Quark gluon tag in HL-LHC Constituent-Based Tagger

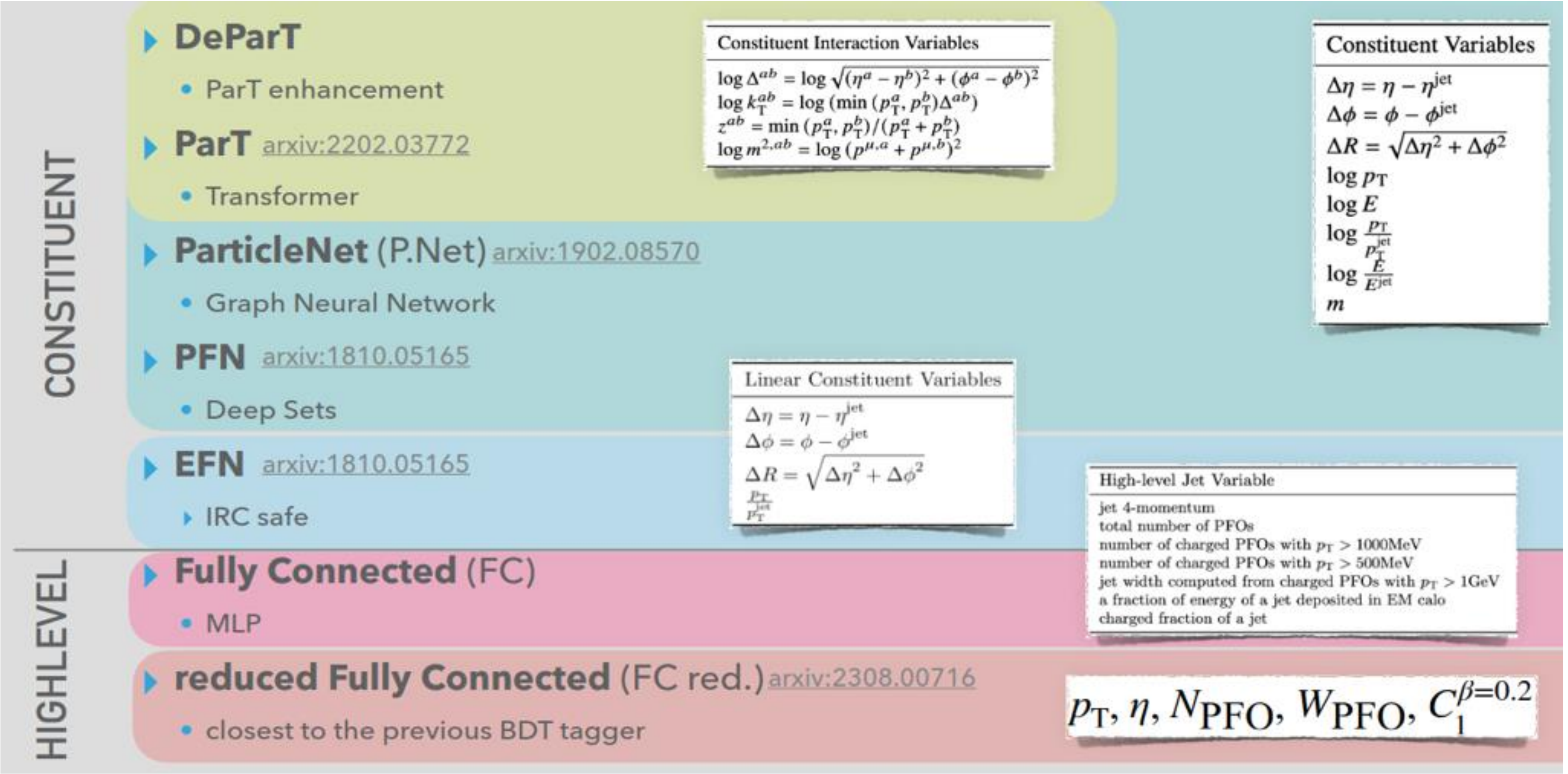
Sabine Elles, Jessica Leveque and Florencia Castillo

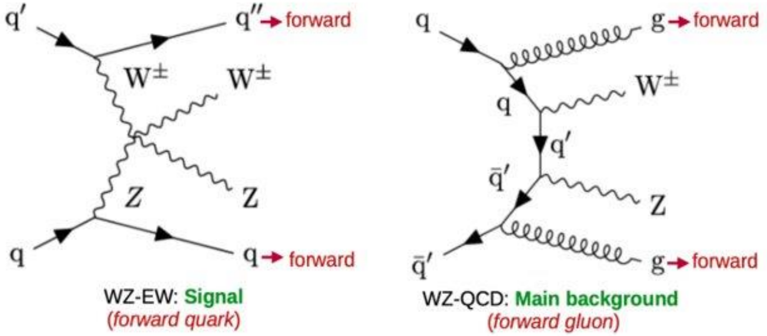
Jet tagging and scale factor meeting

cnrs LAPP Laboratoire d'Annecy-le-Vieux de Physique des Particules anr agence nationale de la recherche ATRAPP

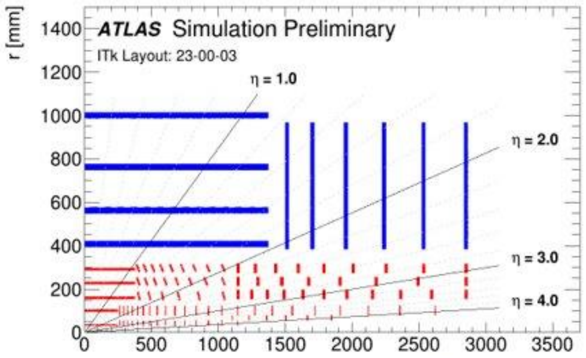
Outlook: quark gluon tag for HL-LHC

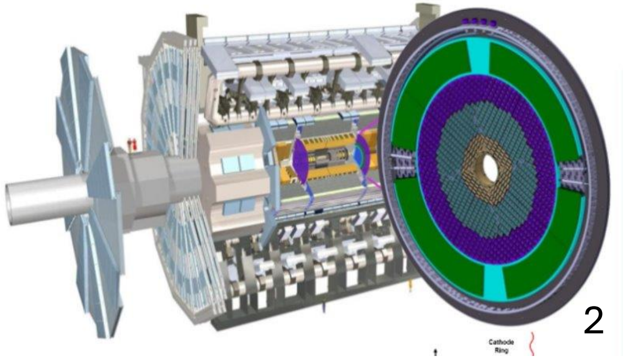
Constituents taggers :





Detector upgrades:





JET Definition and Samples

- **Jets** are reconstructed with the anti-kt with R = 0.4. PFlow jets no Calibrated
- **Cuts**: Only the two jets with the highest pT are keep it. pT > 20 GeV

- PFOs are a collection of **topo-clusters** formed from energy deposits in calorimeter cells and an algorithmic combination of charged-particle **tracks** with those topo-clusters (PFOs as constituents)
- Using official JETM2 production

- **Central Region**: Jets Pt > 20 GeV, Pt < 2.5TeV, |Jets y| < 2.5
- **Forward Region**: Jets Pt > 20 GeV, Pt < 500GeV, |Jets y| > 2.5 and |Jets y| < 4.0
- For training, only jets pt is flatten

- Samples used for training/validation/testing:
  - Dijets, VBF, ttbar(allhad).
  - **PU 0**: Jets for training: 8.6M central, 4M forward
  - **PU 60**: Only evaluation

Taggers from Run 2

Constituent-based

DeParT and ParT

Constituent Interaction Variables
$\log \Delta = \log \sqrt{(\eta^a - \eta^b)^2 + (\phi^a - \phi^b)^2}$
$\log k_\mathrm{T} = \log(\min(p_\mathrm{T}^a, p_\mathrm{T}^b)\Delta)$
$z = \min(p_\mathrm{T}^a, p_\mathrm{T}^b)/(p_\mathrm{T}^a + p_\mathrm{T}^b)$
$\log m^2 = \log(p^{\mu,a} + p^{\mu,b})^2$

Constituent Variables
$\Delta\eta = \eta - \eta^\mathrm{jet}$
$\Delta\phi = \phi - \phi^\mathrm{jet}$
$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$
$\log p_\mathrm{T}$
$\log E$
$\log \frac{p_\mathrm{T}}{p_\mathrm{T}^\mathrm{jet}}$
$\log \frac{E}{E^\mathrm{jet}}$
m

Highlevel variables

FC Network

Highway Network

EMFrac	+ Jet pT
Jet width	
TrackWidthPt1000	
NumChargedPFOWidthPt1000	
chf	

FC Crafted*

$$N_\mathrm{PF0} = \sum_{\mathrm{PFO}\in\mathrm{jet}}$$

+ Jet pT

$$C1^{\beta=0.2} = \frac{\sum_{i,j\in jet}^{i\neq j} p_{T,i} p_{T,j} (\Delta R_{i,j})^{\beta=0.2}}{(\sum_{i,j\in jet} p_T^{PFO})^2}$$

$$w^{PFO} = \frac{\sum_{\mathrm{PFO}\in\mathrm{jet}} p_T^{PFO} \cdot \Delta R_{PFO,jet}}{\sum_{\mathrm{PFO}\in\mathrm{jet}} p_T^{PFO}}$$

*FC crafted mimic Run 2 BDT tagger

Constituents defined by jets_y for HL-LHC

Constituent-based

DeParT and ParT

Constituent Interaction Variables
$\log\Delta = \log\sqrt{(y^a - y^b)^2 + (\phi^a - \phi^b)^2}$
$\log k_T = \log\left(\min\left(p_T^a, p_T^b\right)\Delta\right)$
$z = \min\left(p_T^a, p_T^b\right)/(p_T^a + p_T^b)$
$\log m^2 = \log\left(p^{\mu,a} + p^{\mu,b}\right)^2$

Constituent Variables
$\Delta y = y - y_{\text{jet}}$
$\Delta\phi = \phi - \phi_{\text{jet}}$
$\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2}$
$\log p_T$
$\log E$
$\log\frac{p_T}{p_{T,\text{jet}}}$
$\log\frac{E}{E_{\text{jet}}}$
m

Highlevel variables

FC Network
Highway Network

EMFrac
Jet width
TrackWidthPt1000
NumChargedPFOWidthPt1000
chf

+ Jet pT

FC Crafted*

$$N_{\text{PF0}} = \sum_{\text{PFO}\in\text{jet}}$$

+ Jet pT

$$C1^{\beta=0.2} = \frac{\sum_{i,j\in jet}^{i\neq j} p_{T,i}p_{T,j}(\Delta R_{i,j})^{\beta=0.2}}{(\sum_{i,j\in jet} p_T^{PFO})^2}$$

$$w^{PFO} = \frac{\sum_{\text{PFO}\in\text{jet}} p_T^{PFO}\cdot\Delta R_{PFO,jet}}{\sum_{\text{PFO}\in\text{jet}} p_T^{PFO}}$$

*FC crafted mimic Run 2 BDT tagger

Evaluation. Central region.

Bins pT: (20, 60, 100, 150,500, 1000,2500) [GeV]

- Central region: jets_y < 2.5. Evaluation in jets_y: [0, 0.5 ,1.0, 1.5, 2.0, 2.5]

- For the central region at low pT , we don't see much improvement when using transform-based (low-level variable) taggers compared to NN-based (high-level variable) taggers.

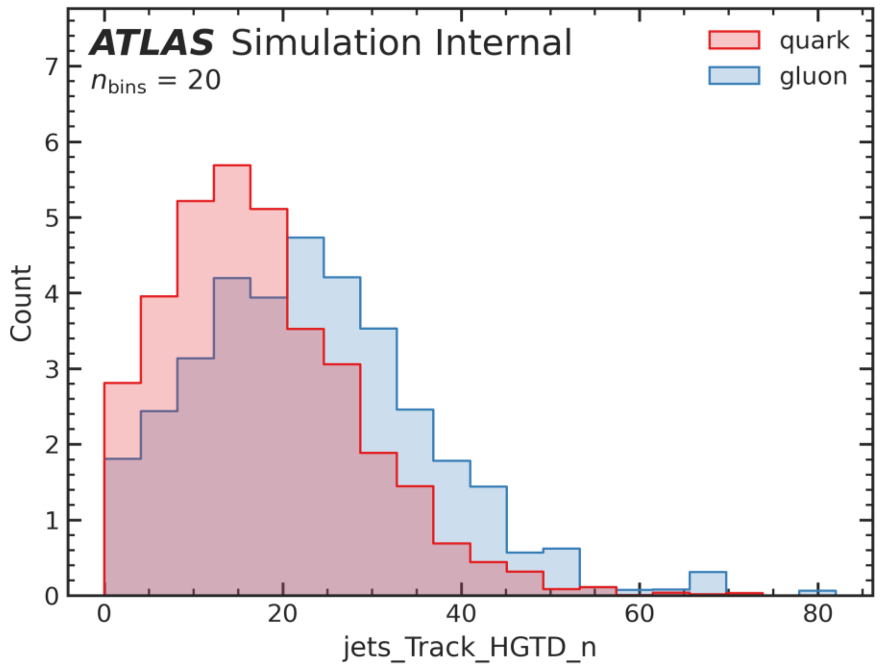
HGTD new variable: jets_Track_HGDT_n

```cpp
const std::vector<float>& times = track->auxdataConst< std::vector<float> >("HGTD_cluster_time");
for (float value : times) {
  if (value != -1.0f) {
    track_HGTD_num++;
  }
}

}
else {

  track_cluster_time_0.push_back(-2);
  track_cluster_time_1.push_back(-2);
  track_cluster_time_2.push_back(-2);
  track_cluster_time_3.push_back(-2);
}
```

HGTD has 4 clusters



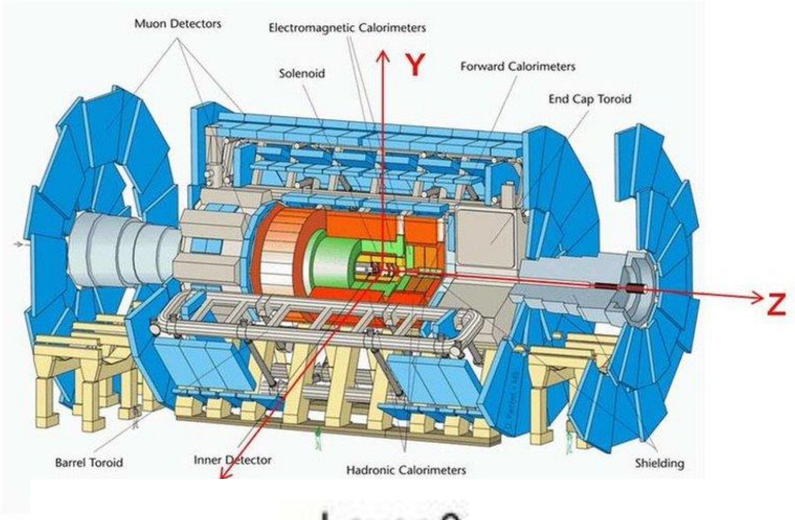
d	1
u	2
s	3
c	4
b	5
t	6
g	21

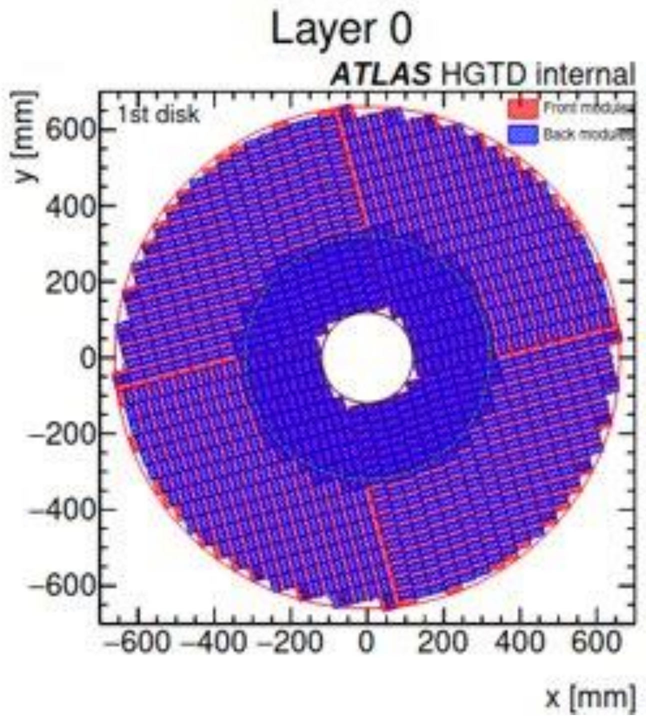
```
jets_PartonTruthLabelID { 3, 1 }
jets_Track_HGTD_cluster_time { {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, { -0.0868788f, -1.00000f, -1.00
000f, -1.00000f }, { -0.0626068f, -0.133852f, -1.00000f, -1.00000f }, { -0.0399399f, -0.131161f, -0.0712185f, -0.201877f }, {
-1.00000f, -1.00000f, -1.00000f, -1.00000f }, { -1.00000f, -0.0767756f, -0.0902195f, -1.00000f }, { -0.104699f, -1.00000f, -0.
0932989f, -0.0751143f } }
jets_Track_HGTD_cluster_truth_class { {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, {}, { 4, -1, -1, -1 }, { 1,
1, -1, -1 }, { 1, 1, 1, 1 }, { -1, -1, -1, -1 }, { -1, 1, 1, -1 }, { 1, -1, 1, 1 } }
jets_Track_n { 16, 6 }
```

HGTD low level variable

$$\log \Delta_{x,y}^{HGTD} = \log \sqrt{(x^a - x^b)^2 + (y^a - y^b)^2}$$

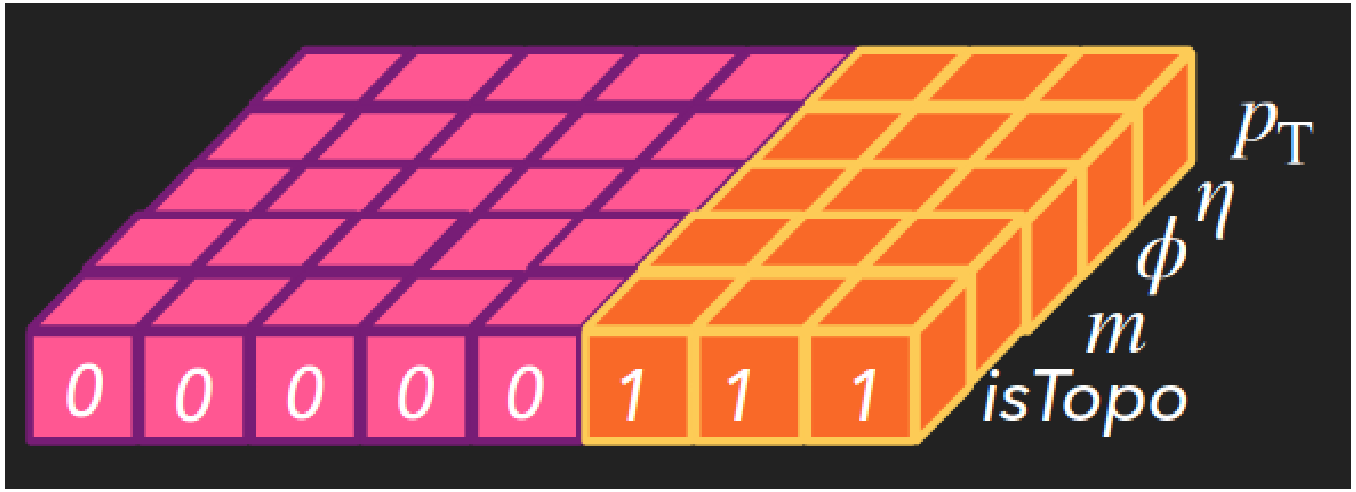
InDetTrackParticlesAuxDyn.HGTD_cluster_merged
InDetTrackParticlesAuxDyn.HGTD_cluster_raw_time
InDetTrackParticlesAuxDyn.HGTD_cluster_shadowed
InDetTrackParticlesAuxDyn.HGTD_cluster_time
InDetTrackParticlesAuxDyn.HGTD_cluster_truth_class
InDetTrackParticlesAuxDyn.HGTD_extension_chi2
InDetTrackParticlesAuxDyn.HGTD_extrap_x
InDetTrackParticlesAuxDyn.HGTD_extrap_y
InDetTrackParticlesAuxDyn.HGTD_has_extension
InDetTrackParticlesAuxDyn.HGTD_primary_expected





Input variables

- **PFOs are always used as input.**
- **For the forward region**, Topo Towers and Track information are included, along with the new HGTD variable.
- **For the central region**, Tracks are also added.
- These are concatenated with the PFOs, and an additional variable (isTopo or isTrack, depending on the object) is introduced. This approach follows the implementation by Samuel: slides



Constituent Interaction Variables
$\log \Delta = \log \sqrt{(y^a - y^b)^2 + (\phi^a - \phi^b)^2}$
$\log \Delta_{x,y}^{HGTD} = log\sqrt{(x^a - x^b)^2 + (y^a - y^b)^2}$
$\log k_T = \log\left(\min\left(p_T^a, p_T^b\right)\Delta\right)$
$z = \min\left(p_T^a, p_T^b\right) / (p_T^a + p_T^b)$
$\log m^2 = \log\left(p^{\mu,a} + p^{\mu,b}\right)^2$

Constituent Variables
$\Delta y = y - y_{\text{jet}}$
$\Delta\phi = \phi - \phi_{\text{jet}}$
$\Delta\phi^{HGTD} = \phi^{HGTD} - \phi_{\text{jet}}$
$\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2}$
$\log p_T$
$\log E$
$\log \frac{p_T}{p_{T,\text{jet}}}$
$\log \frac{E}{E_{\text{jet}}}$
m

Evaluation. Forward region.

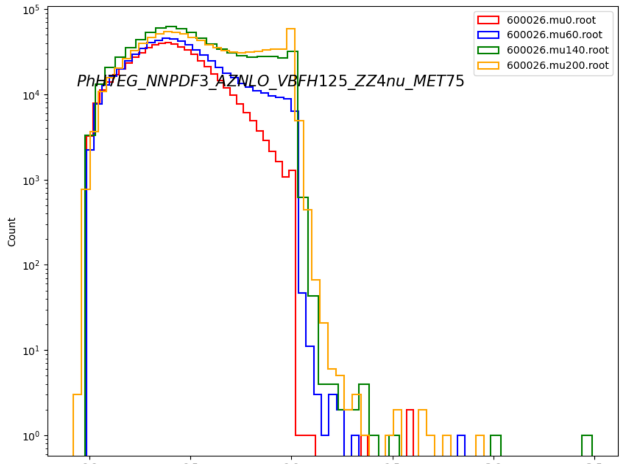
Bins split PT: (20, 60, 100,150, 500) [GeV]

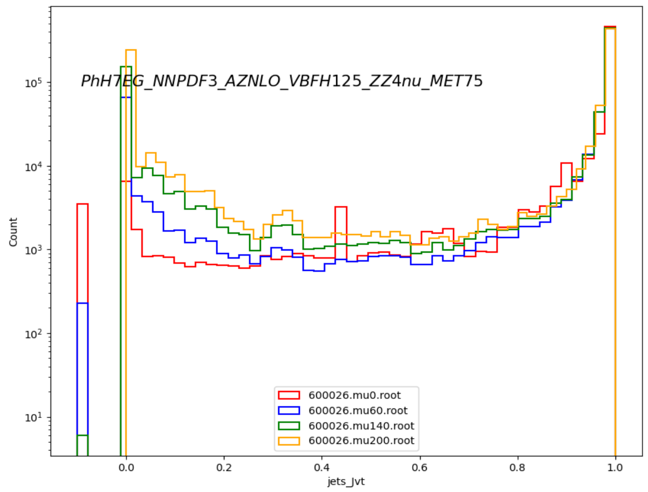
- Forward region: jets_y > 2.5. Evaluation in jets_y: [2.5, 3.0, 3.5, 4.0]

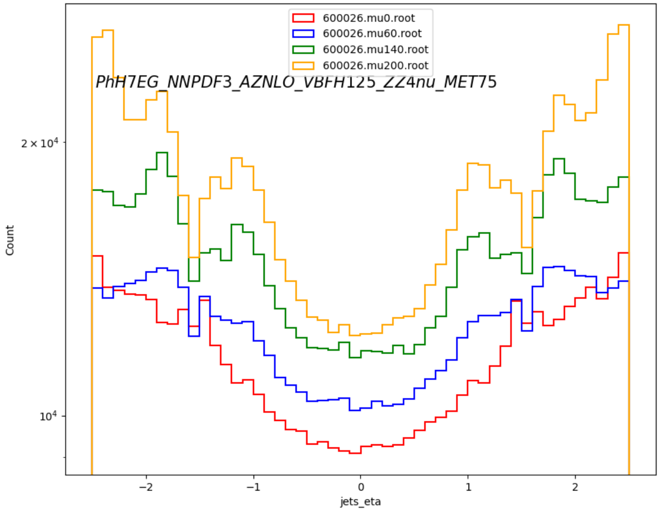
Pile-up studies

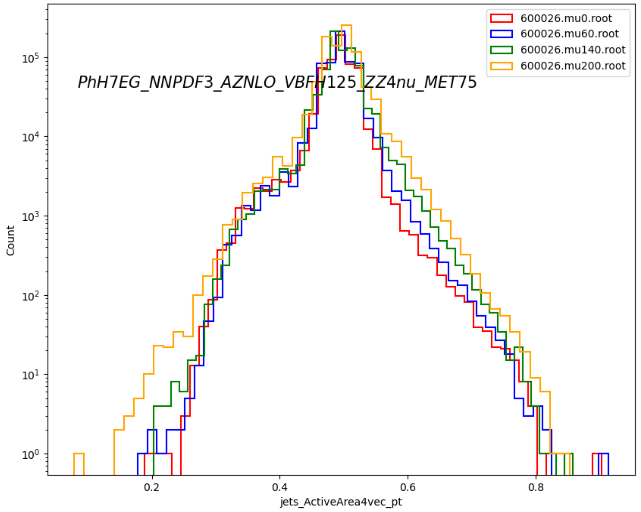
Test mu60

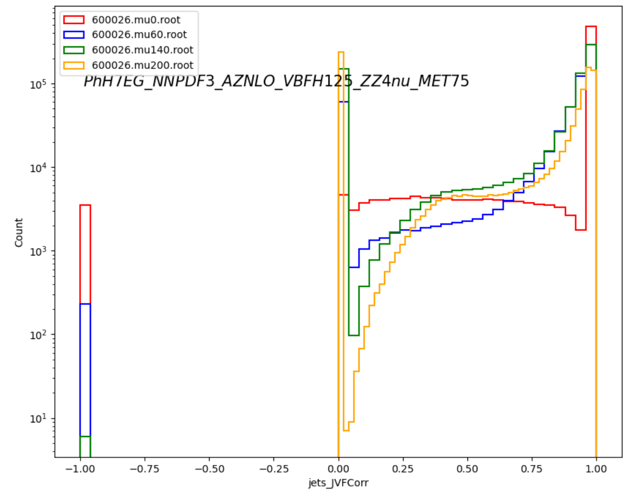
New input variables? Central region

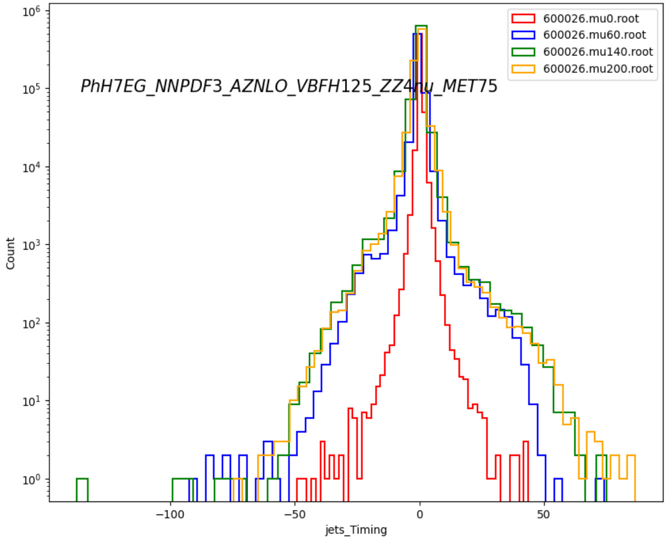




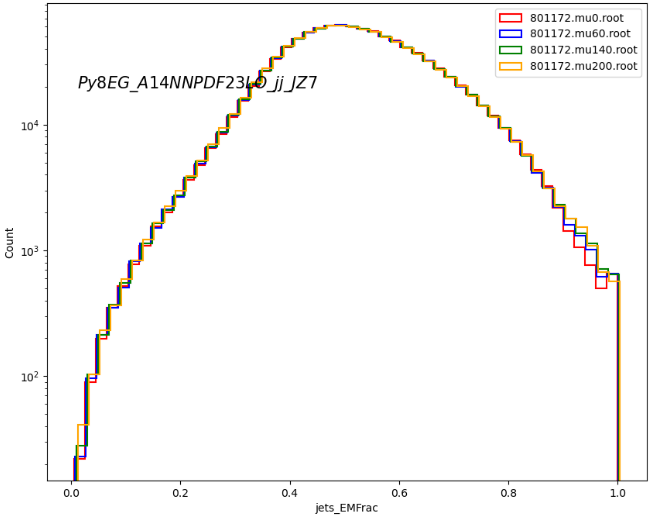


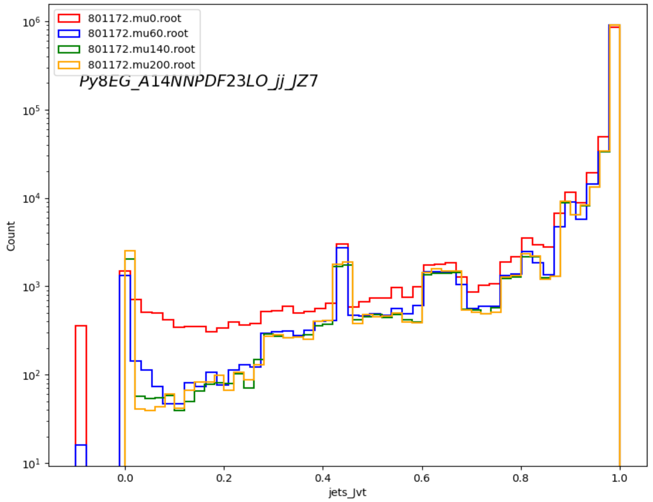


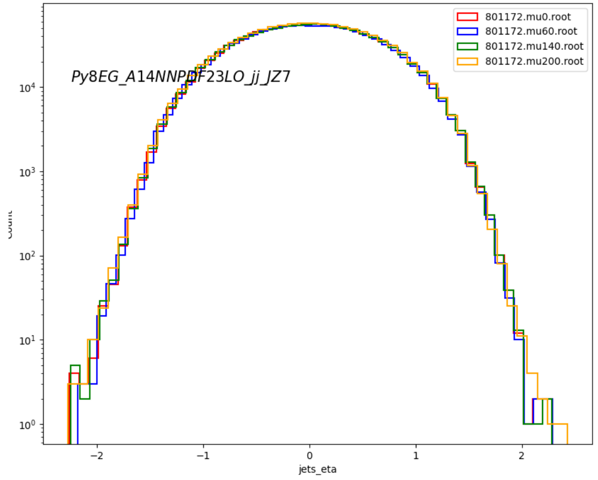


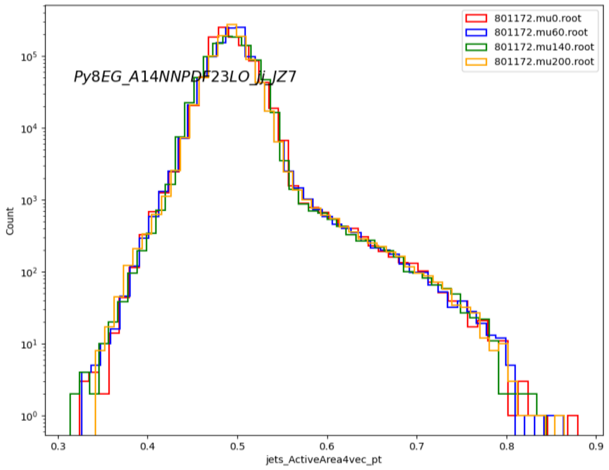


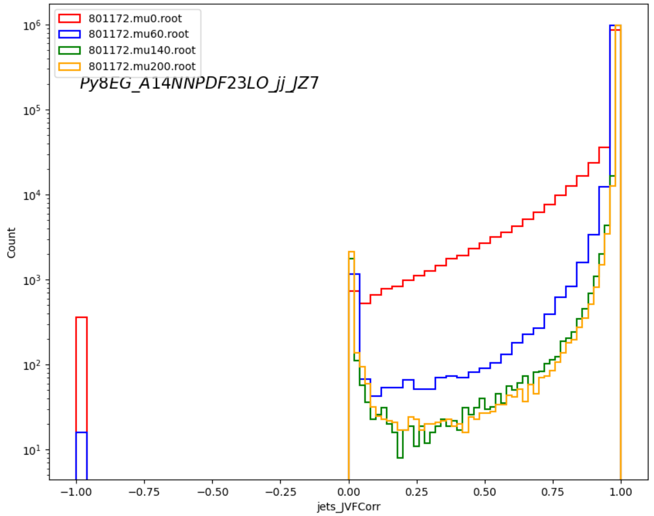
New input variables? Central region

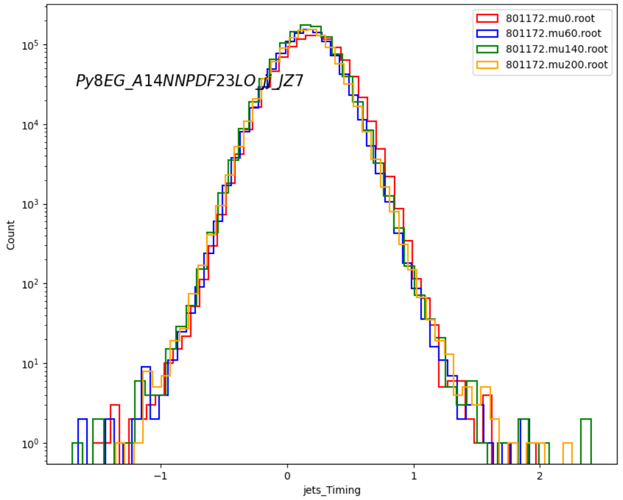




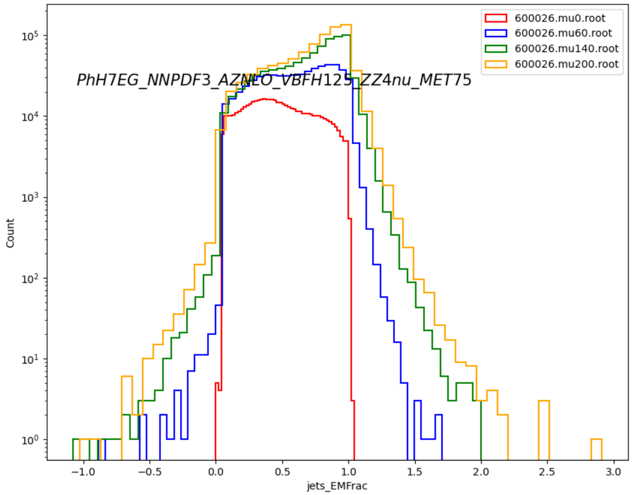


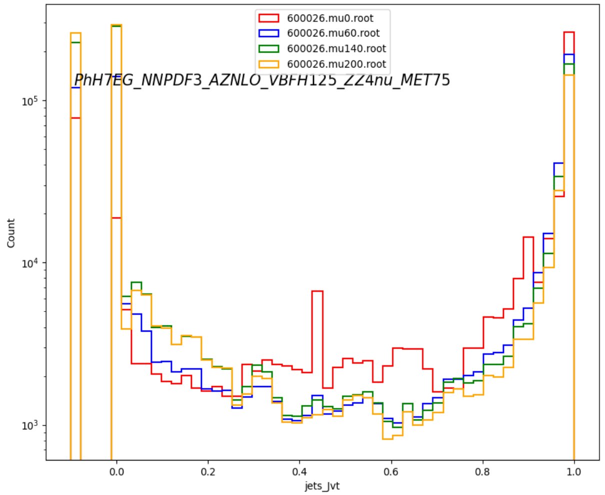


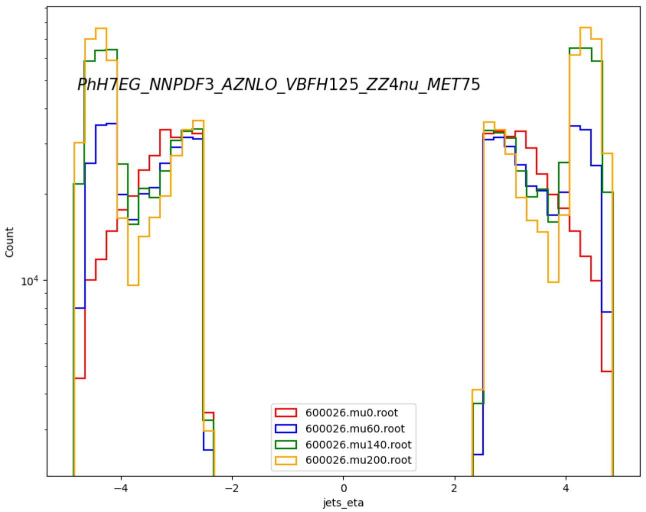


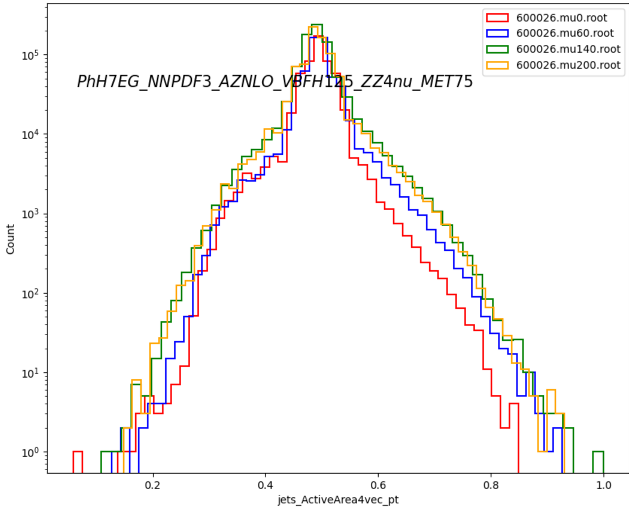


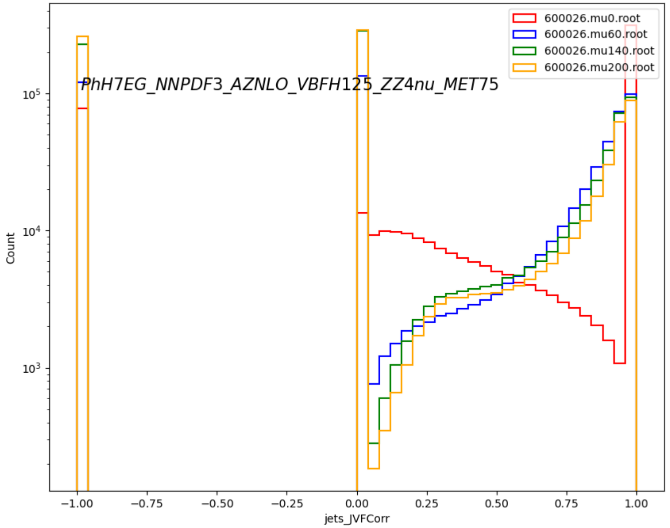
New input variables? Forward region

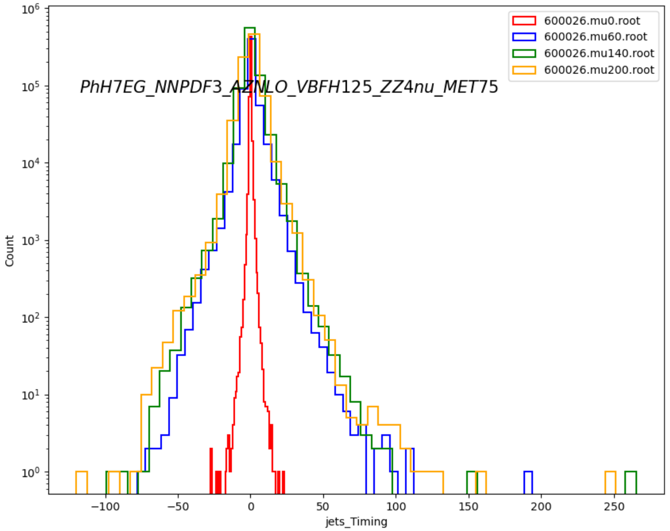




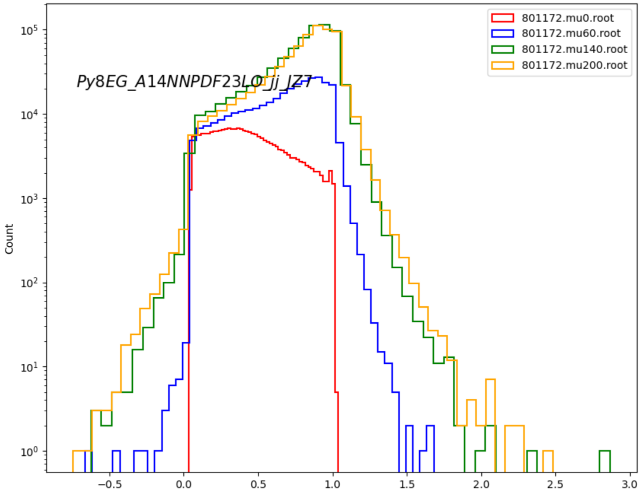


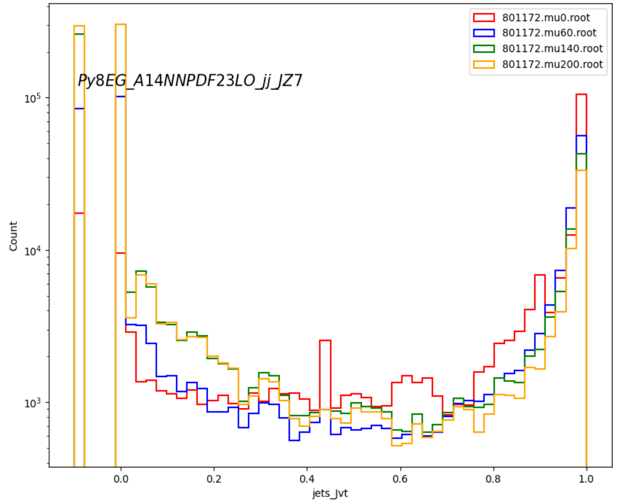


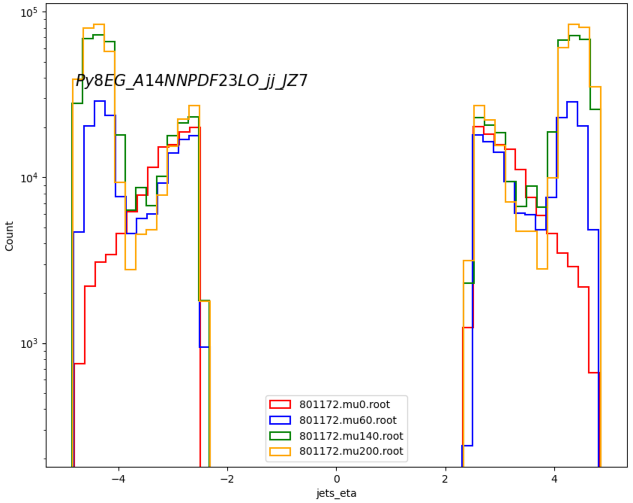


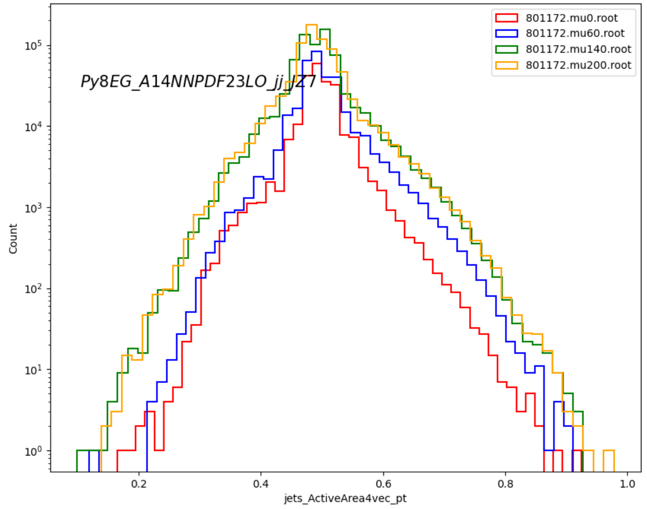


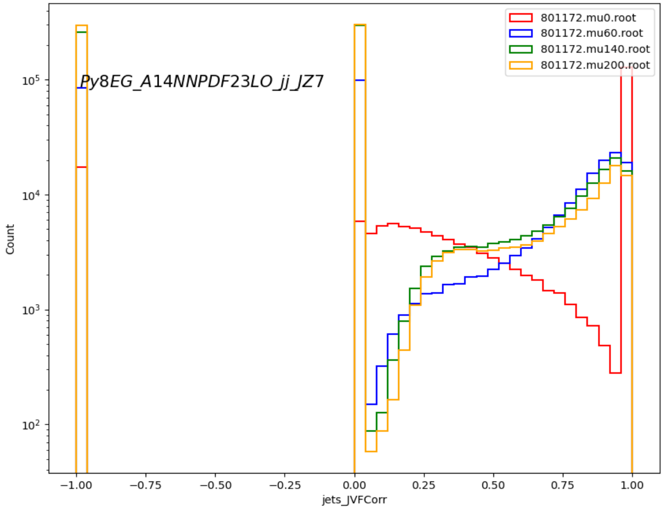
New input variables? Forward region

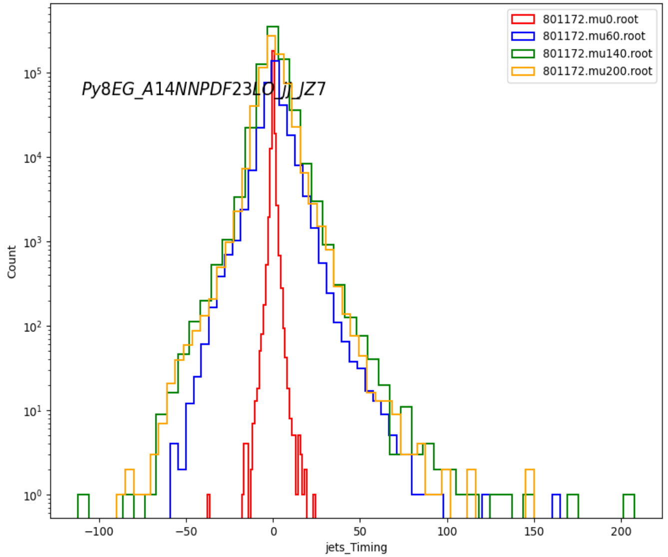




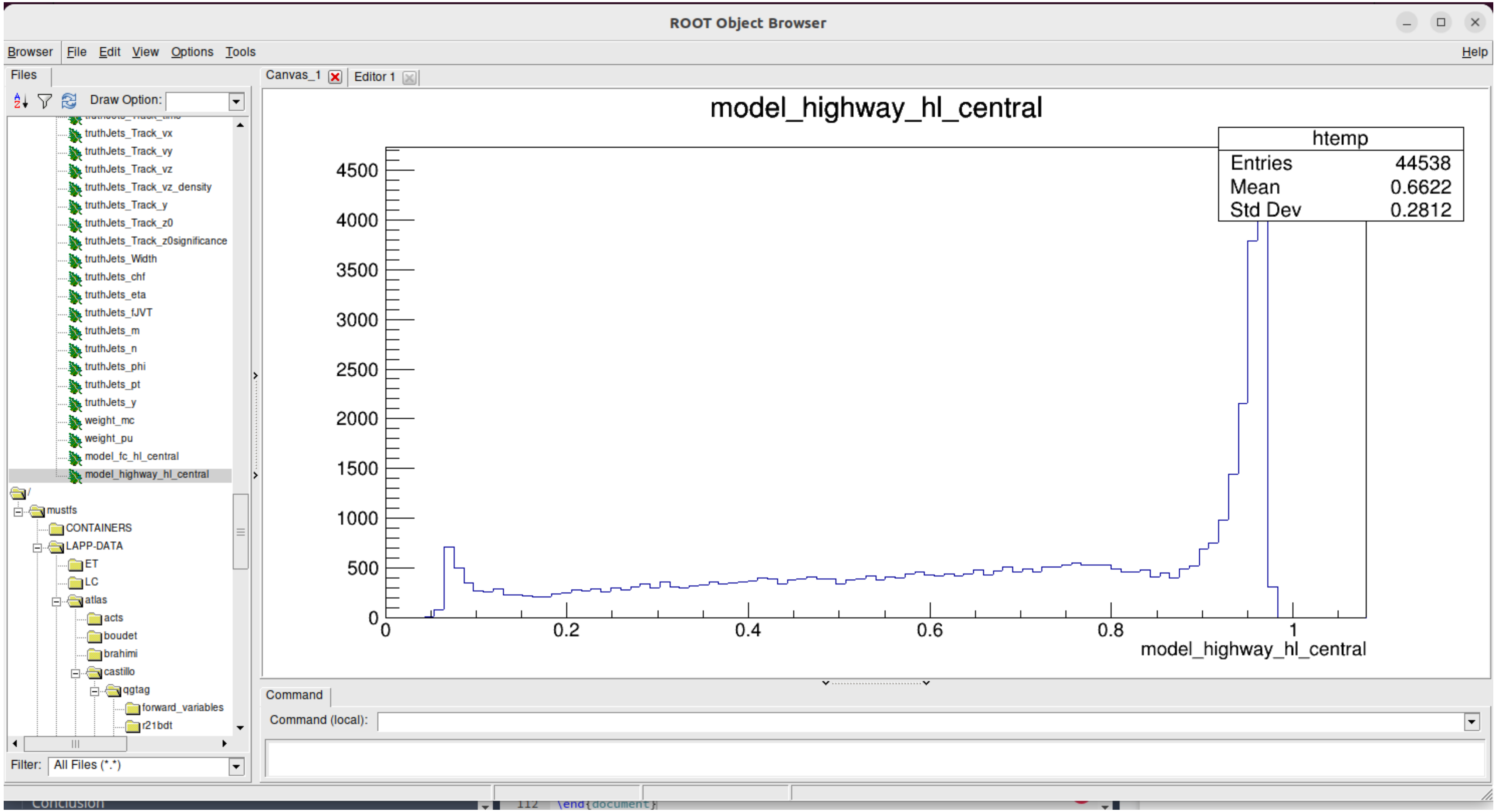








Scores in root files



Scores in root files