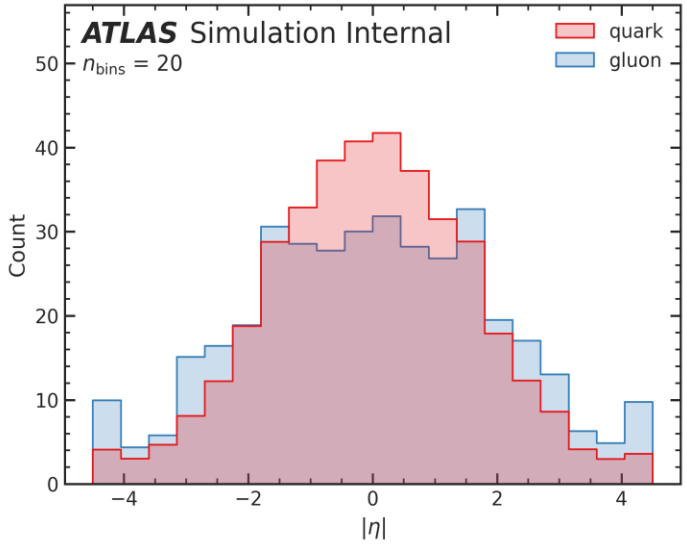
Exploring input variables

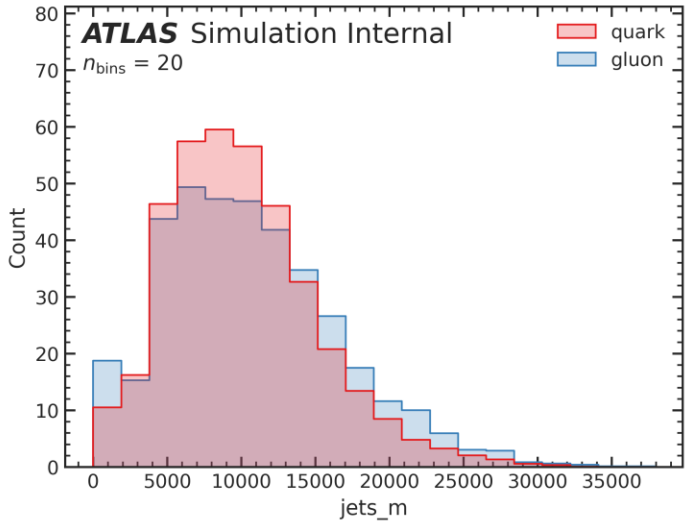
- mc21_14TeV.601230.PhPy8EG_A14_ttbar_hdamp258p75_dil.





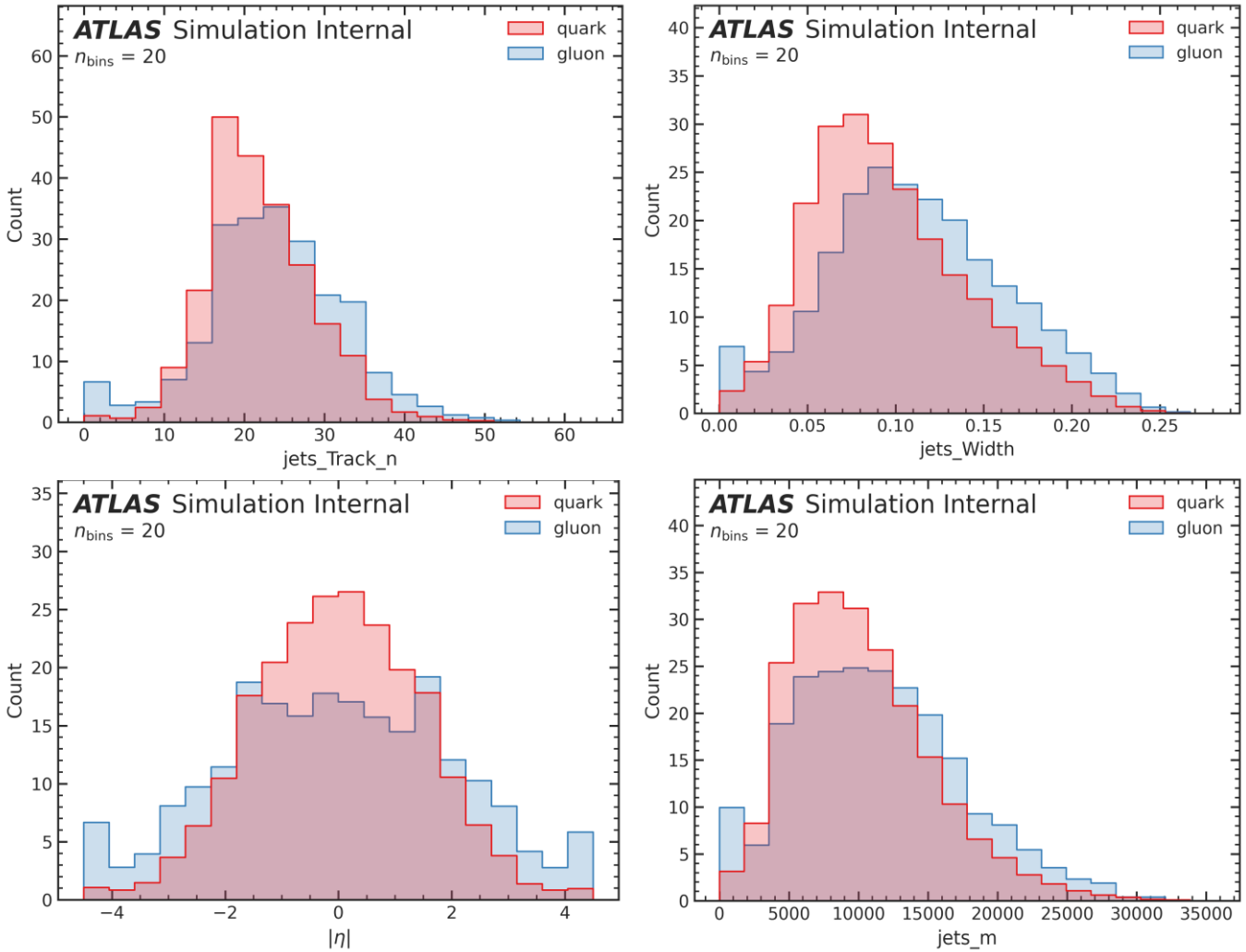


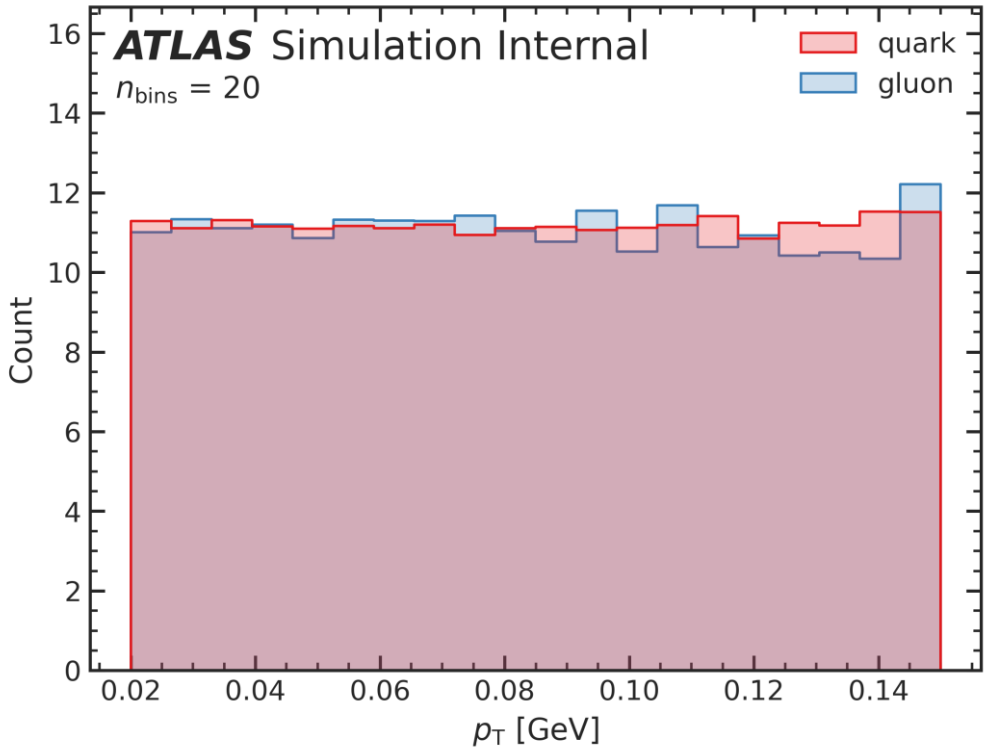




Exploring input variables

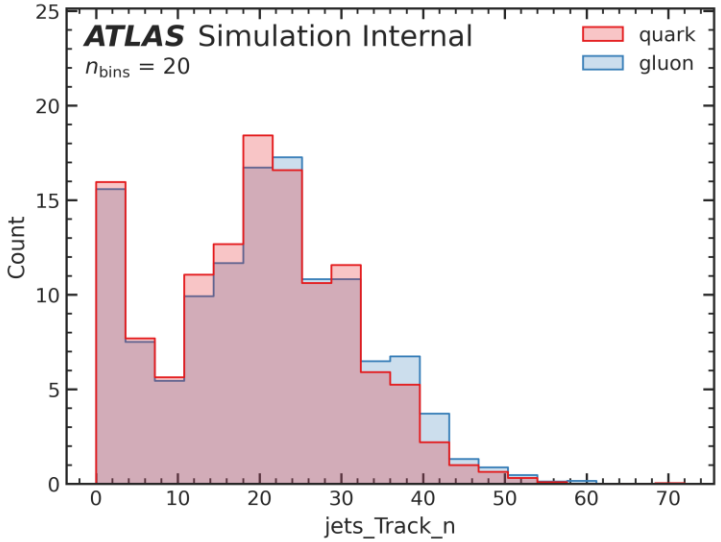
- mc21_14TeV.601237.PhPy8EG_A14_ttbar_hdamp258p75_allhad.recon

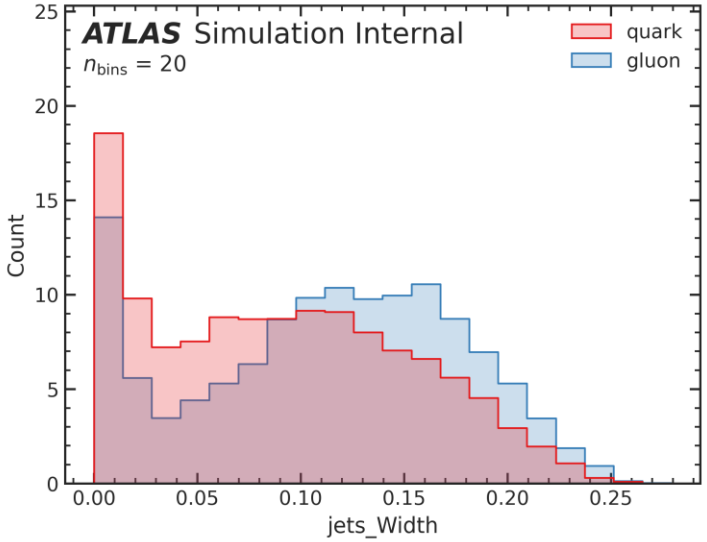


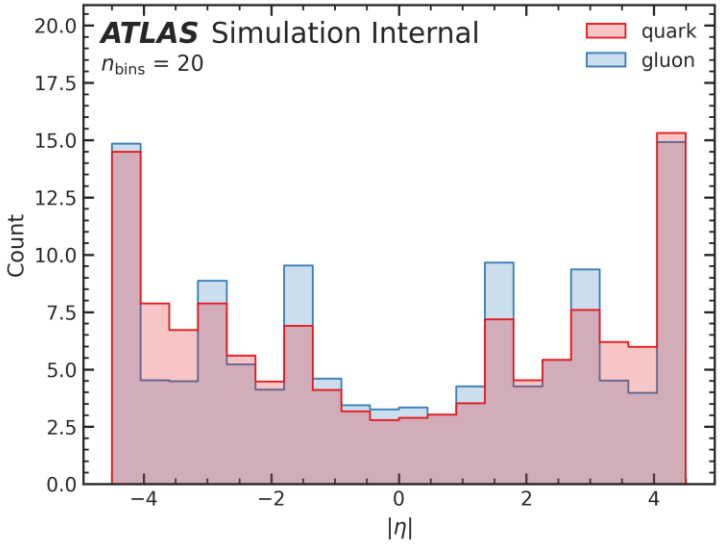


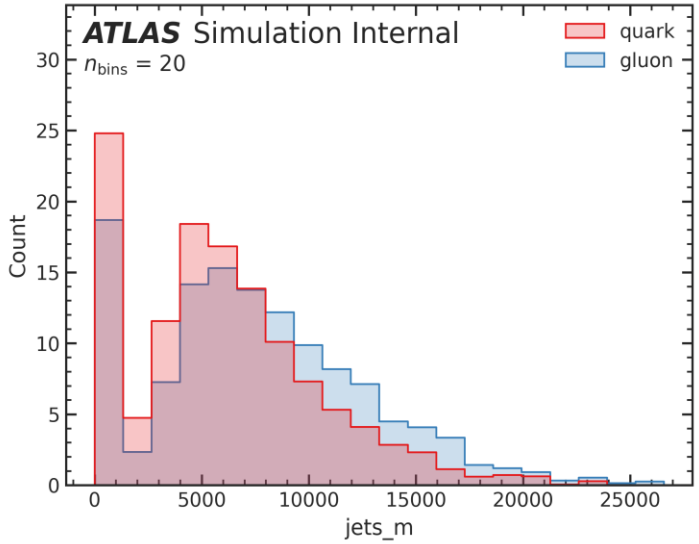
Exploring input variables

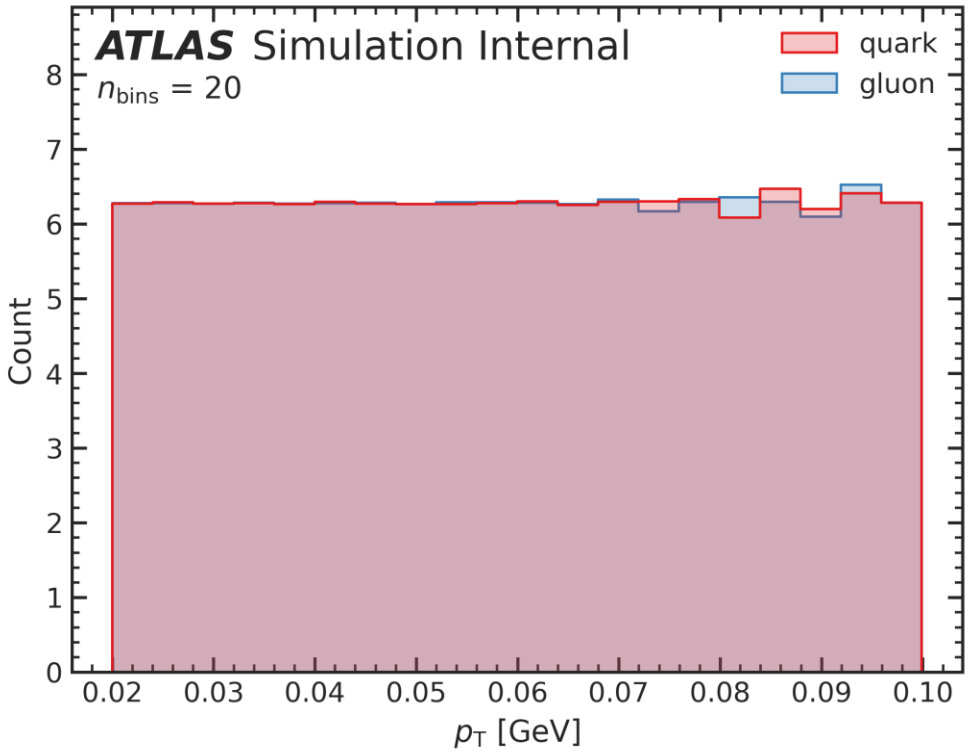
- mc21_14TeV.801166.Py8EG_A14NNPDF23LO_jj_JZ1





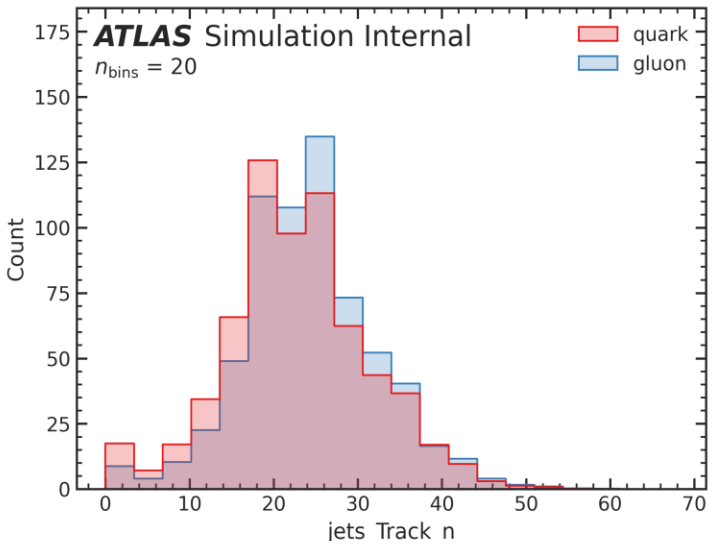


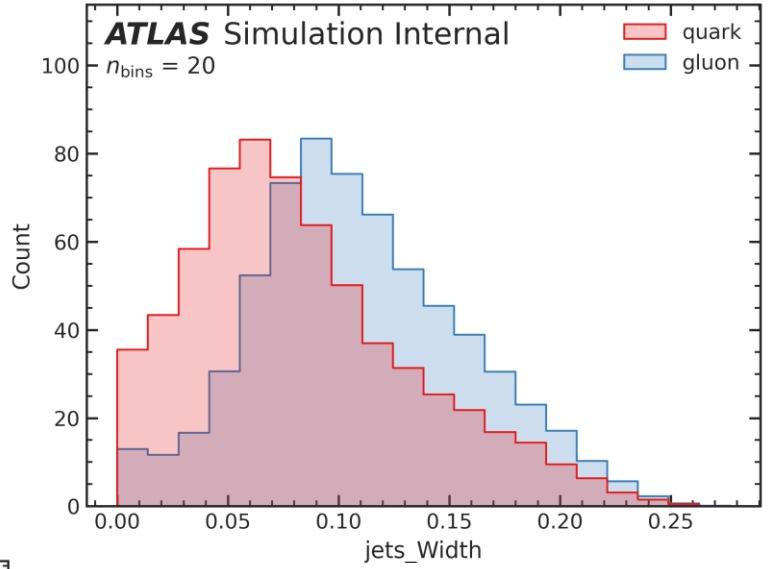


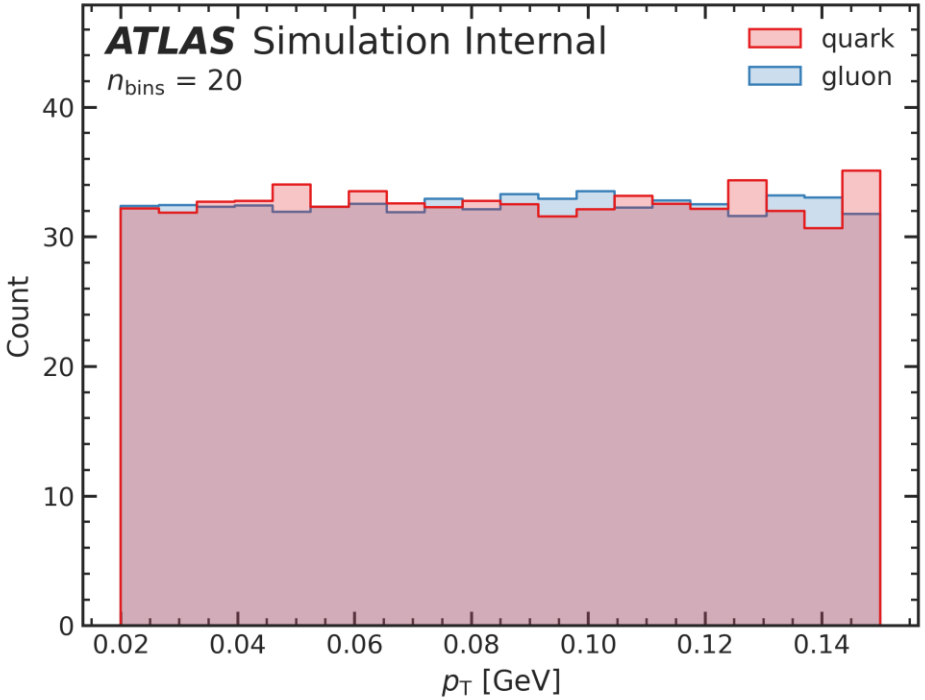


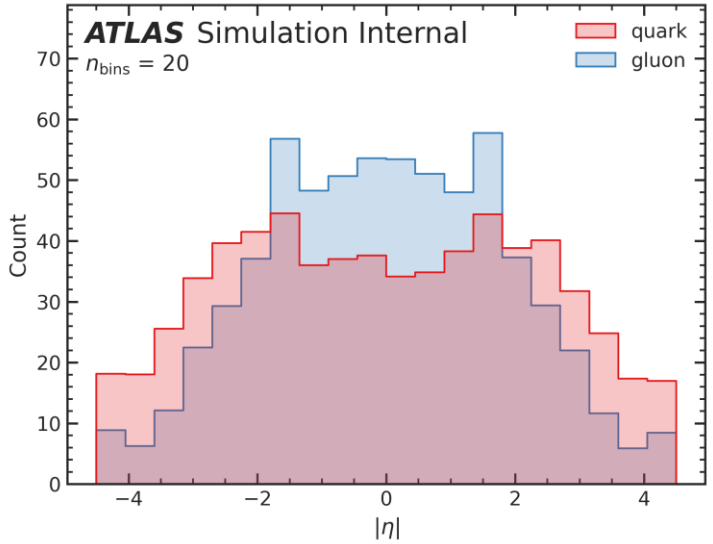
Exploring input variables

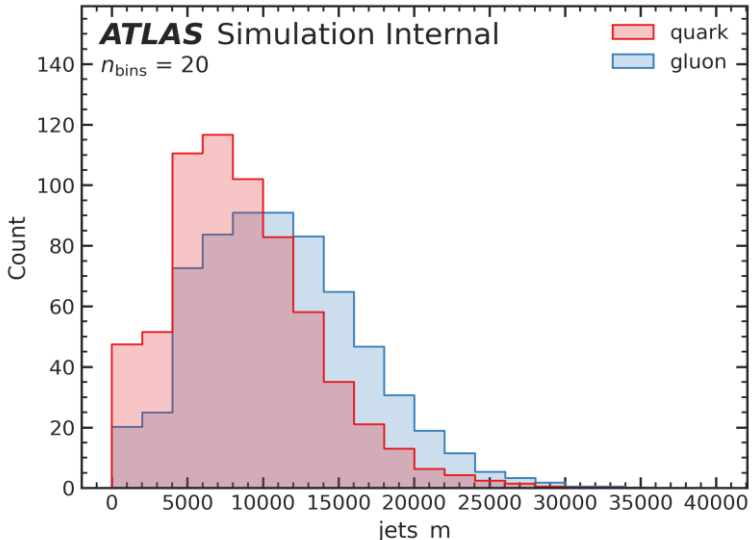
- mc21_14TeV.801167.Py8EG_A14NNPDF23LO_jj_JZ2





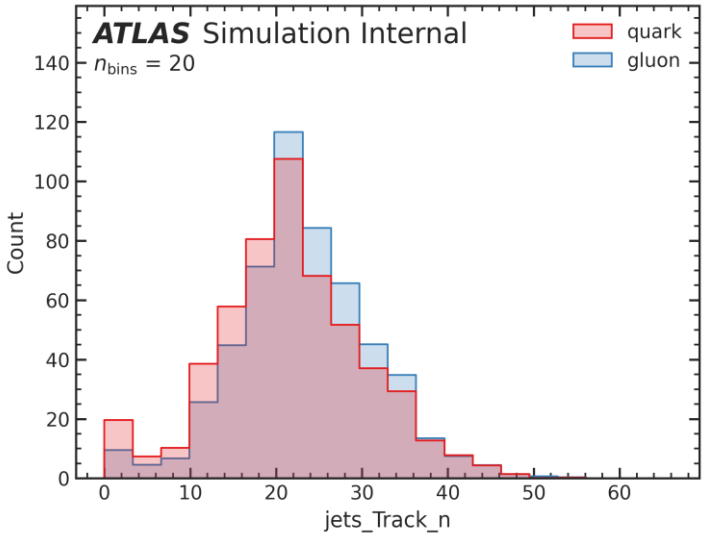


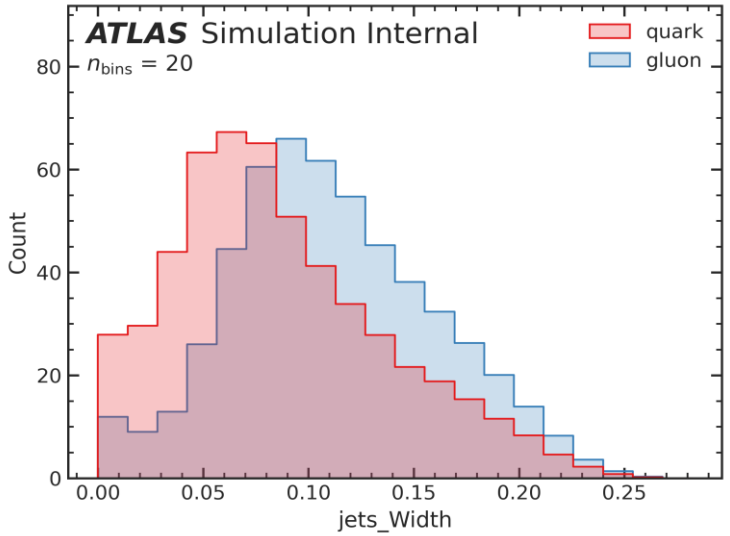


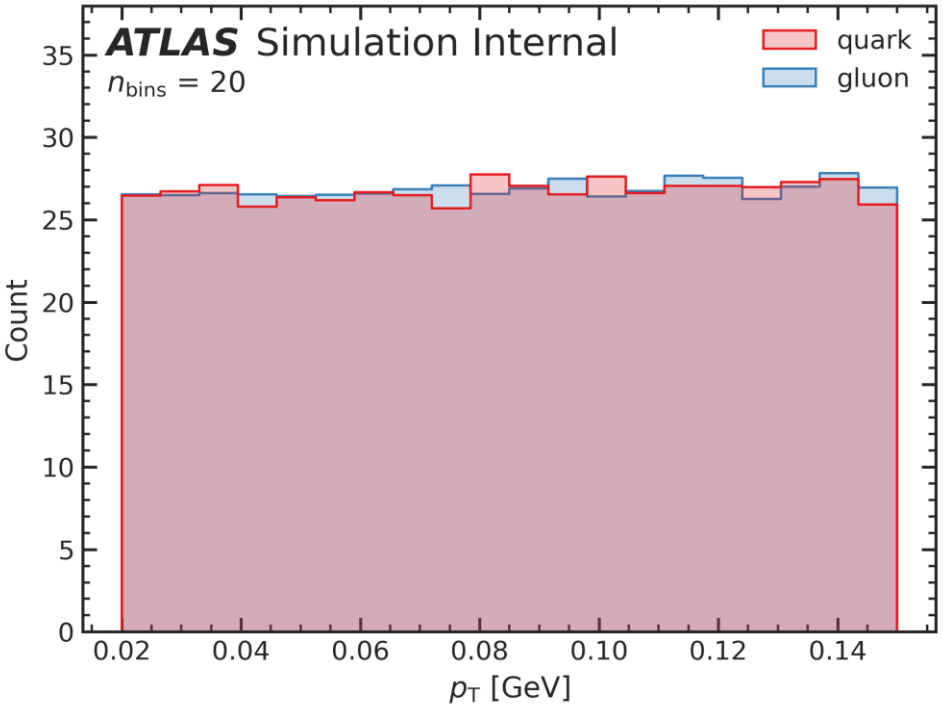


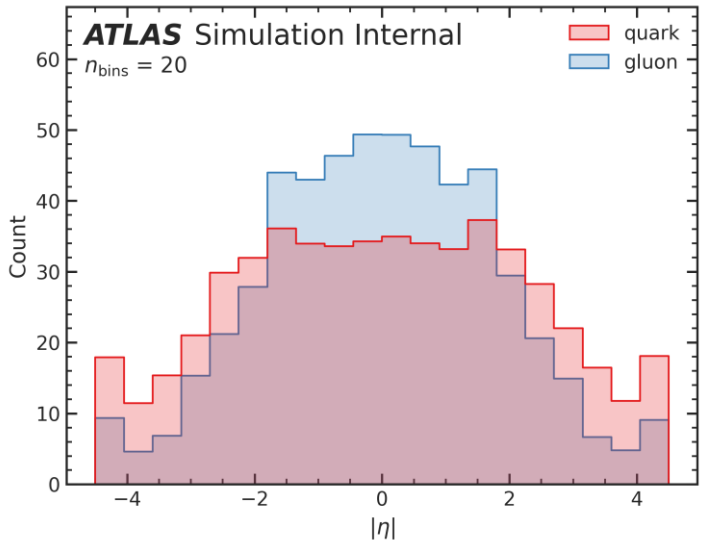
Exploring input variables

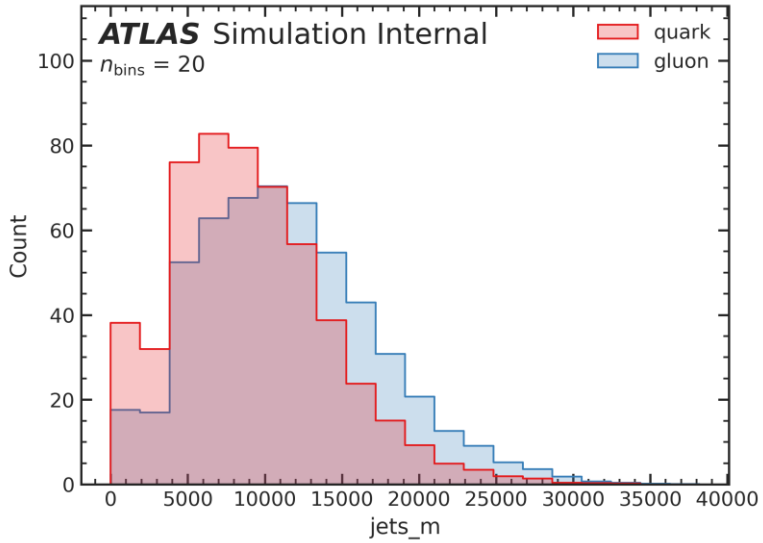
- mc21_14TeV.801168.Py8EG_A14NNPDF23LO_jj_JZ3.





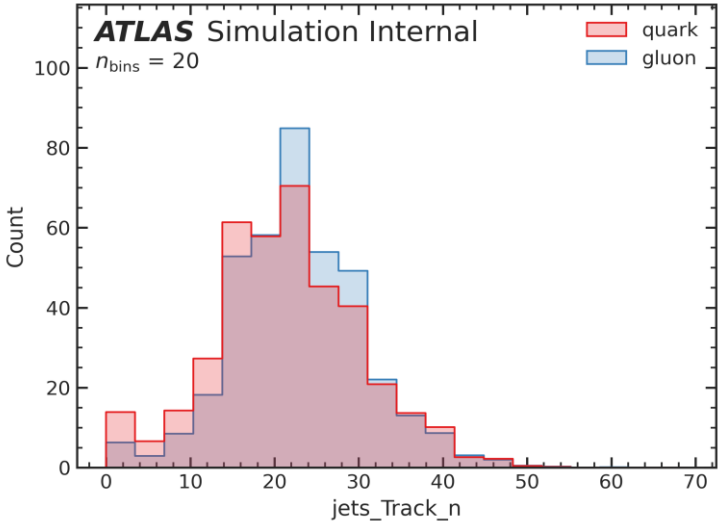


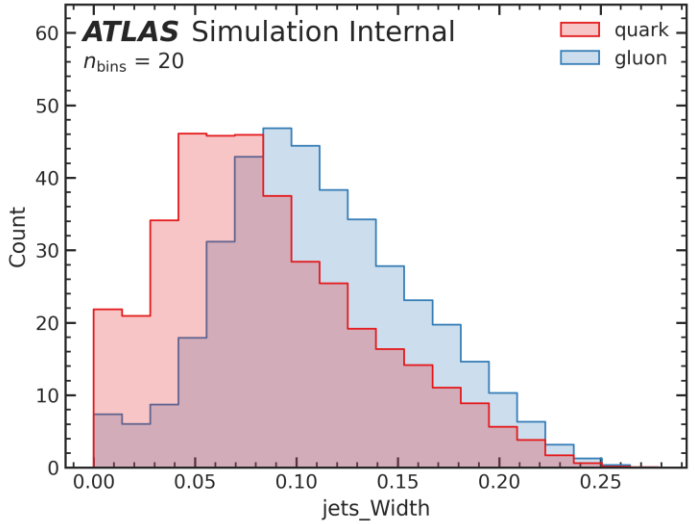


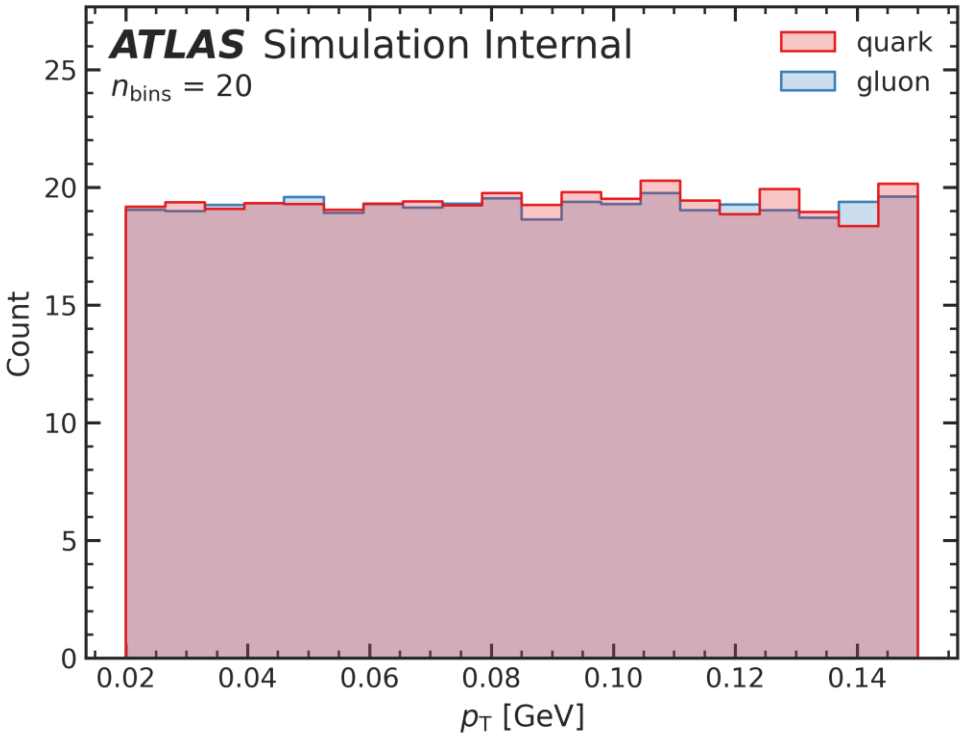


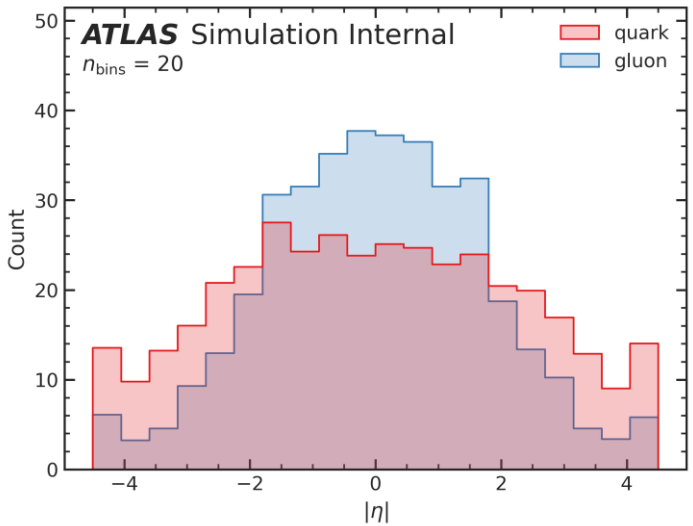
Exploring input variables

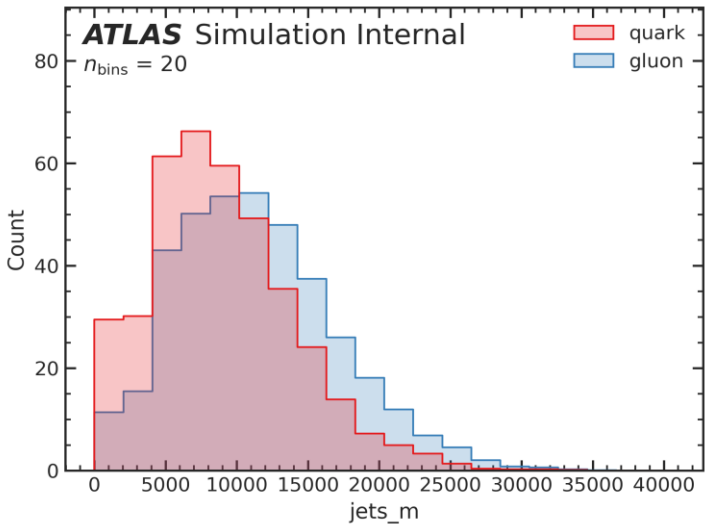
- mc21_14TeV.801169.Py8EG_A14NNPDF23LO_jj_JZ4.





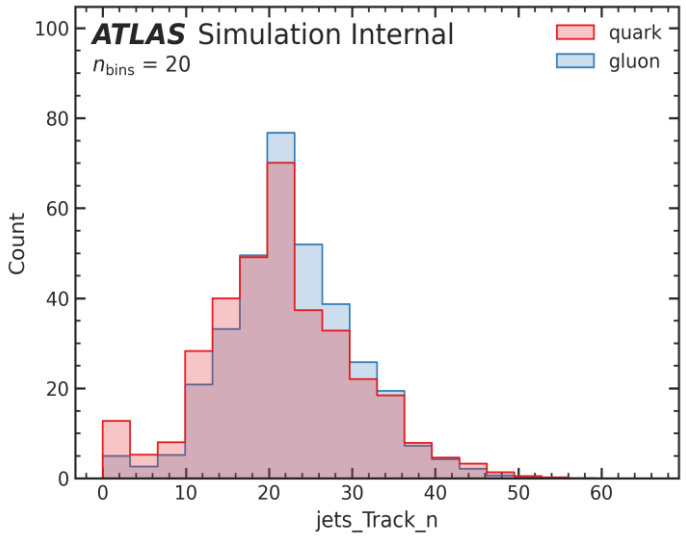


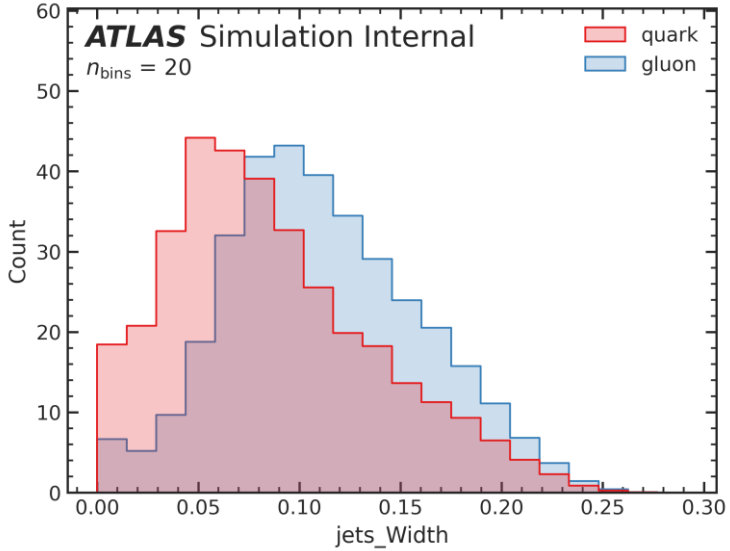


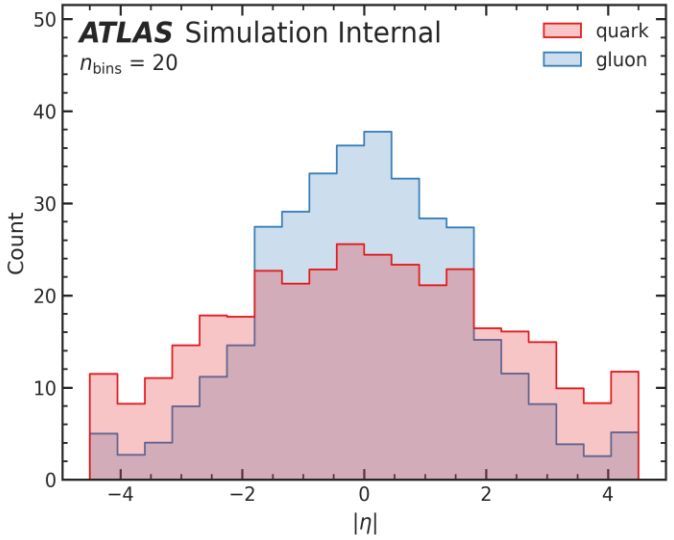


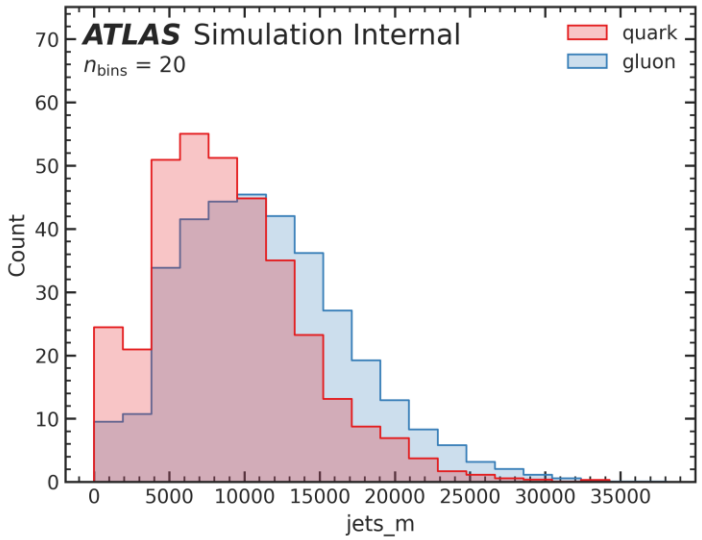
Exploring input variables

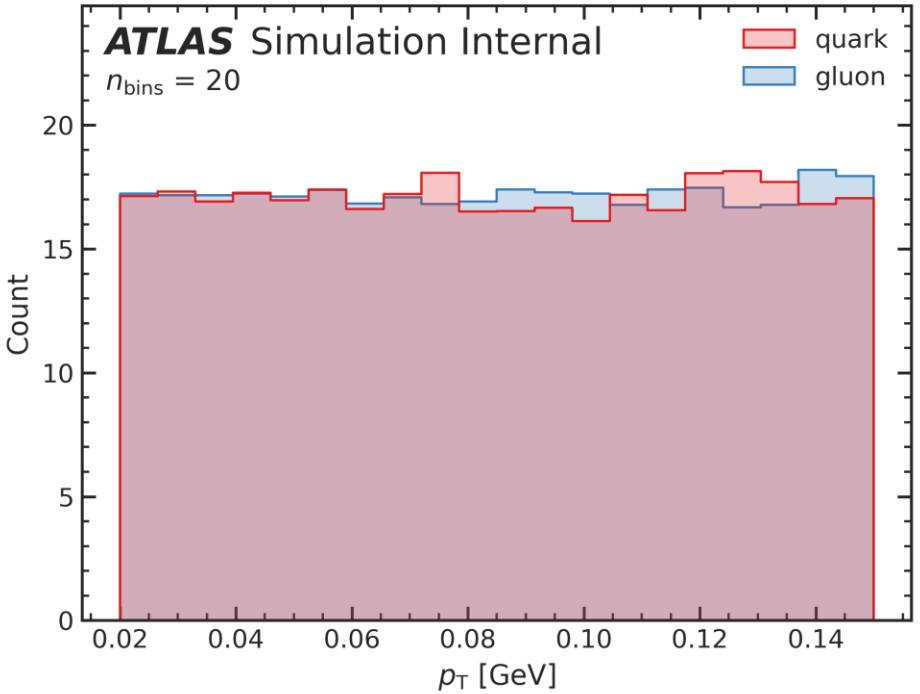
- mc21_14TeV.801170.Py8EG_A14NNPDF23LO_jj_JZ5



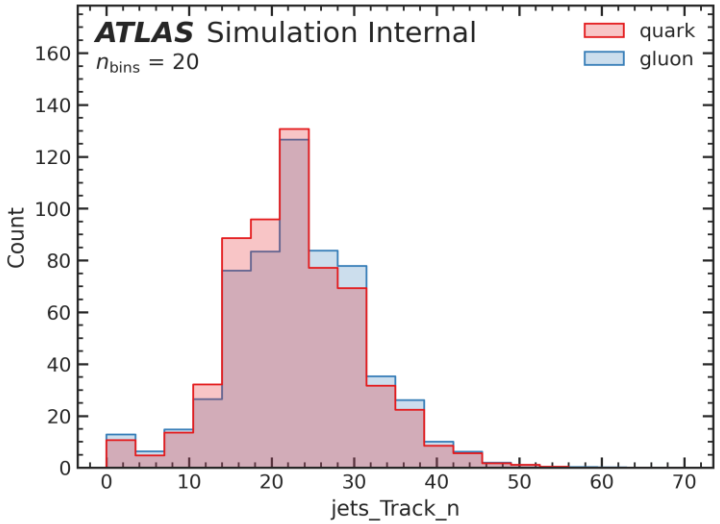


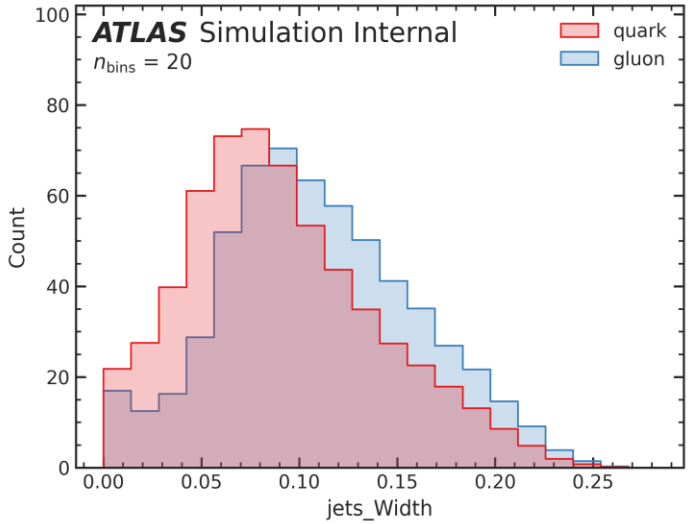


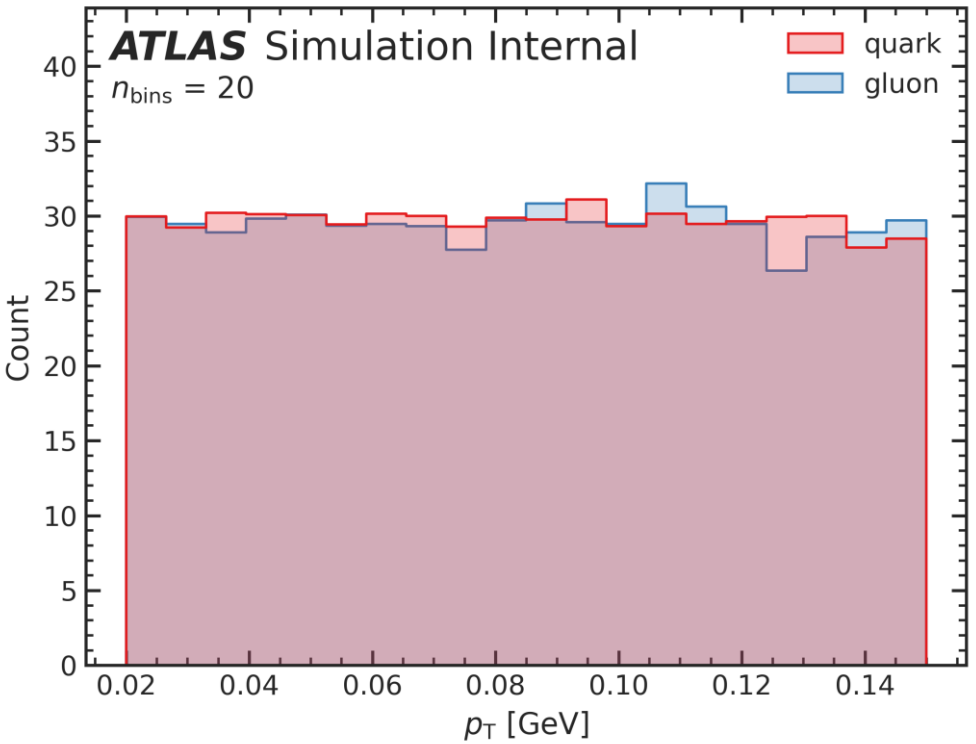


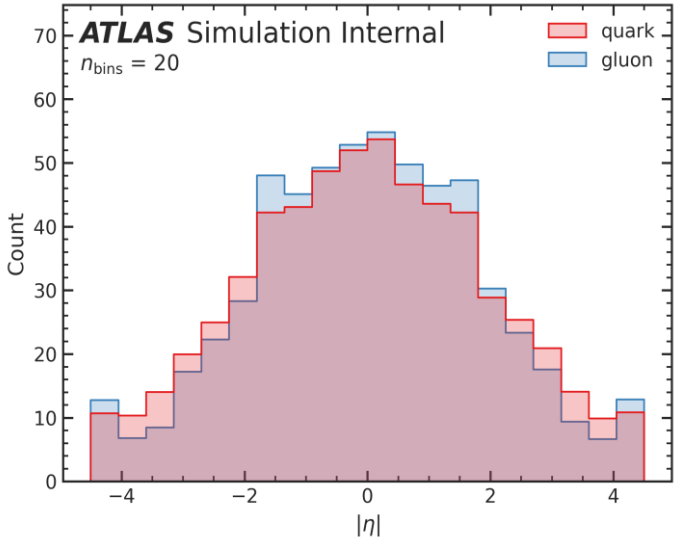


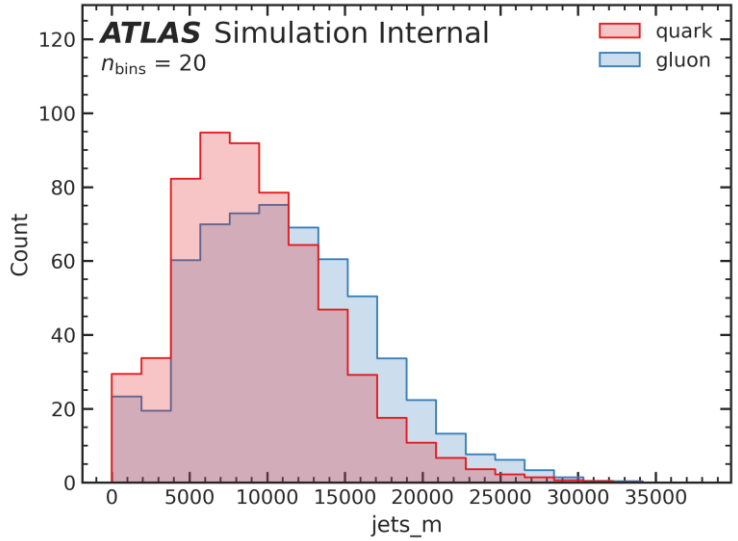
Merging samples (not 801167 (train))



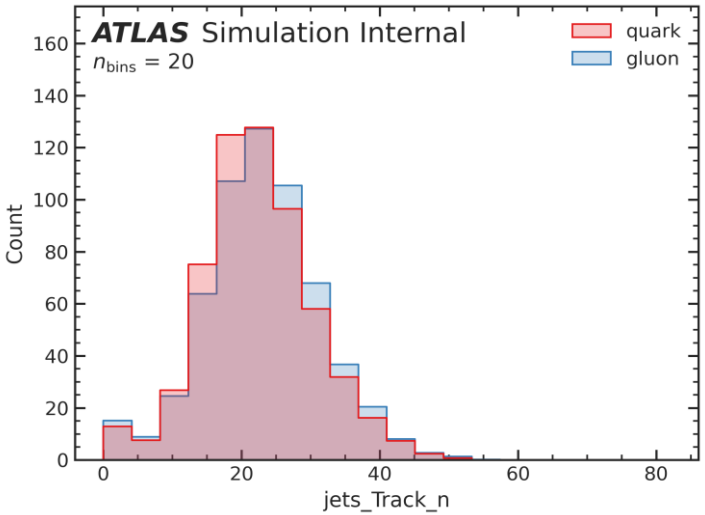


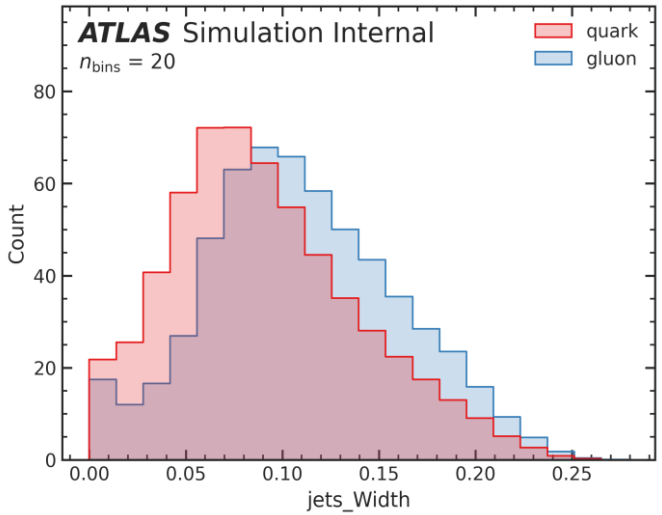


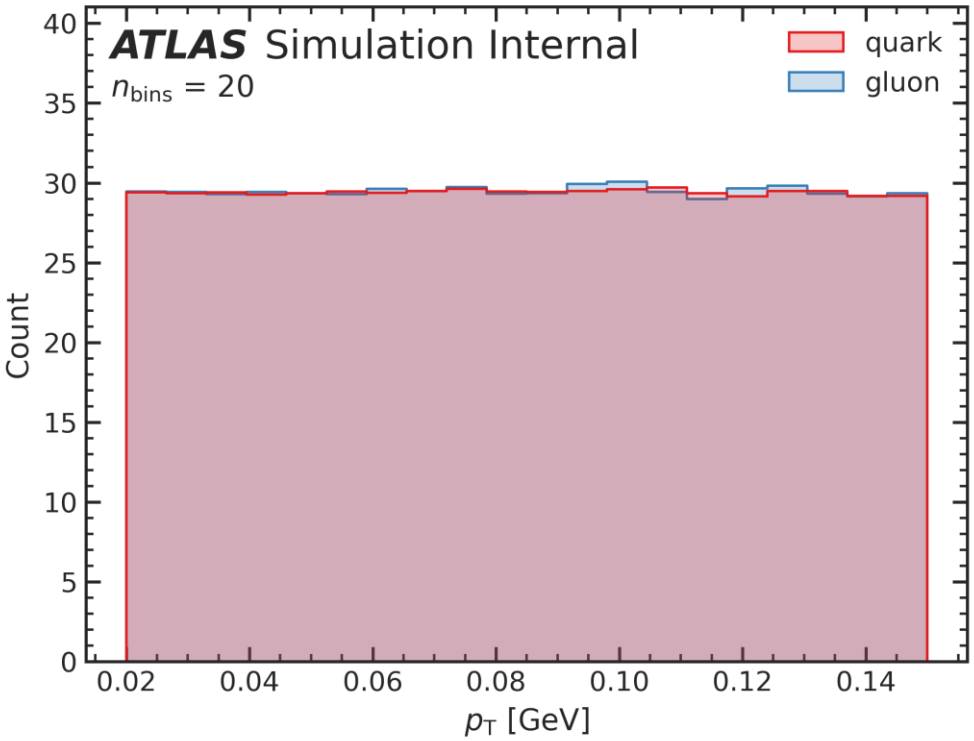


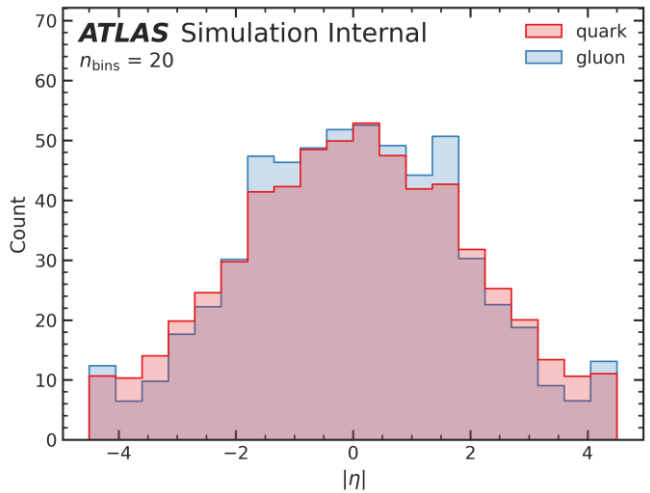


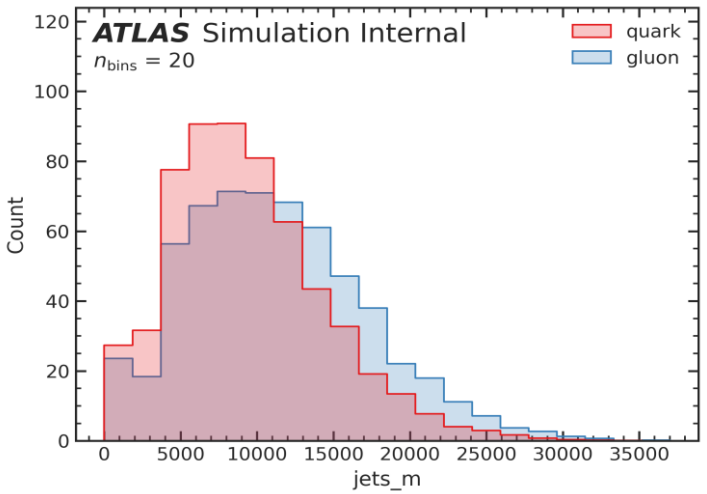
Merging samples (not 801167 (val))



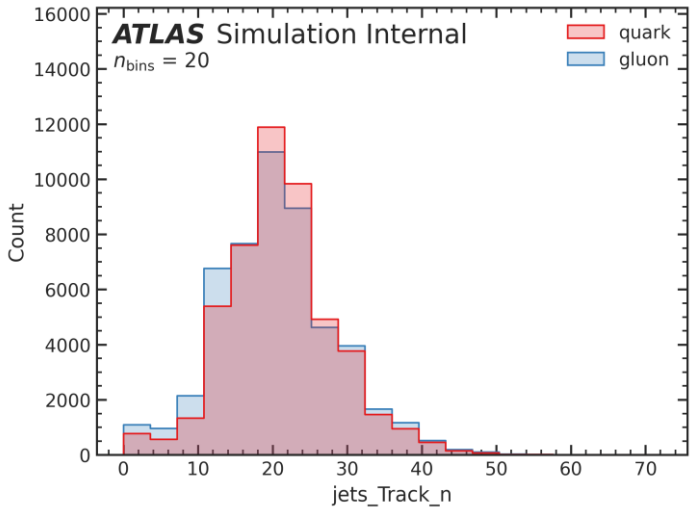


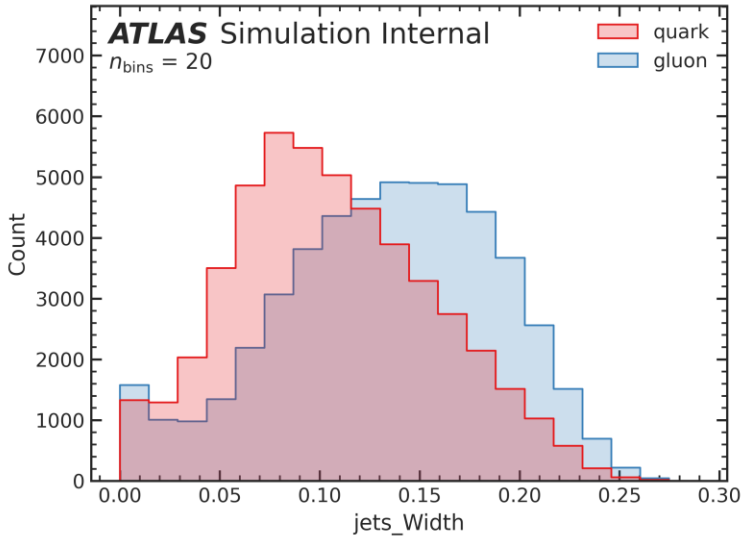


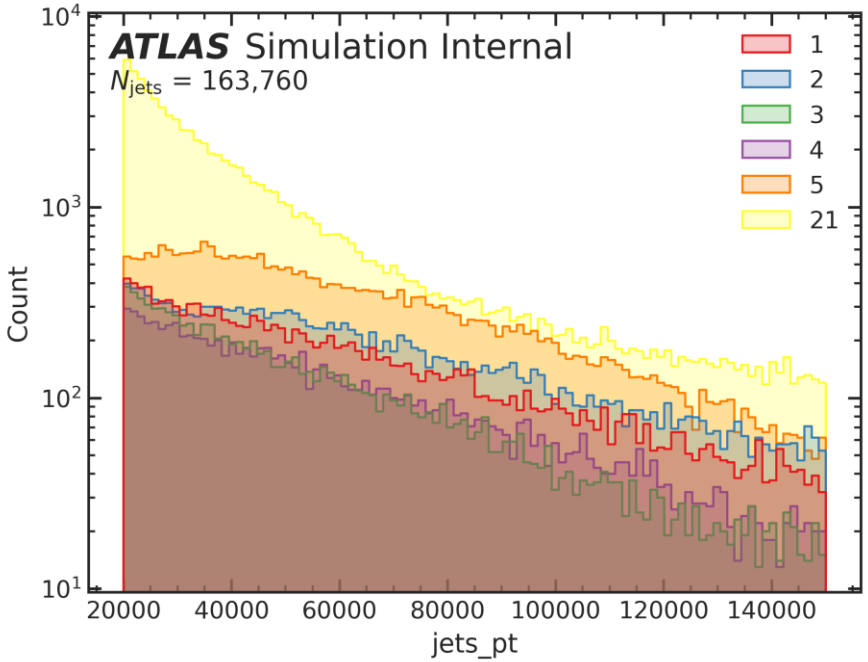


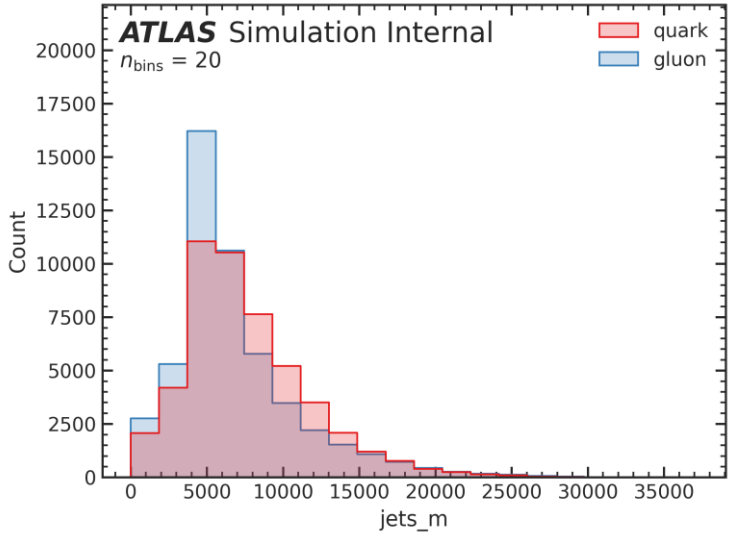


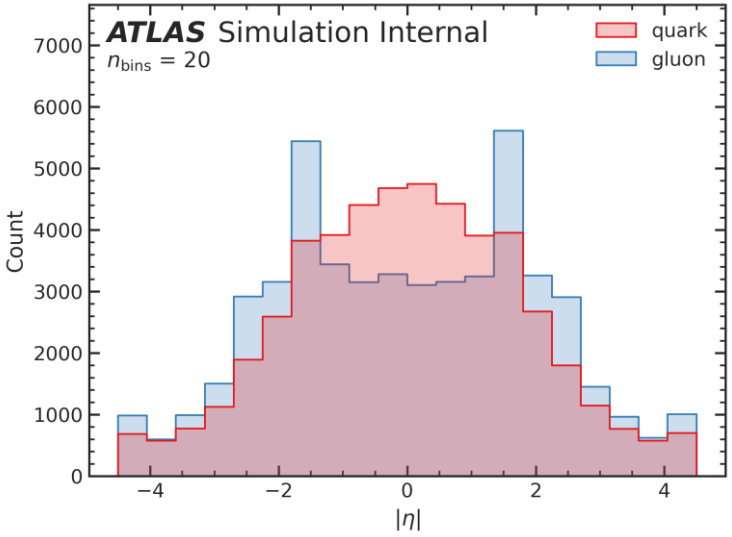
Merging samples (not 801167 (test))

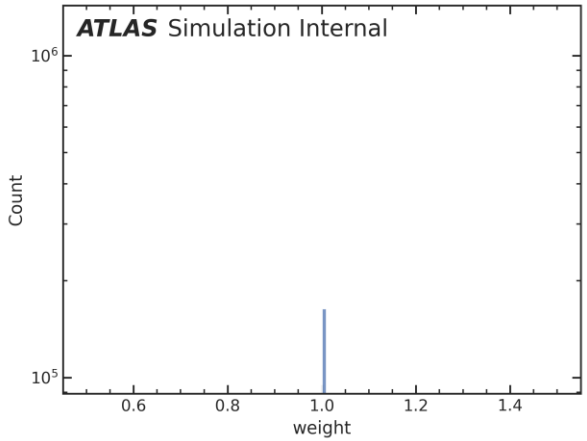












One test, Fully Connected.

4.4 High-level Jet Variables

High-level jet variables are calculated from the jet constituents and describe the jet as a whole. They are used as input to the FC and Highway networks. There are 10 of these variables, forming a 10-dimensional vector as the input to the neural network. They are:

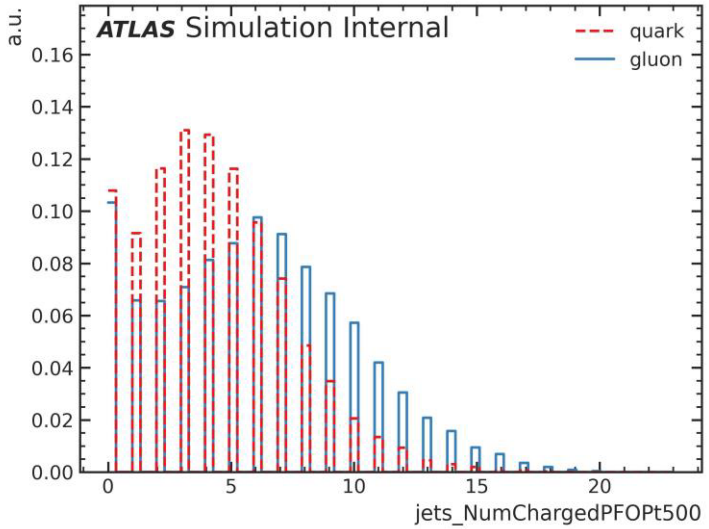
- Calibrated jet four-momentum.
- Number of PFOs.
- Numbers of charged PFOs passing two different pT thresholds (500 and 1000 MeV).
- Jet width [15] calculated using PFOs with $p_{\mathrm{T}} > 1000$ MeV.
- Fraction of the jet energy deposited in the EM calorimeter. Fraction of the jet energy carried by charged particles.

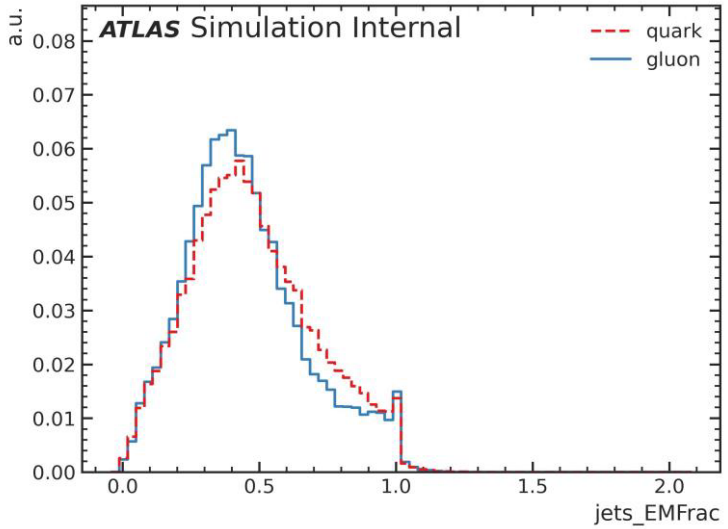
In addition to these variables, a second set of 5 high-level jet variables is employed. These variables correspond to the ones utilized in the prior ATLAS q/g tagger, as detailed in the reference [15]. Concretely, these encompass jet p_{T} and η and the count of PFOs within a jet. The last two variables are the jet width, and the two-point correlation function [13, 16], both calculated using the PFOs.

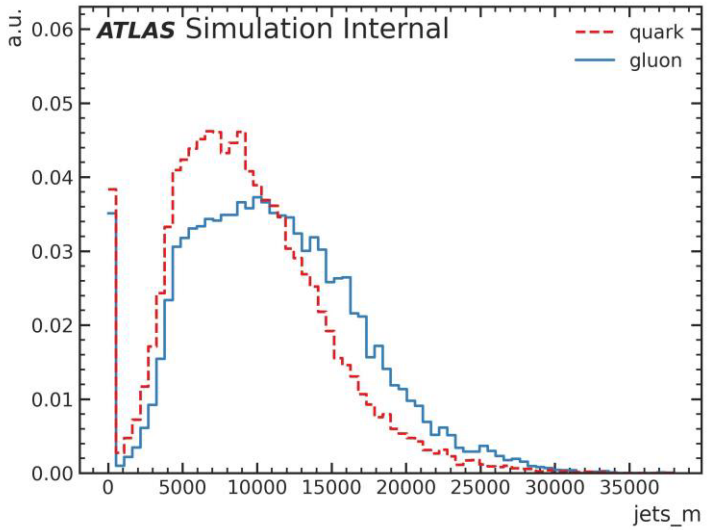
My test--> Goal find a good set

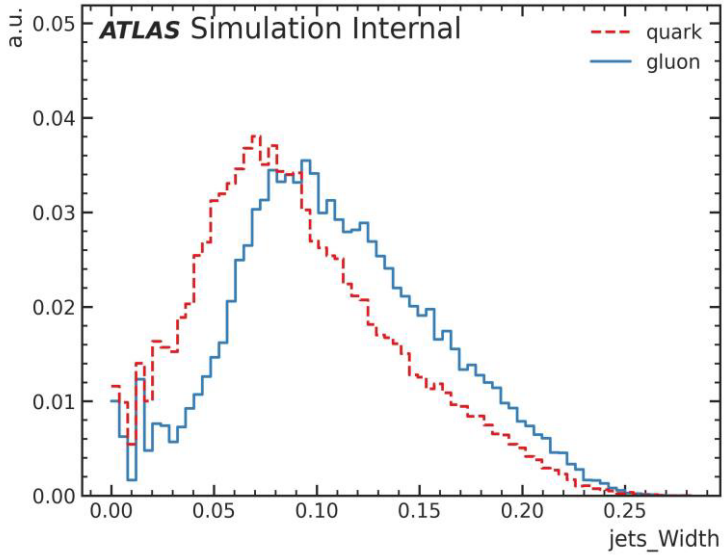
```
def __init__(self, *args, **kwargs):
    super().__init__(*args, **kwargs)
    if self.variables is None:
        self.variables = [
            'jets_EMFrac',
            #'jets_chf',
            'jets_m',
            'jets_phi',
            'jets_pt',
            'jets_Width',
        ]

        self.idxd_variables = [
            'jets_NumChargedPFOPt500',
        ]
```

Distributions for training

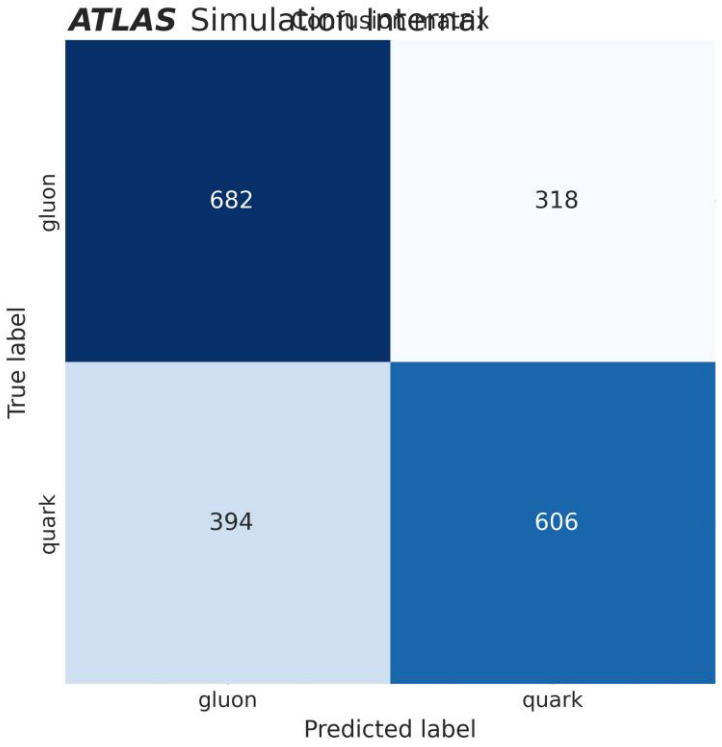


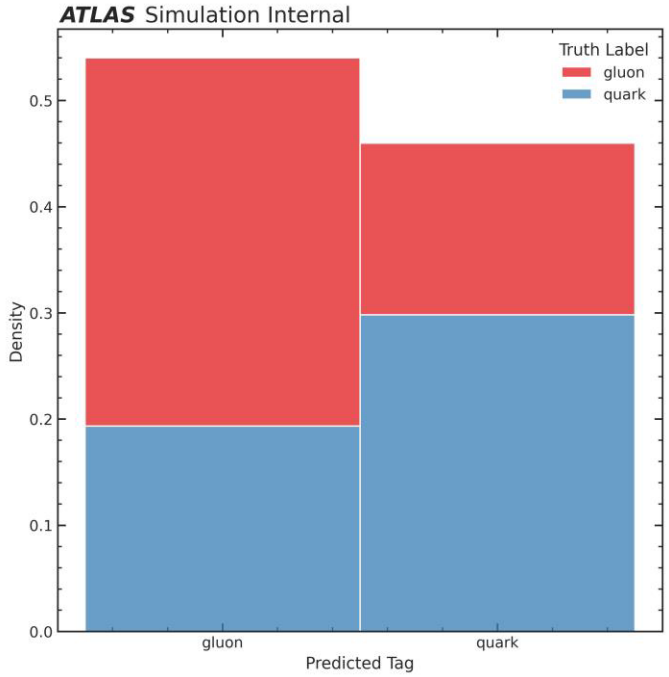


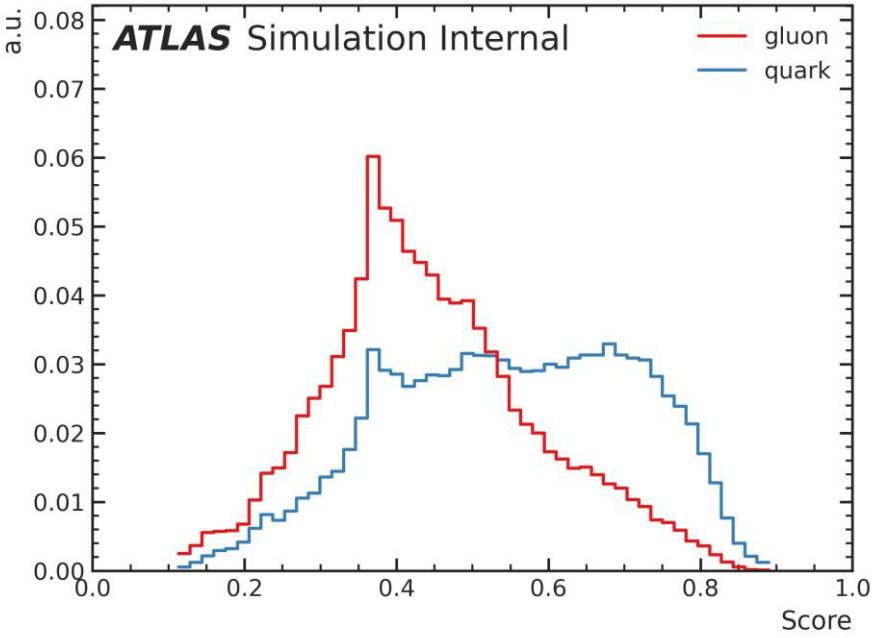




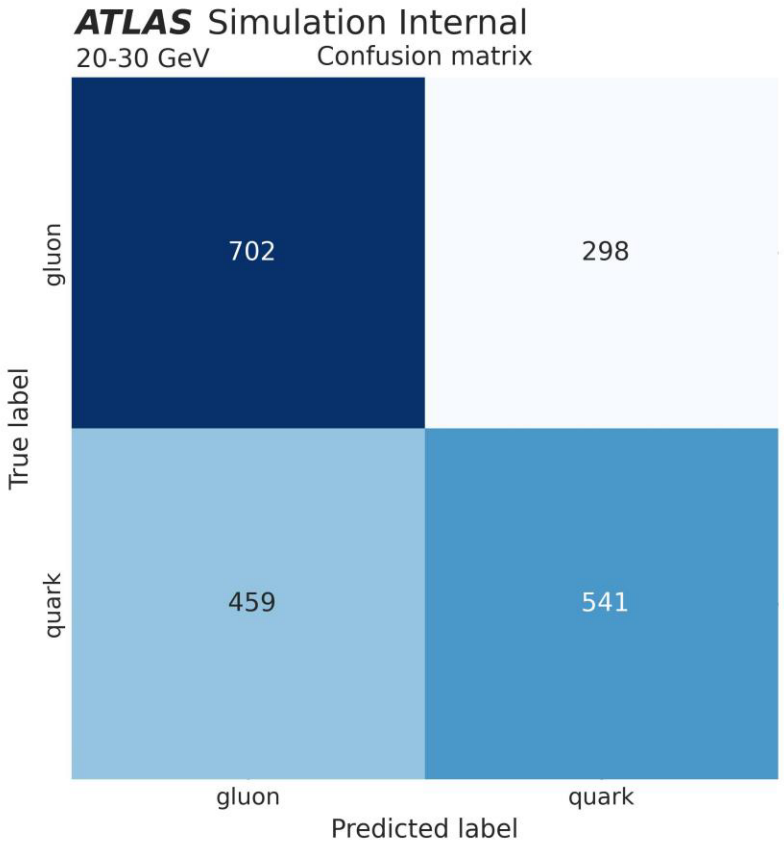
Evaluation (20 GeV to 150 GeV)

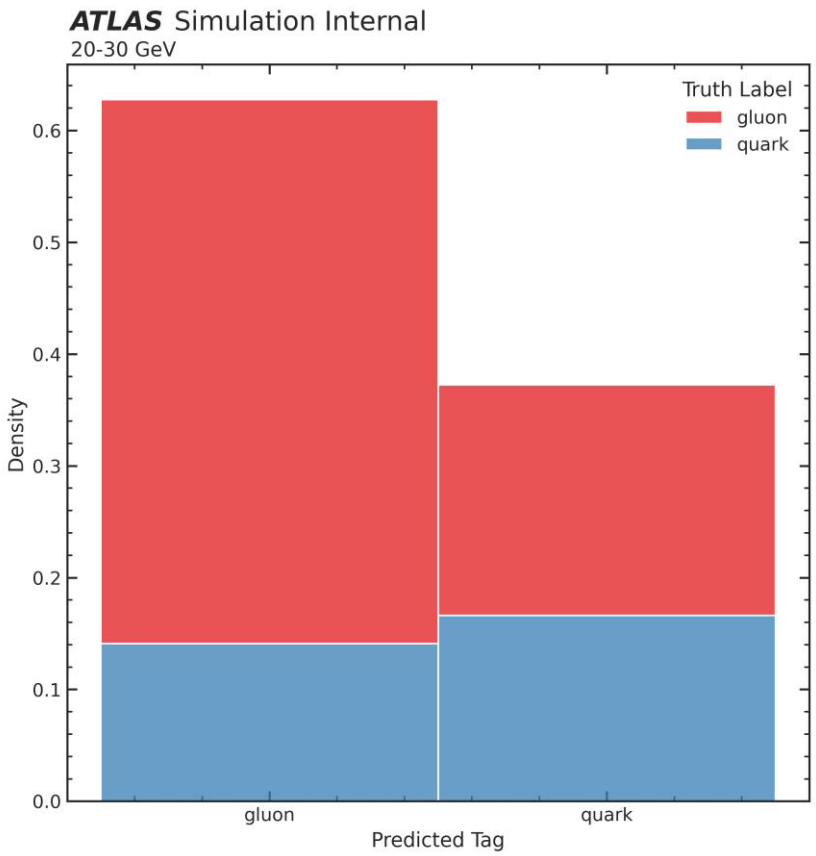


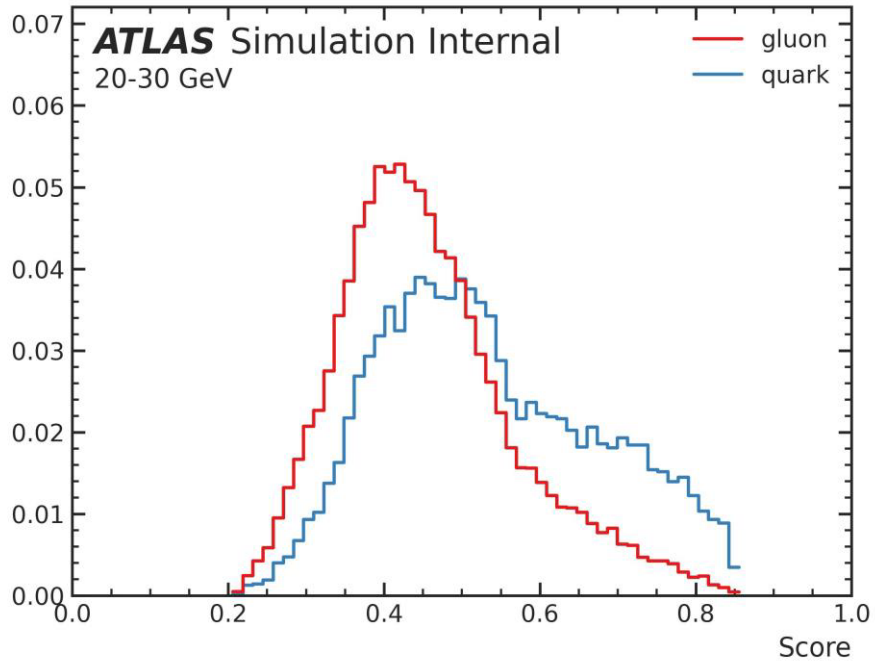




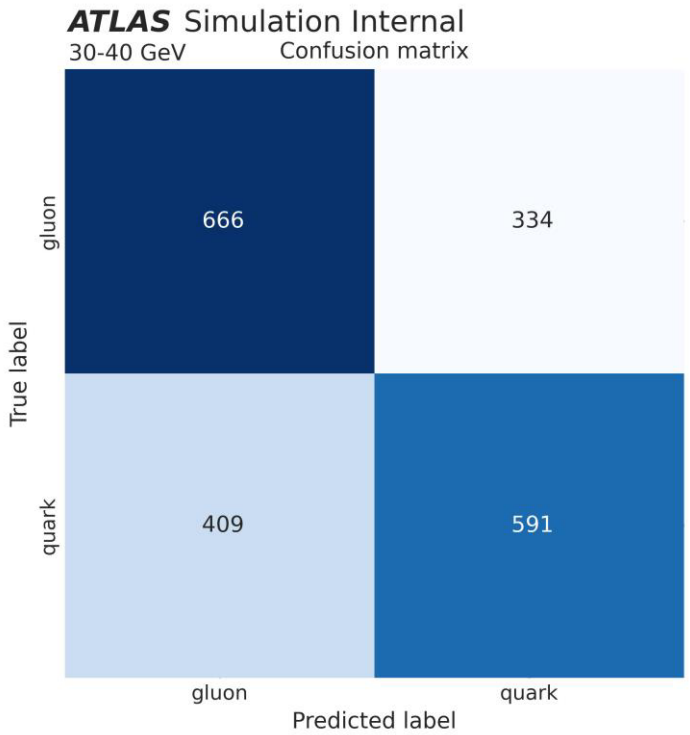
Evaluation 20-30 GeV

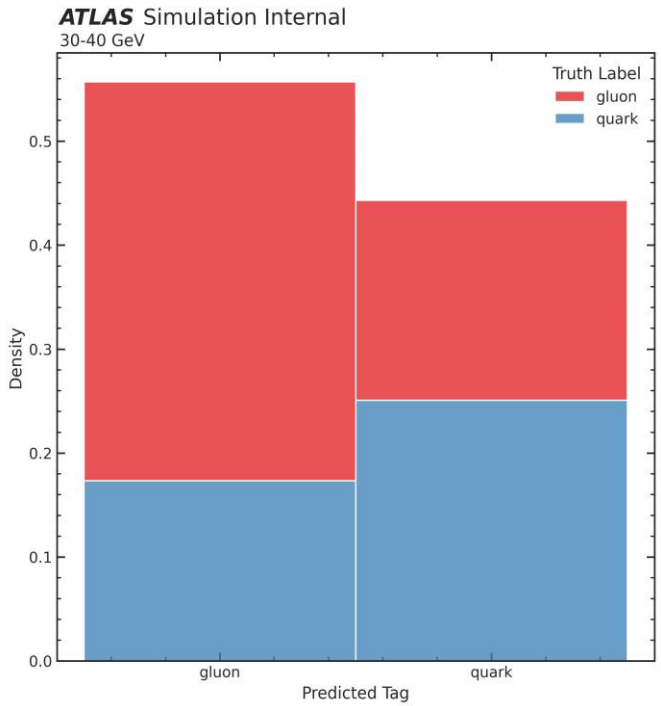


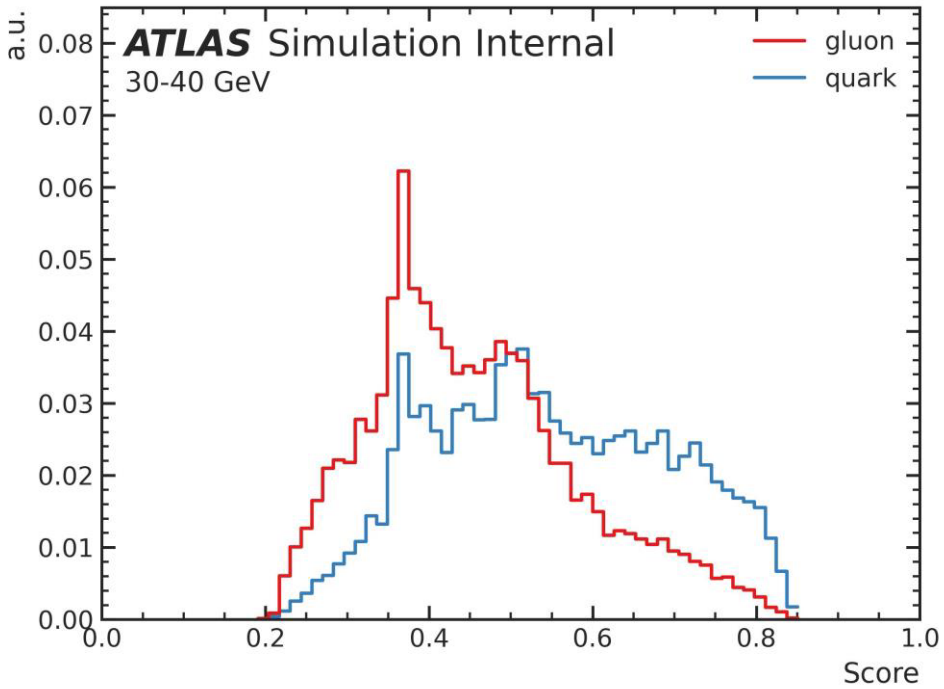




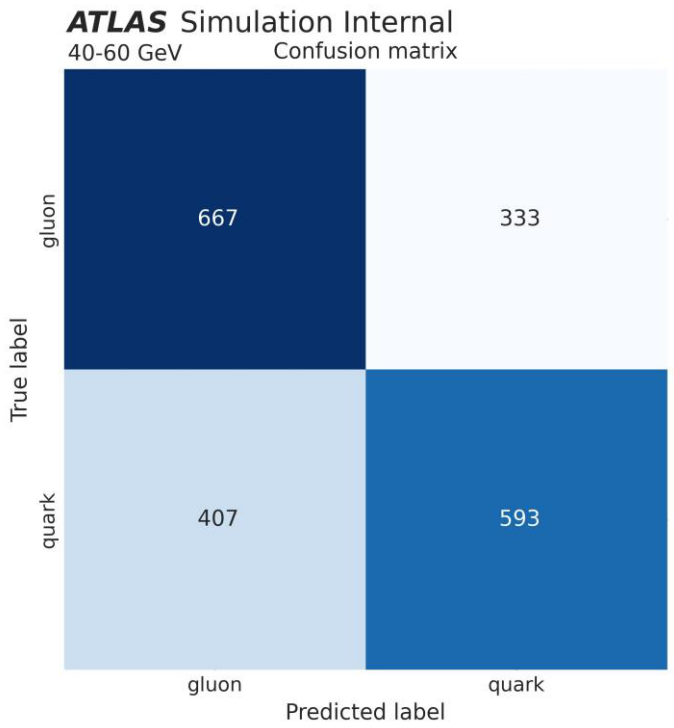
Evaluation 30-40 GeV

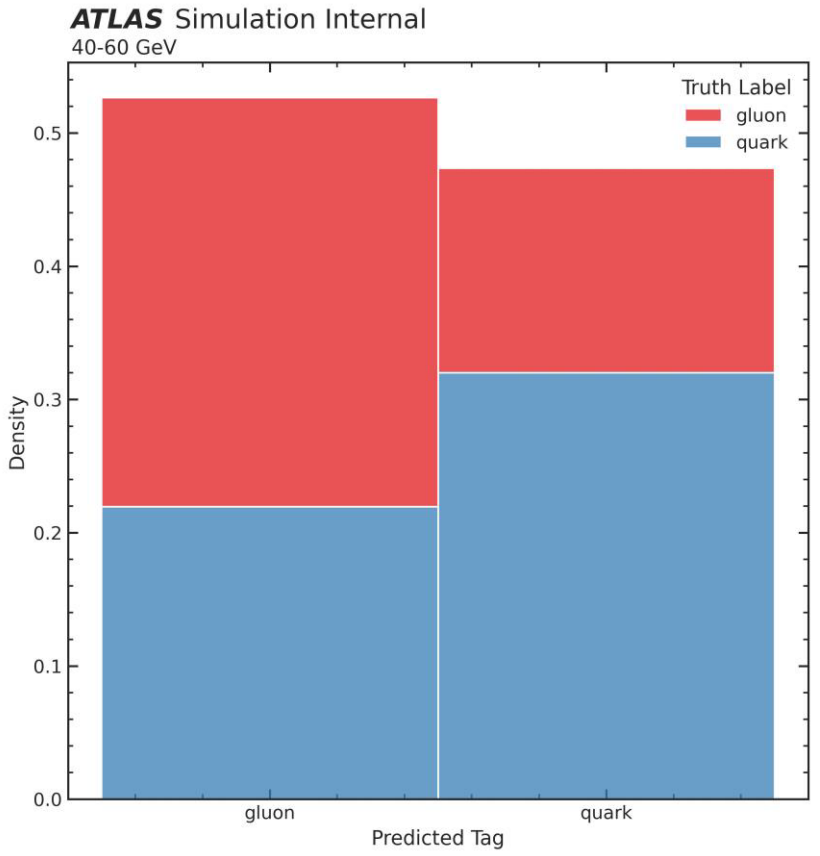


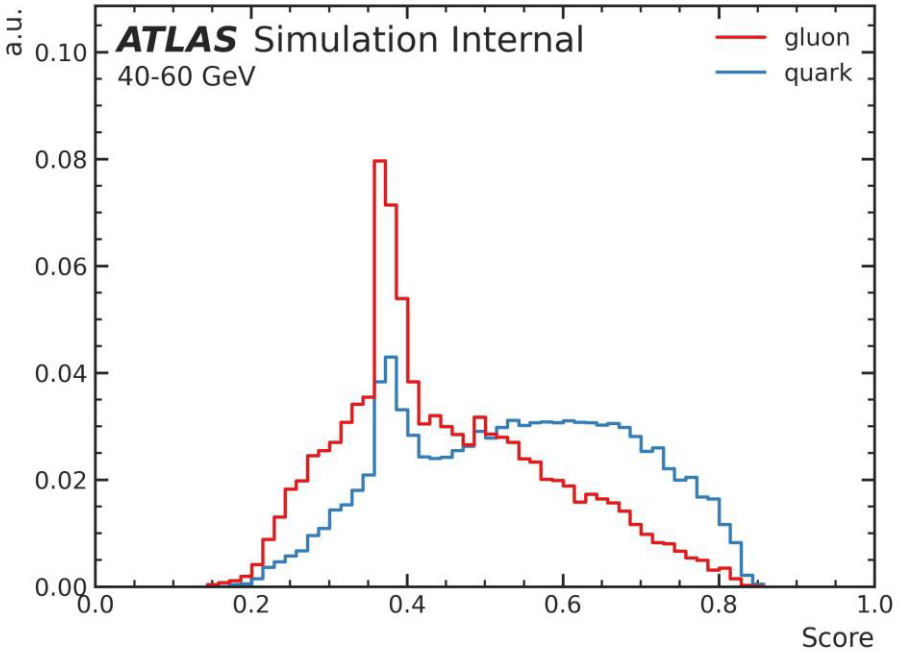




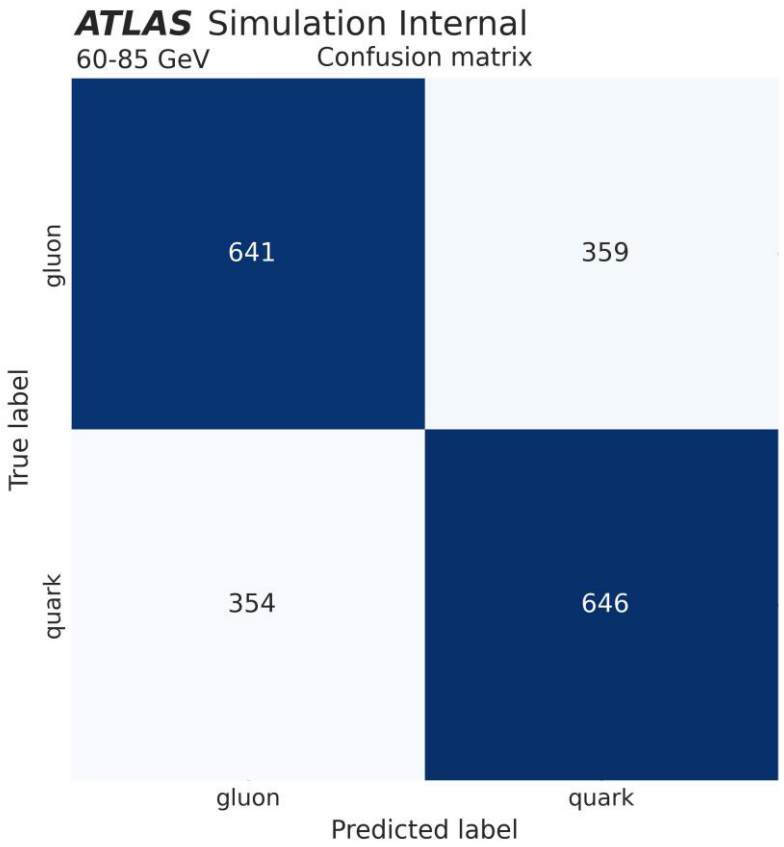
Evaluation 40-60 GeV

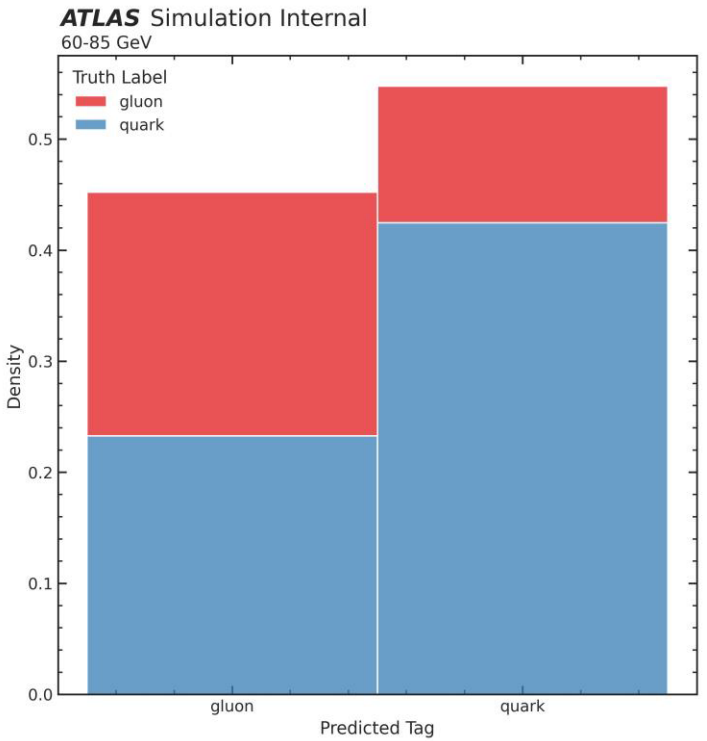


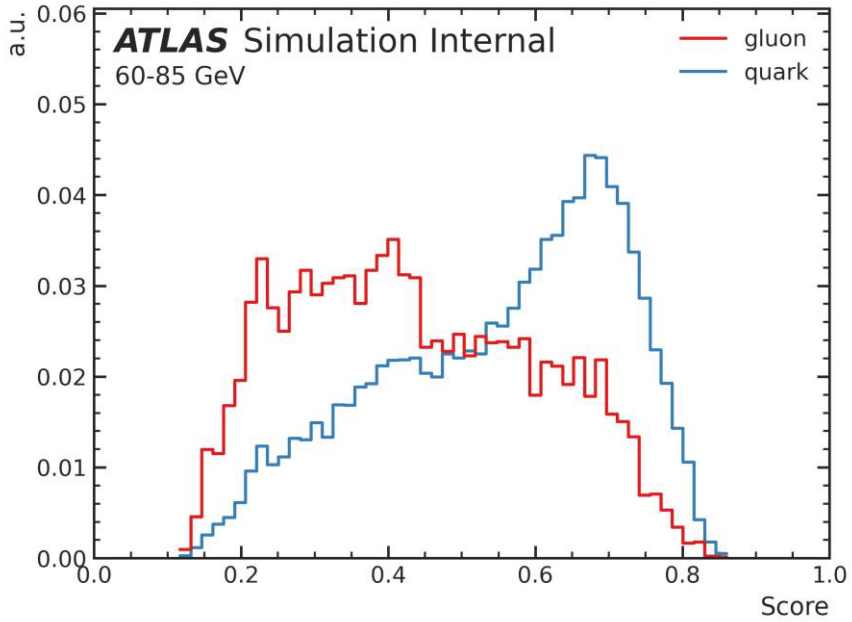




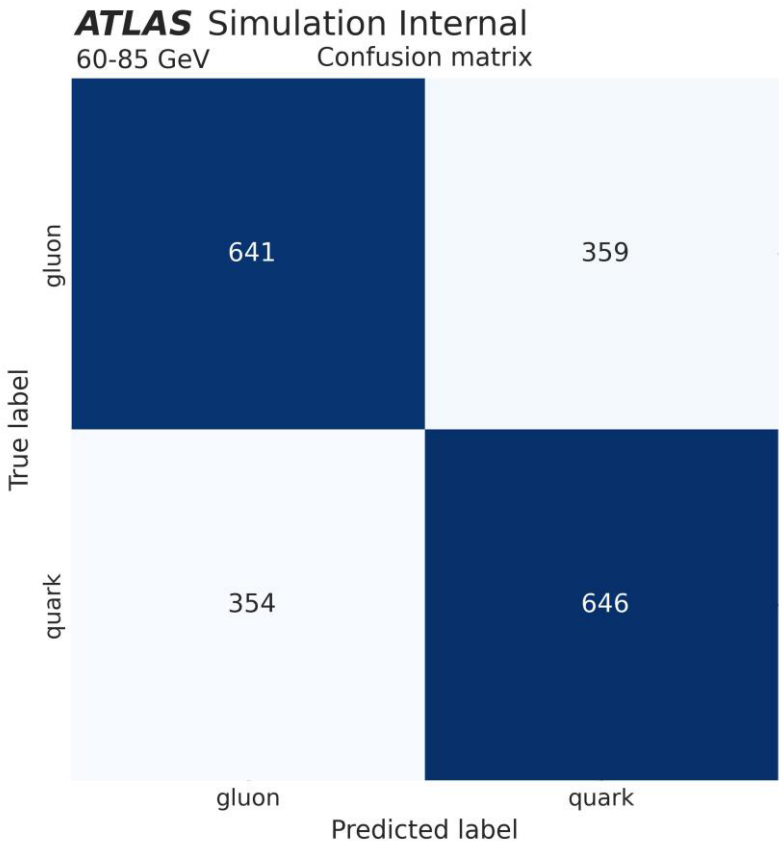
Evaluation 60-85 GeV

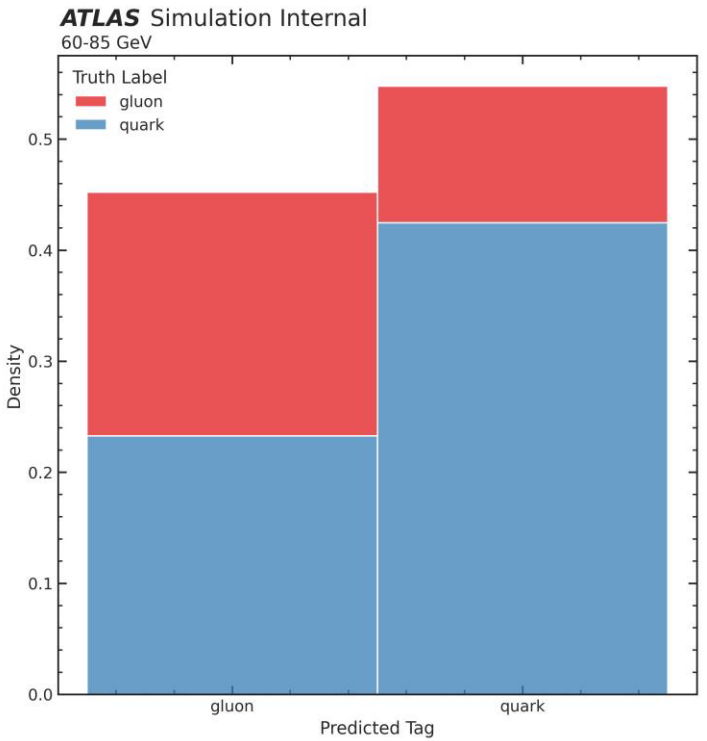


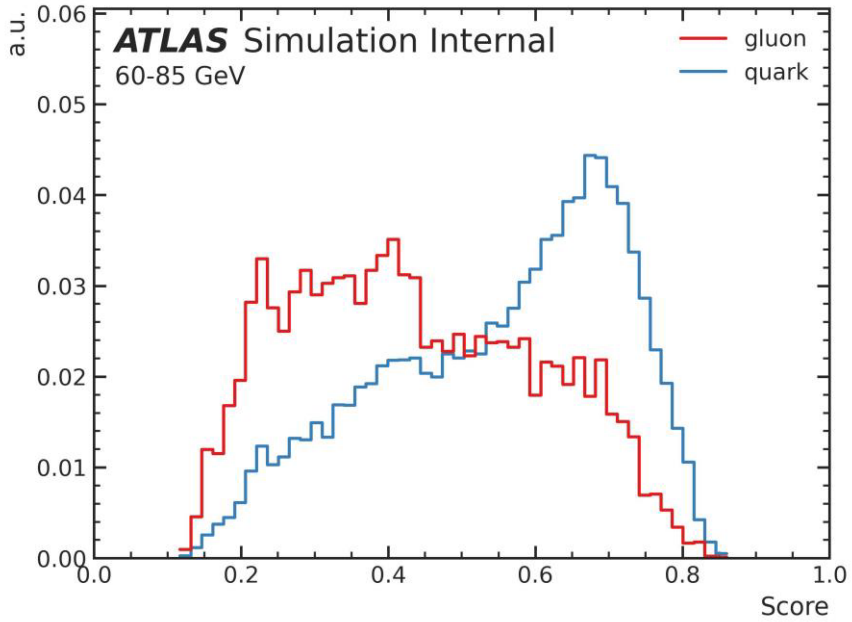




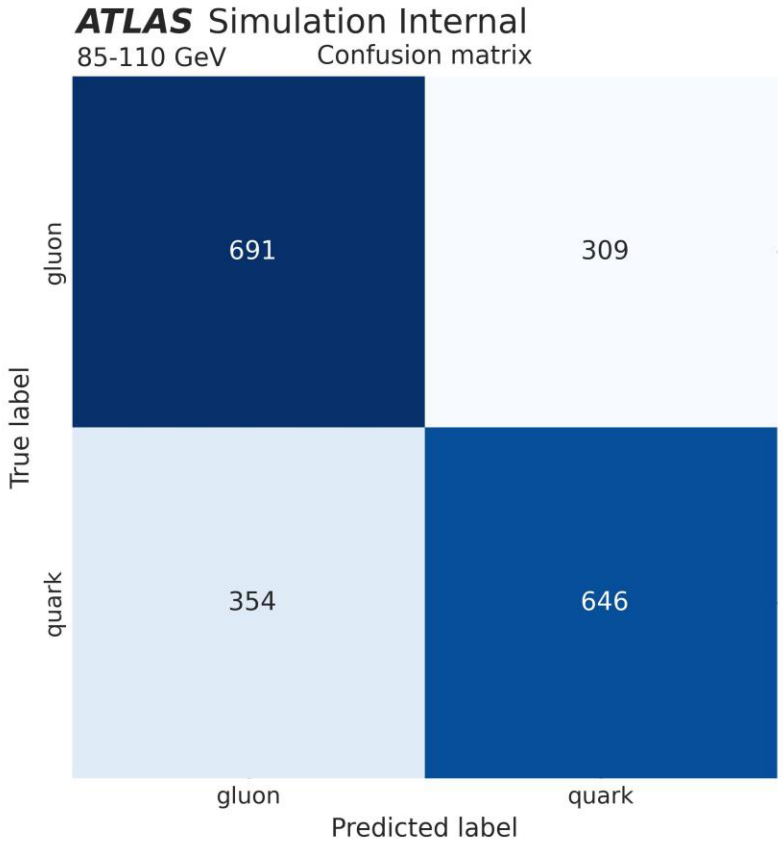
Evaluation 60-85 GeV

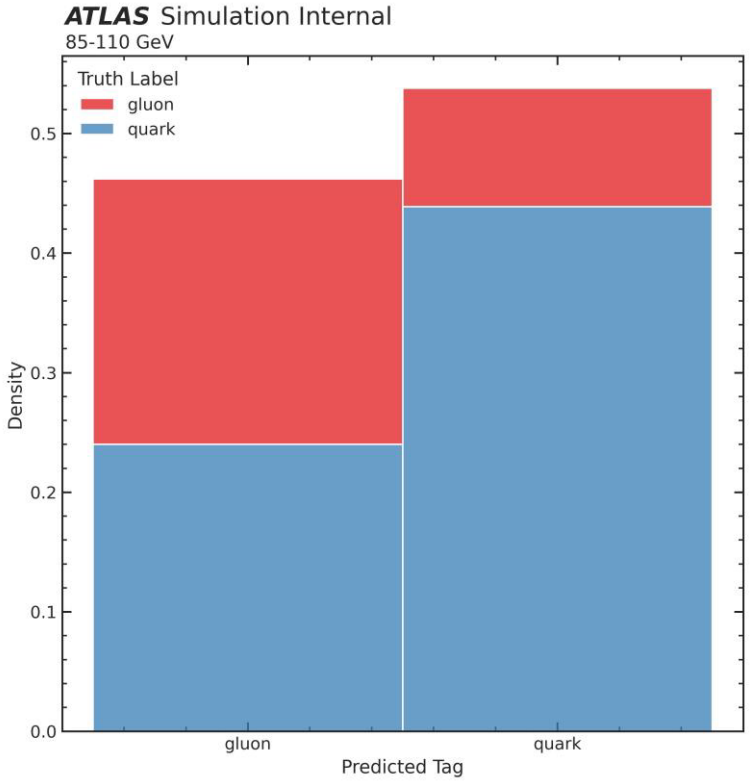


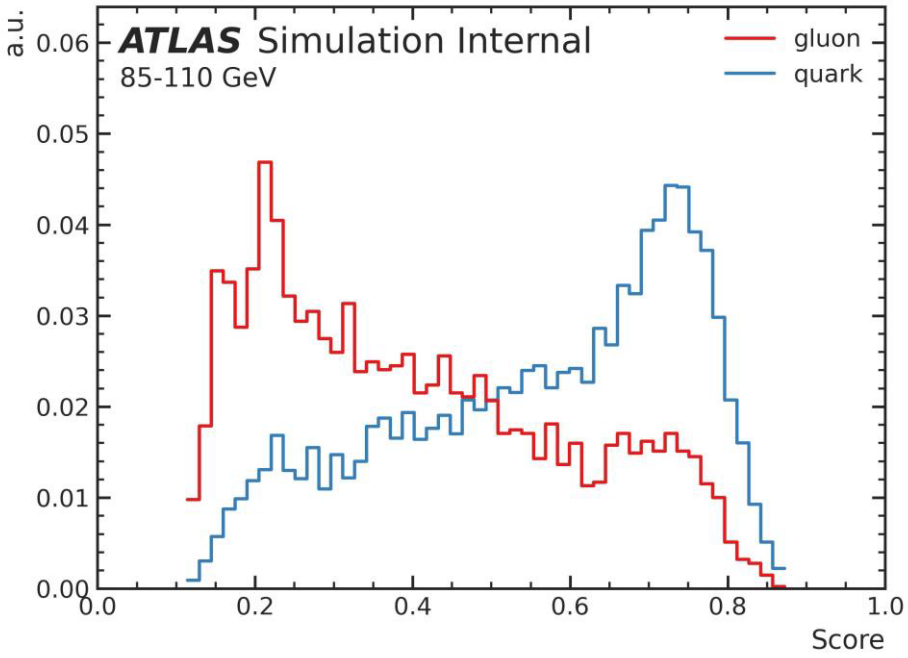




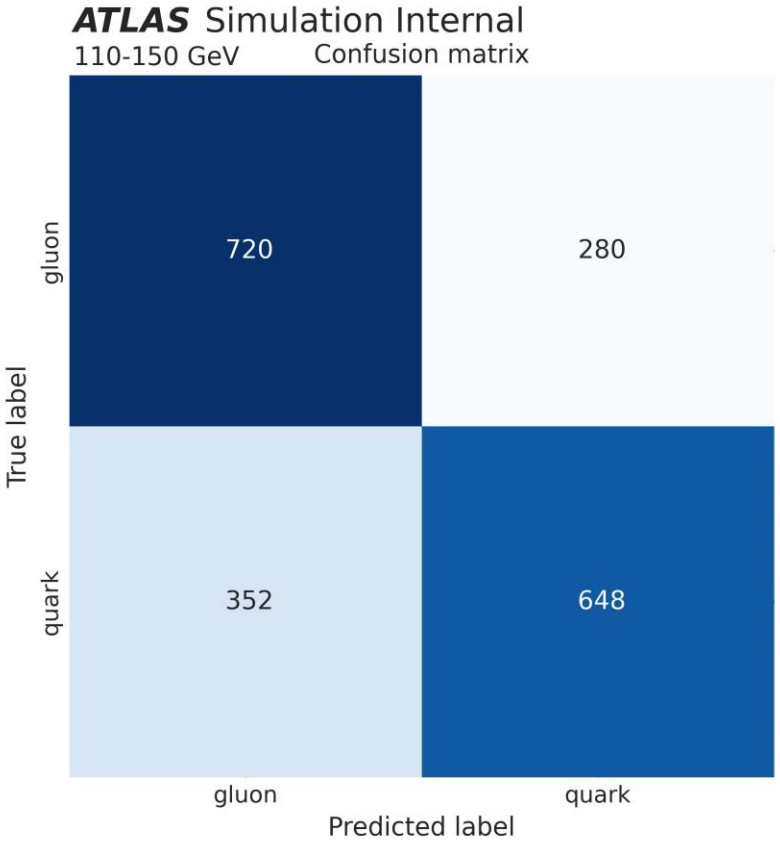
Evaluation 85-110 GeV

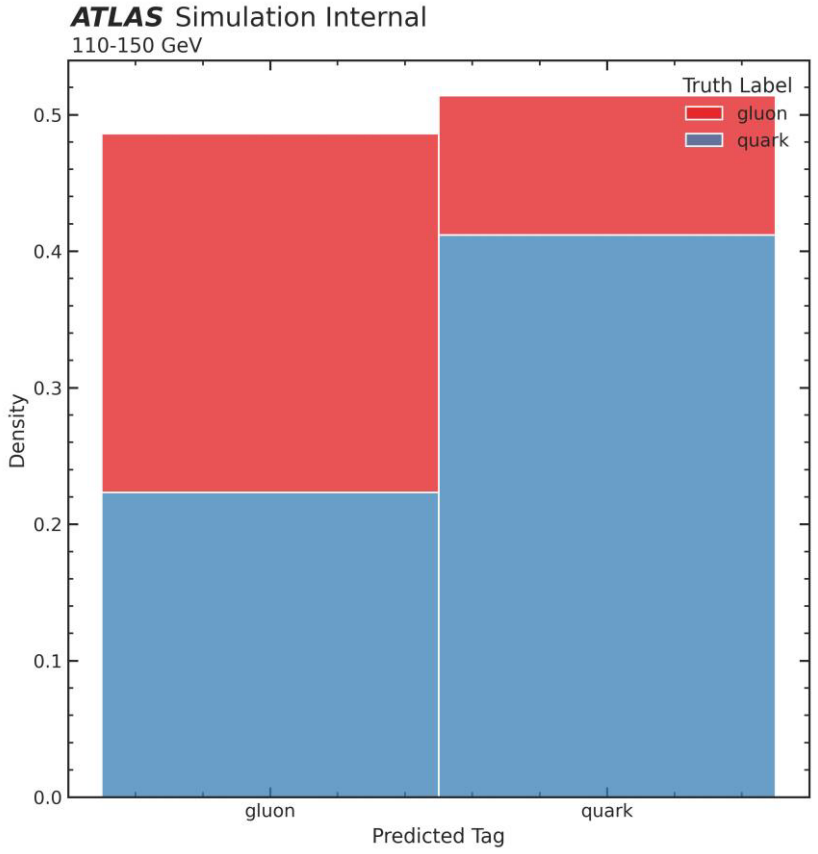


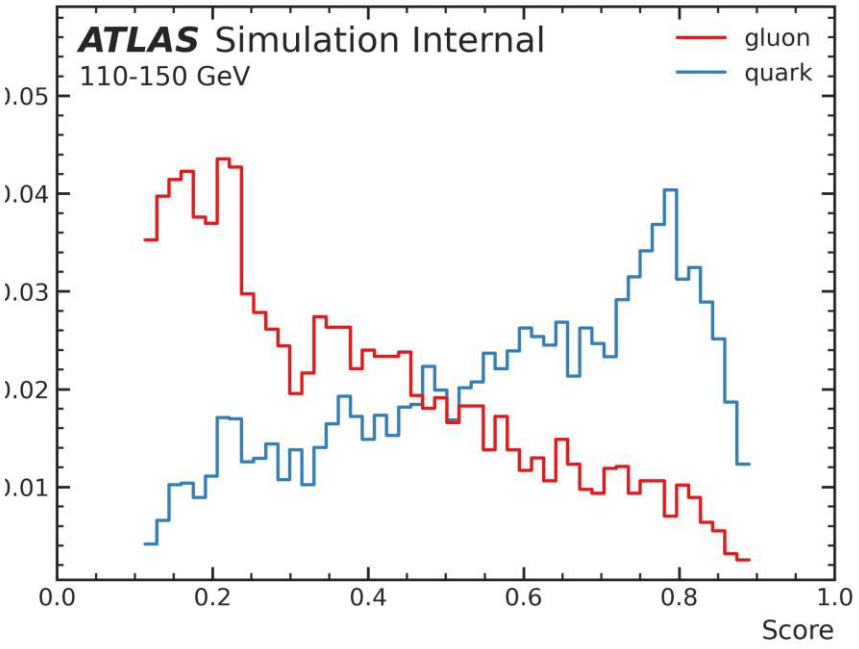




Evaluation 110-150 GeV







Next steps

- Explore new variables (high and low level)
- Add definition
- 2d plot eta and pt vs vars
- Central and forward
- Extend to new samples Zee, Zmumu, ggH
- Feature importance