Quark gluon tag – First look in the HL-LHC

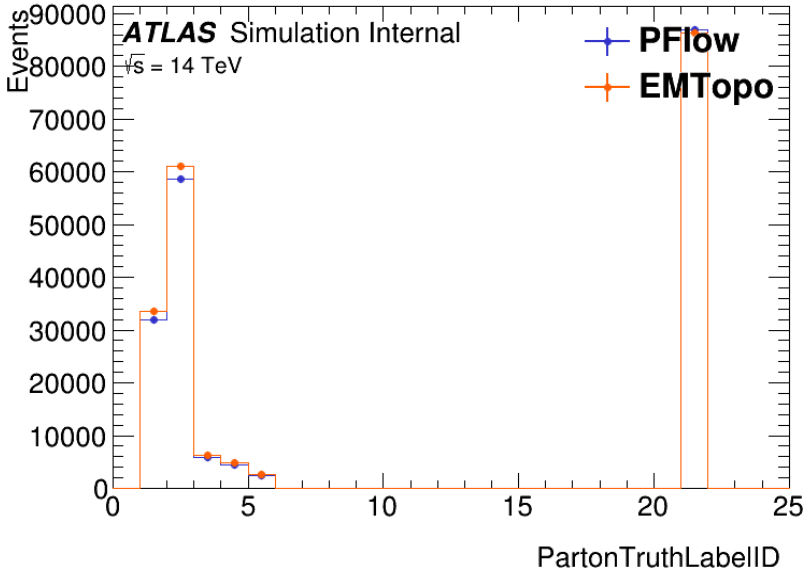
Florencia Castillo

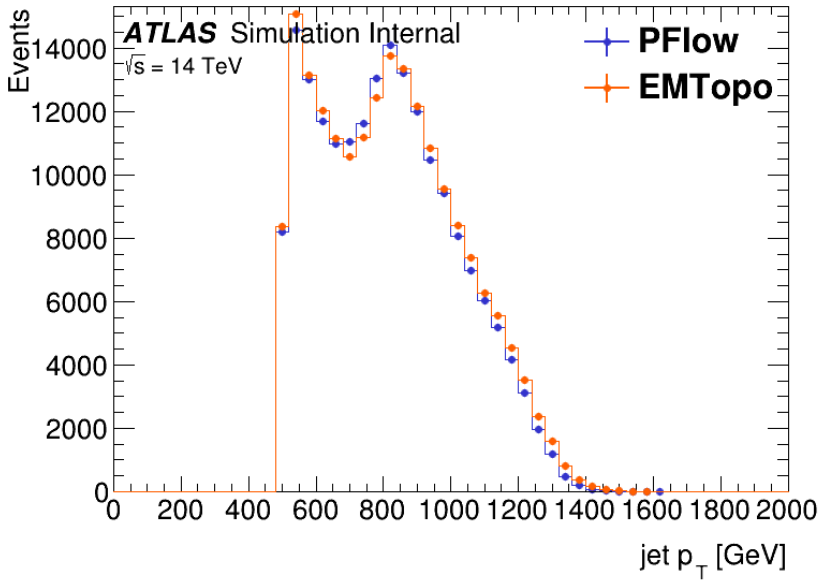
Setup and derivations

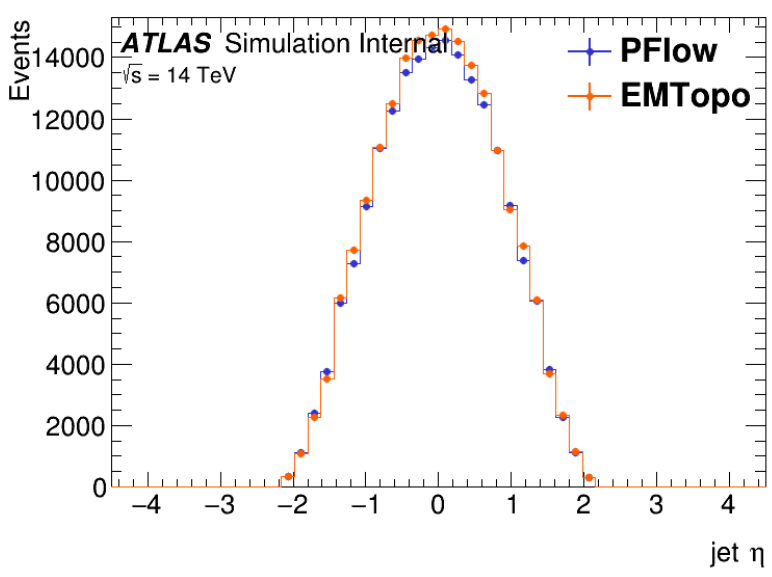
- Running over di-jet production
  - mc21_14TeV.801165.Py8EG_A14NNPDF23LO_jj_JZ0.deriv.DAOD_JETM1.e8481_s4038_r14365_r14367_p5742
  - mc21_14TeV.801166.Py8EG_A14NNPDF23LO_jj_JZ1.deriv.DAOD_JETM1.e8481_s4038_r14365_r14367_p5742
  - mc21_14TeV.801167.Py8EG_A14NNPDF23LO_jj_JZ2.deriv.DAOD_JETM1.e8481_s4038_r14365_r14367_p5742
  - mc21_14TeV.801168.Py8EG_A14NNPDF23LO_jj_JZ3.deriv.DAOD_JETM1.e8481_s4038_r14365_r14367_p5742
  - mc21_14TeV.801169.Py8EG_A14NNPDF23LO_jj_JZ4.deriv.DAOD_JETM1.e8481_s4038_r14365_r14367_p5742
  - mc21_14TeV.801170.Py8EG_A14NNPDF23LO_jj_JZ5.deriv.DAOD_JETM1.e8481_s4038_r14365_r14367_p5742
- Creation of a minitree algorithm
  - https://gitlab.cern.ch/fcastill/hl-lhc-tagger
  - Running using: asetup AnalysisBase,main,latest

Jets -> EMTopo vs PFlow (no calibrated)

- Cuts: pt > 500 GeV, abs(eta)<2.1. ghost particles associated to the tracks



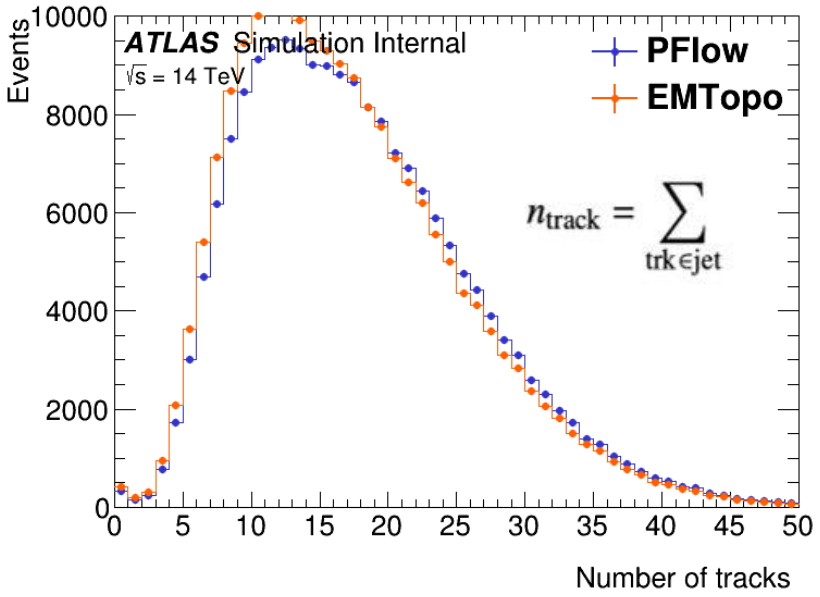


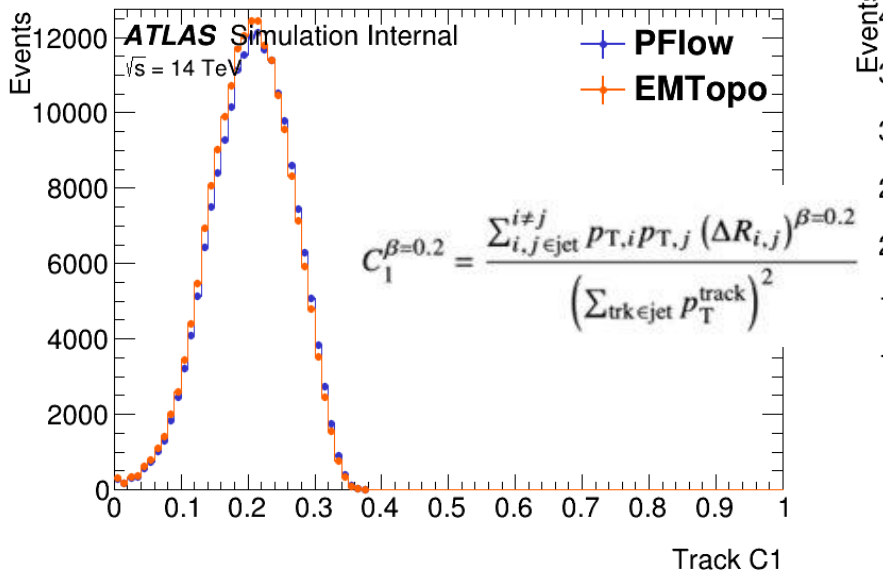


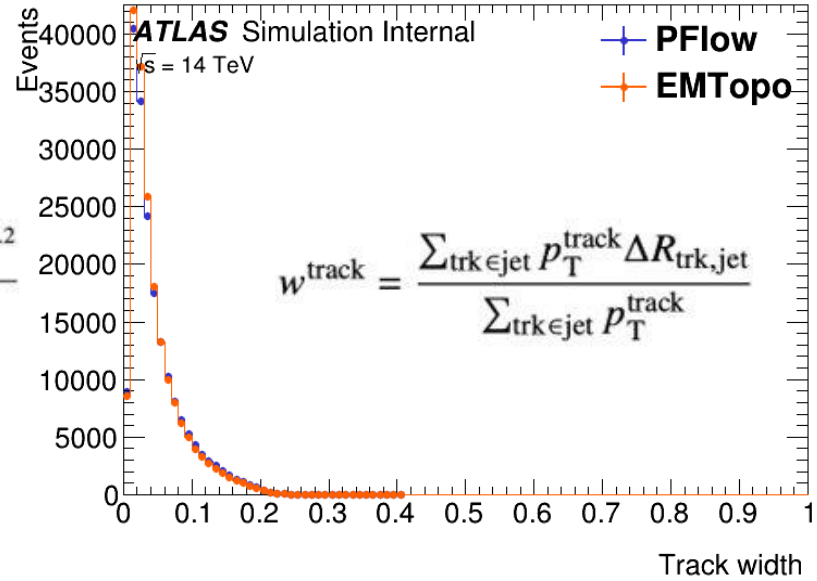
- Not big discrepancy in the central region for high pt jets

Jets -> EMTopo vs PFlow (no calibrated)

- Cuts: pt > 500 GeV, abs(eta)<2.1. ghost particles associated to the tracks





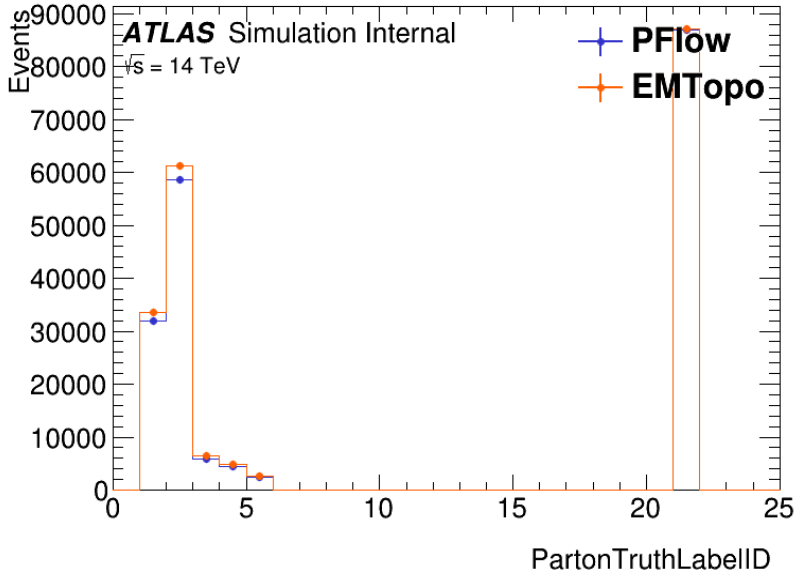


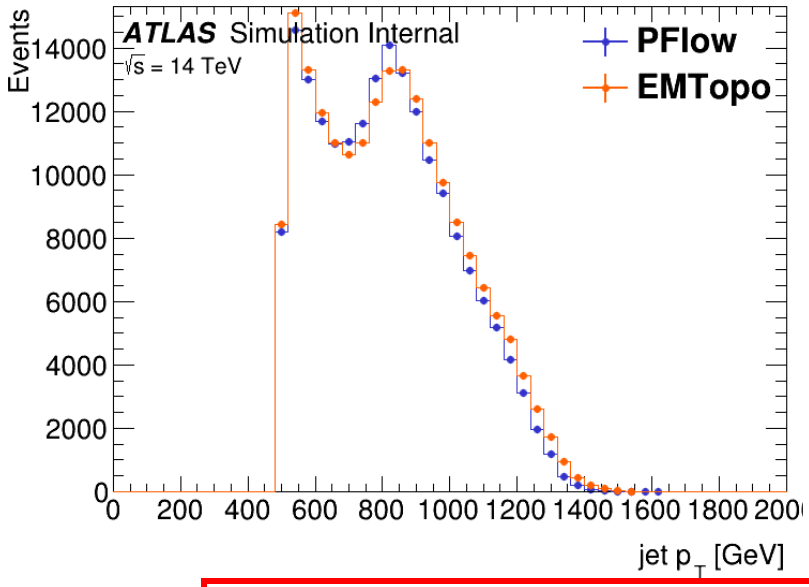
- bdt input variables (along with flat pt)

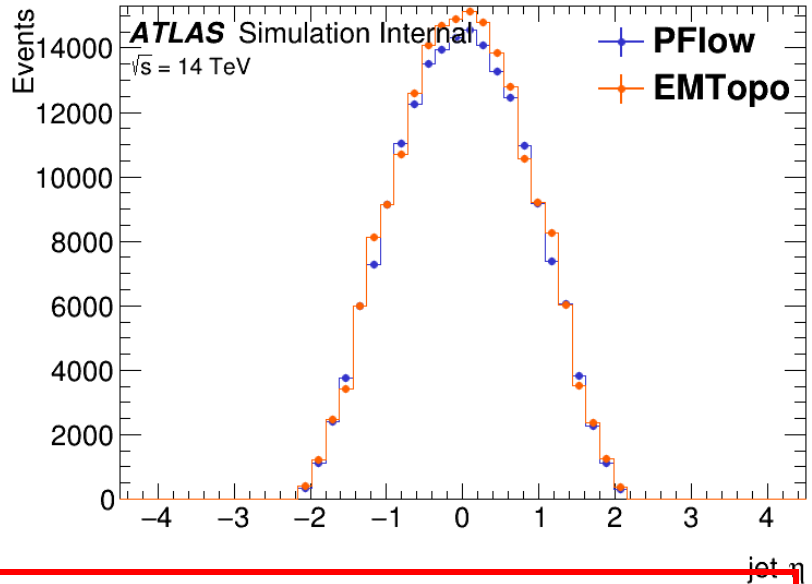
EMTopo calibrated vs PFlow uncalibrated

- Default option is implemented (same used for HL-LHC analyses):
  - HLLHC/JES_MC16_HLLHC_June2021.config
- PFlow not ready jet
  - They are very close to get the samples
    - https://its.cern.ch/jira/browse/ATLGBLCONDTAGS-135
  - Once we have the samples, it is still at least a month to get calibrations... it could easily be several months

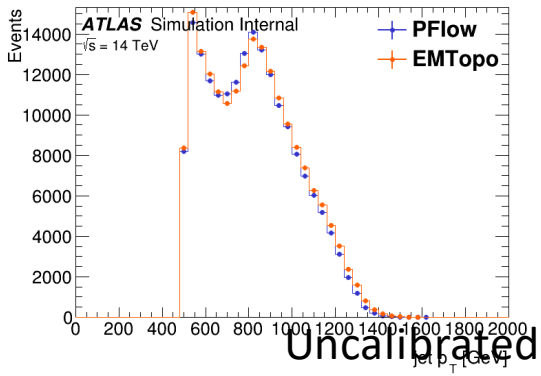
EMTopo calibrated vs PFlow uncalibrated

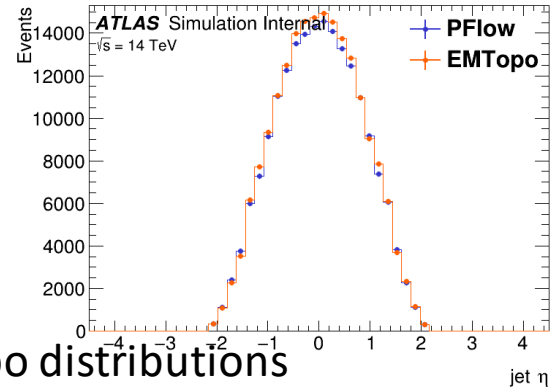






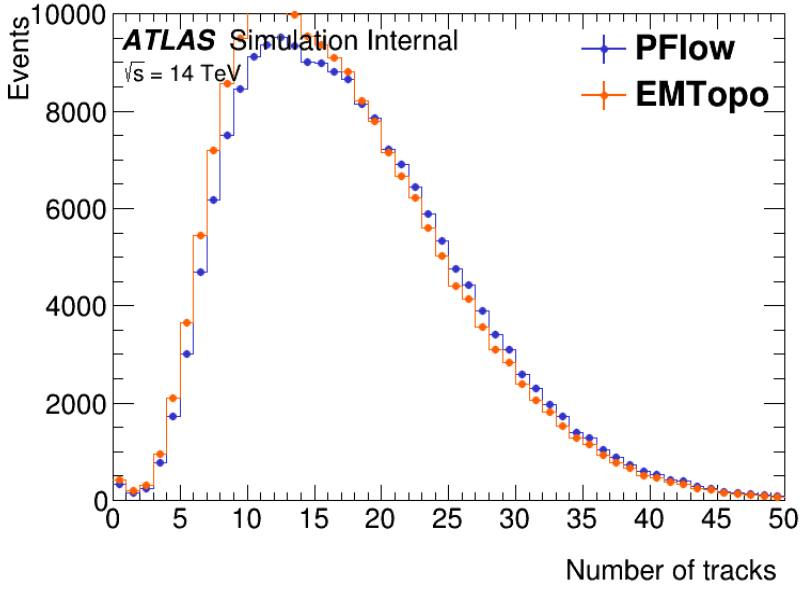
Calibrated EMTopo differs more from PFlow

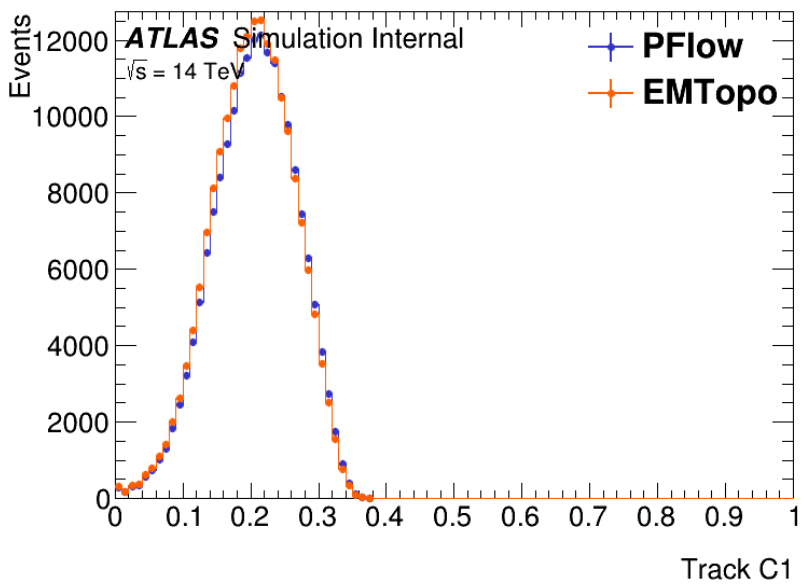


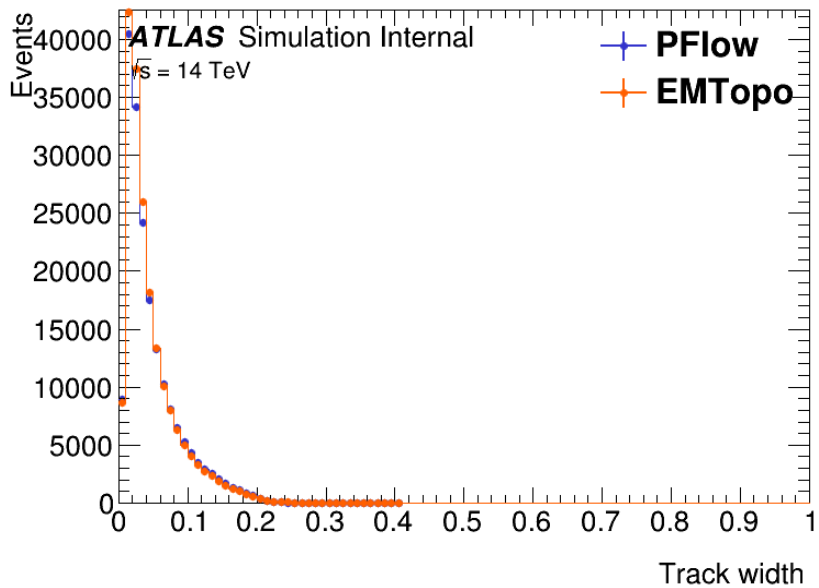


Uncalibrated EM topo distributions

EMTopo calibrated vs PFlow uncalibrated







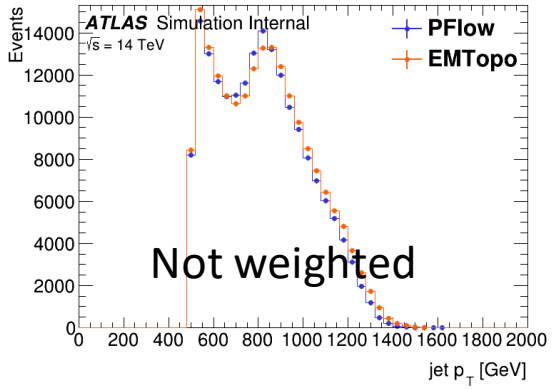
Calibrated EMTopo differs more from PFlow

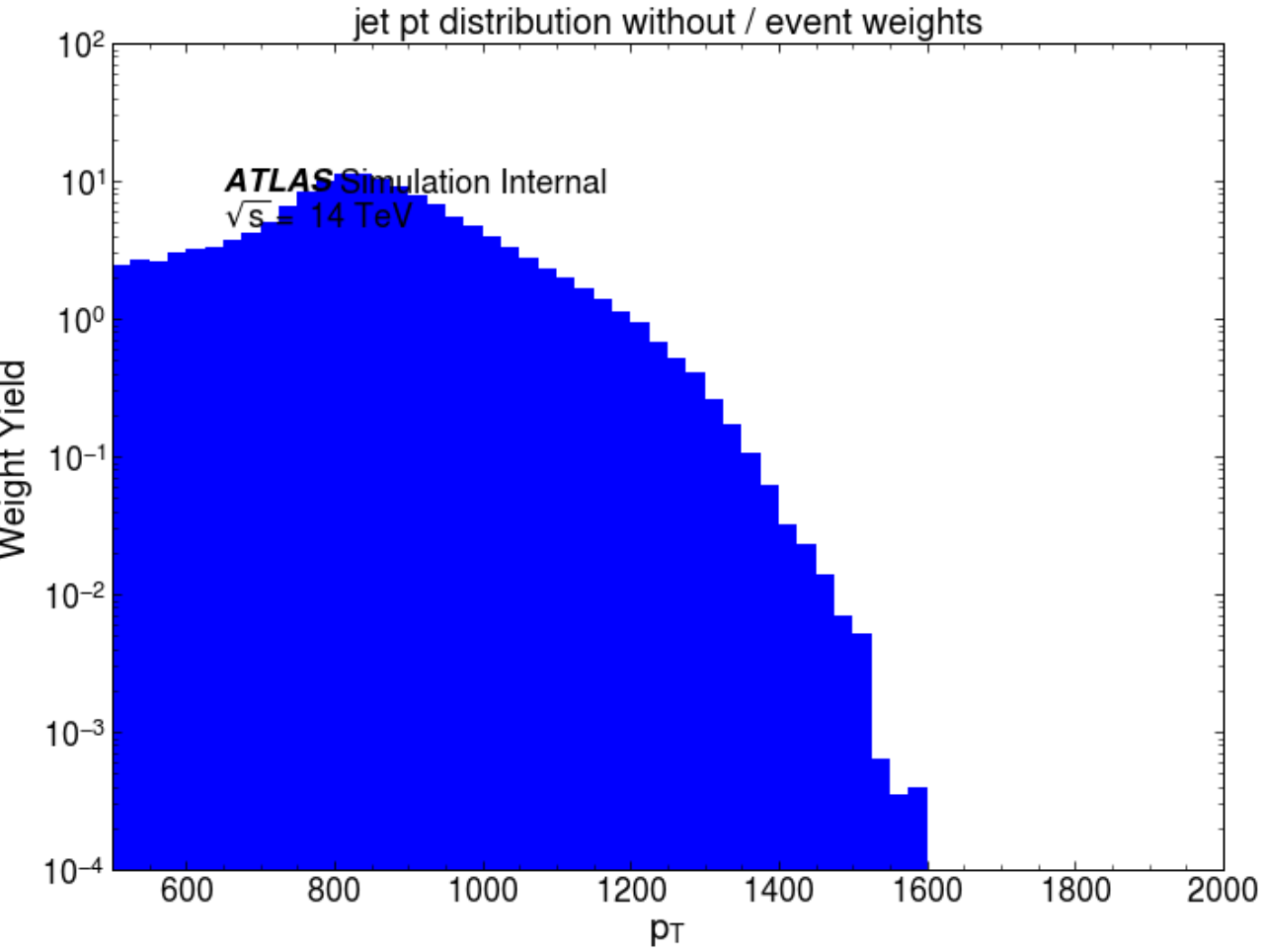
BDT performance

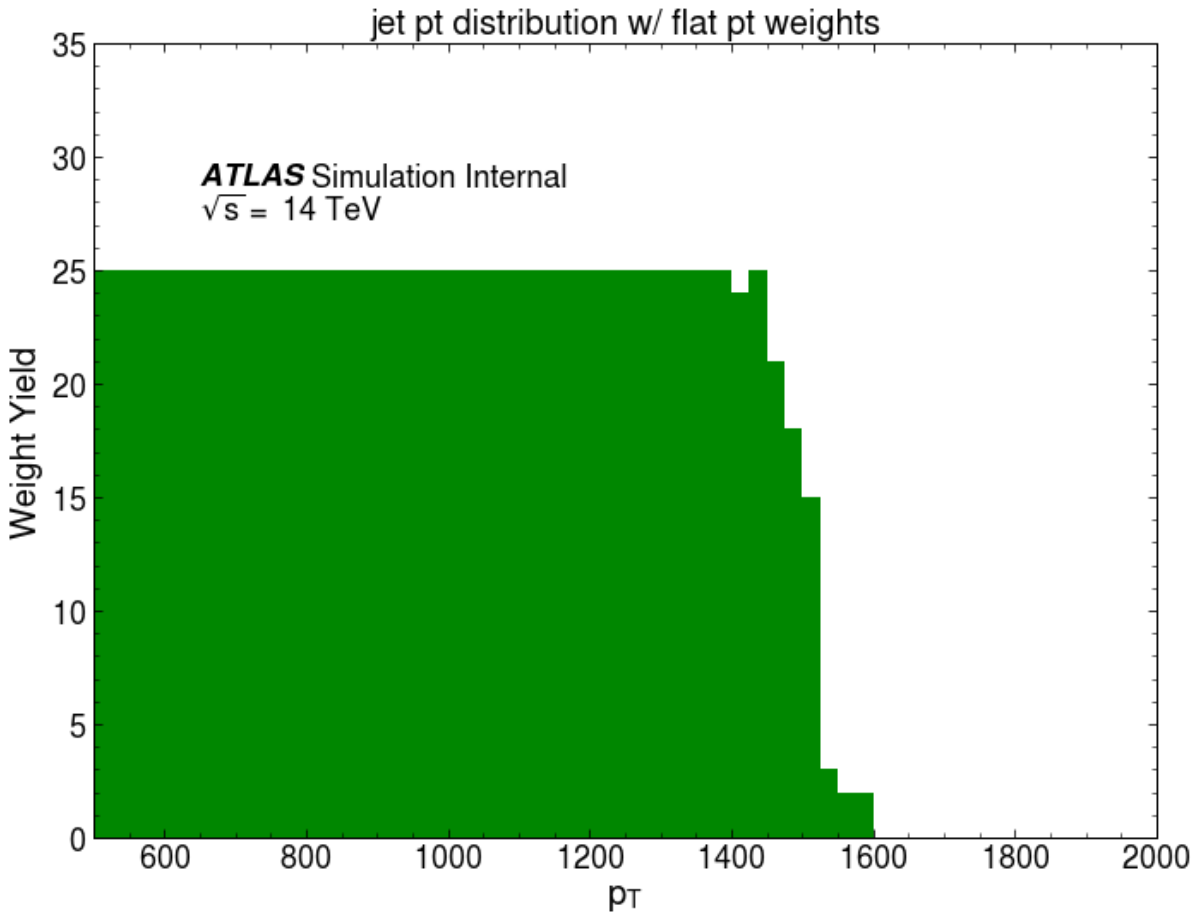
- Input varabiles: Ntracks, Track C1 (energy correlation), Track With and flat pT
- Jets: Calibrated EM topo jets
- Cuts: pT > 500 GeV and abs(eta)<2.1
- Weights (PU not included/it was in the run 2 version):

```
aux =  sample_ak["mconly_weight"]
sumweight=ak.sum(aux, axis=0)
event_weight = sample_ak["xsec"] * sample_ak["mconly_weight"] / sumweight
```

Pt weighted and flat



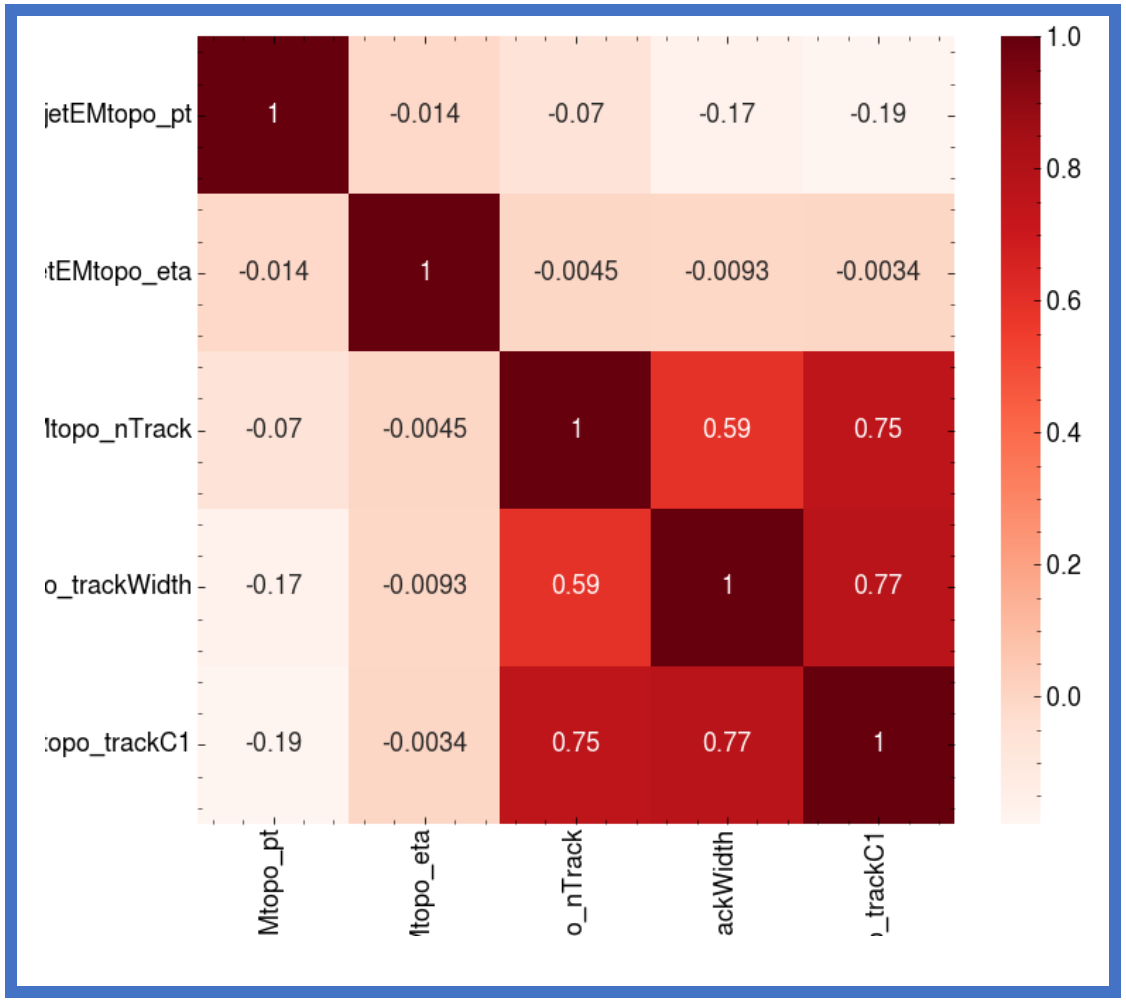




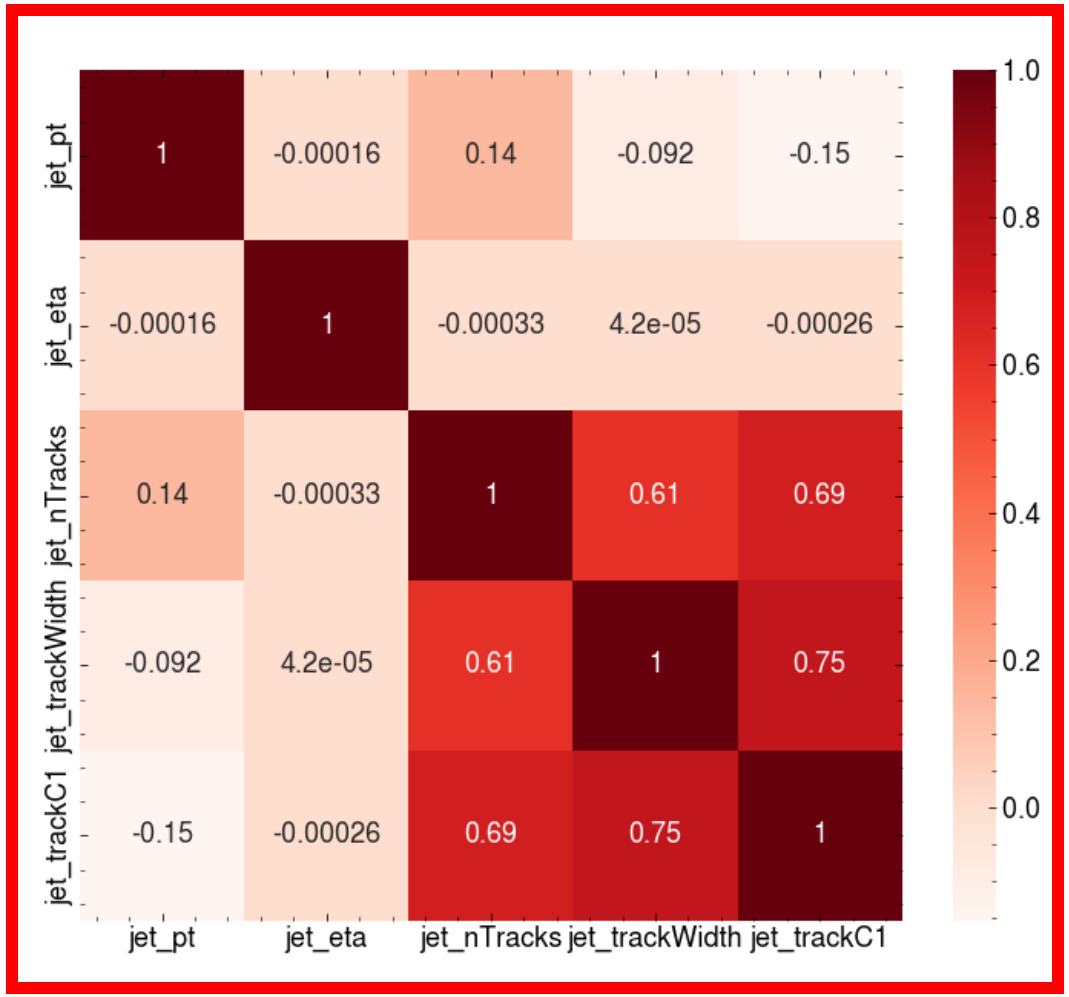
Correlation

Discrepancies might due to low stats for HL-LHC and new geometry of the detector

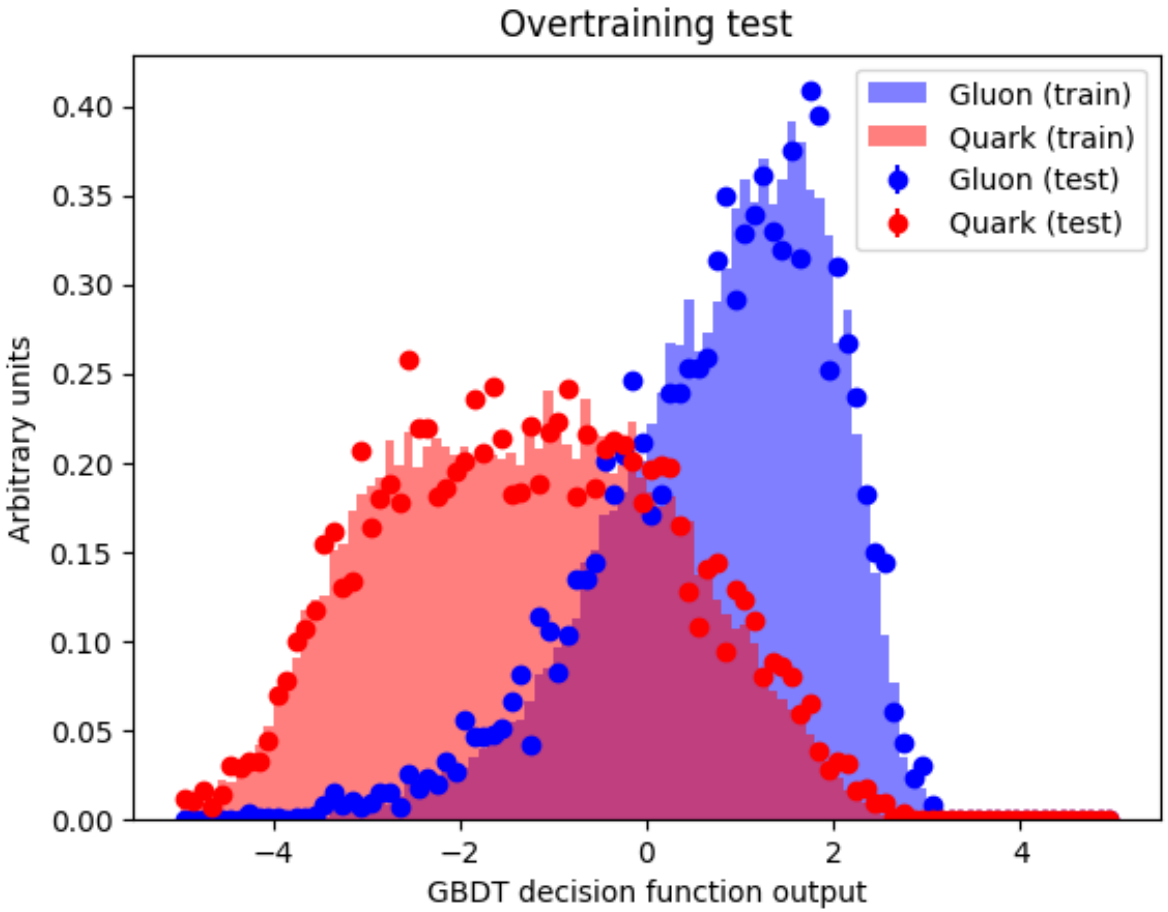
HL-LHC simulation

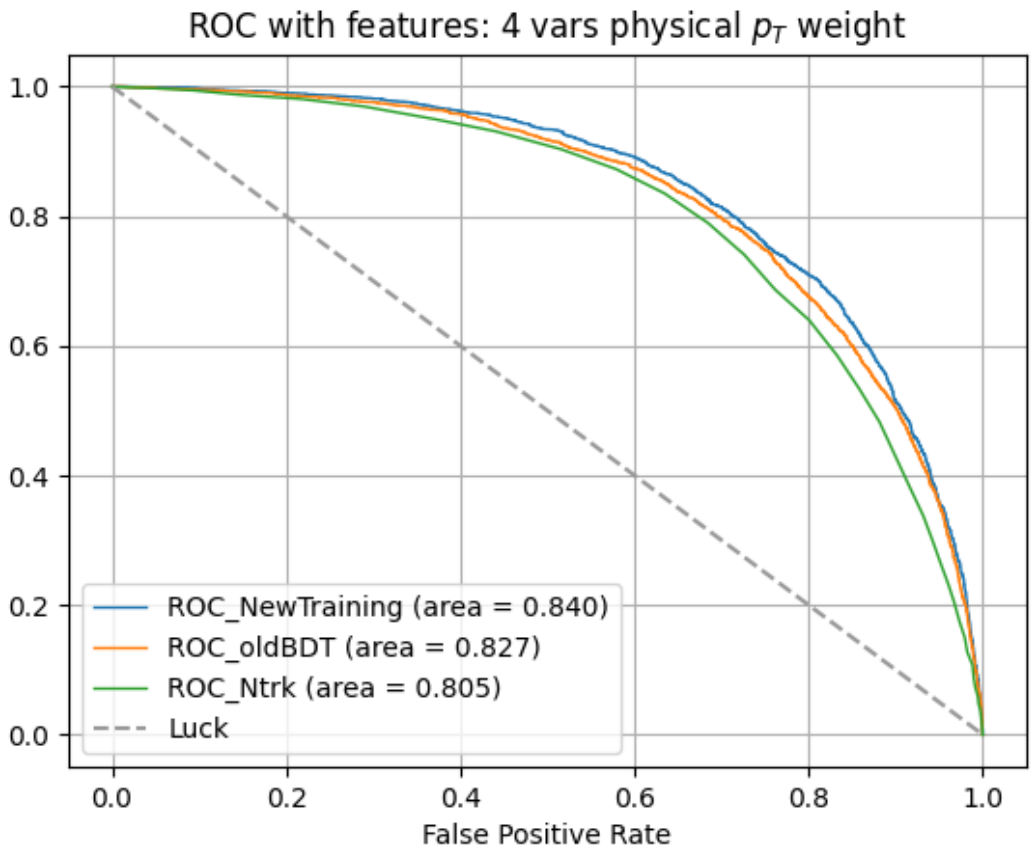


Run 2 data and simulation



BDT performance





Low stats, need to request more samples

Outlook & Next steps

- Fist look to BDT in HL-LHC environment looks ok!
- Present these results when I come from holidays in jetetmiss tag meeting
  - Probably move to PFlow when they are ready
  - Ask for more stats
  - Samples with more populations of forward jets
- Update git lab with instructions
  - https://gitlab.cern.ch/fcastill/hl-lhc-tagger
- Start using transformers
  - https://gitlab.cern.ch/atlas-jetetmiss/tagging/quarkgluontagging/qgtraining