First Look at VBF_Hbb with 2017 Data

A. Raspereza[1)], A. Tumasyan[2)]

1) **DESY**, Hamburg, Germany
2) **ANSL**, Yerevan, Armenia

Outline

- MC and DATA samples
- Triggers
- Analysis strategy and status
- Event preselection
- Jet combination
- Signal vs Background discrimination

MC-Samples *(MiniAod)*

Signal

- VBF_Hbb_M125
- ggF_Hbb_M125
- ttH_Hbb_M125
- VH_Hbb_M125

Background

- QCD_HT* / QCD_bEnriched_HT*
- TTbar
- SingleTop
- DYJetsToQQ *(NO PRODUCTION FOUND)*
- WJetsToQQ *(FROM HTT400)*
- WW, WZ, ZZ

MC- campaign:

- RUNIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017realistic_v14-v*

GolobalTag:

- 94X_mc2017_realistic_v14

CMSSW_9_4_6_patch1

Data-Samples

Dataset: **/BTagCSV/Run2017F-31Mar2018-v1/MINIAOD**

RunRange: **305405 – 306460**

Integral Luminosity: **7.73 fb^{-1}**

JSON:
/afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions17/13TeV/ReReco/Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt

GlobalTag: **94X_dataRun2_v6**

CMSSW_9_4_6_patch1

Triggers

Single_BTag: **(M_{qq} > 460 GeV, $\Delta\eta_{qq}$ > 3.5)**

HLT_QuadPFJet105_88_76_15_BTagCSV_p013_VBF2_v5

with L1_TripleJet_100_85_72_VBF OR *

HLT_QuadPFJet111_90_80_15_BTagCSV_p013_VBF2_v5

with L1_TripleJet_105_85_76_VBF OR *

Double_BTag: **(M_{qq} > 240 GeV, $\Delta\eta_{qq}$ > 2.1)**

HLT_QuadPFJet105_90_76_15_DoubleBTagCSV_p013_p08_VBF1_v5

with L1_TripleJet_100_85_72_VBF OR *

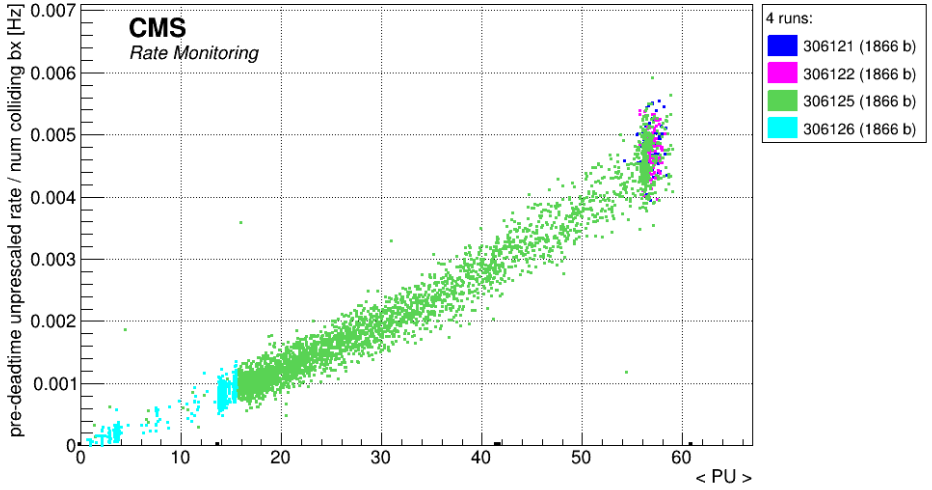
HLT_QuadPFJet111_90_80_15_DoubleBTagCSV_p013_p08_VBF1_v5

with L1_TripleJet_105_85_76_VBF OR *

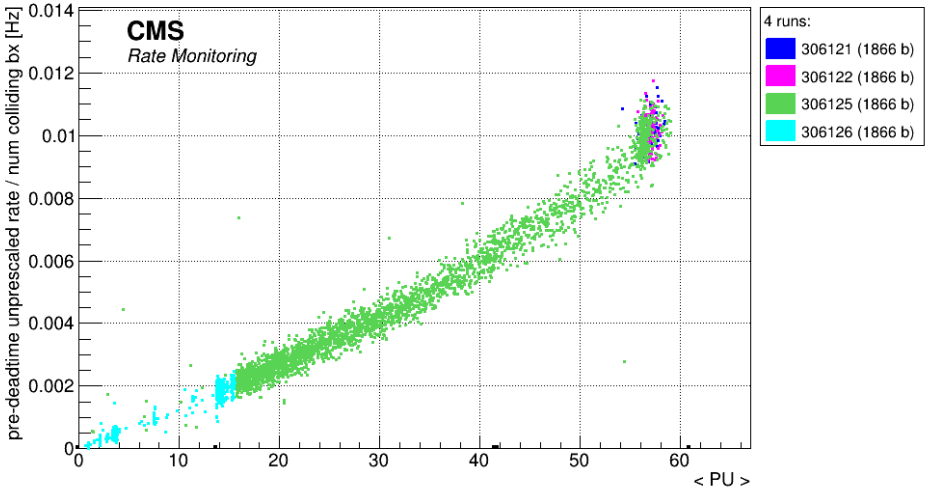
* L1_HTT280er OR L1_HTT300er OR L1_HTT320er OR L1_SingleJet170 OR L1_SingleJet180 OR L1_SingleJet200

Triggers Performance

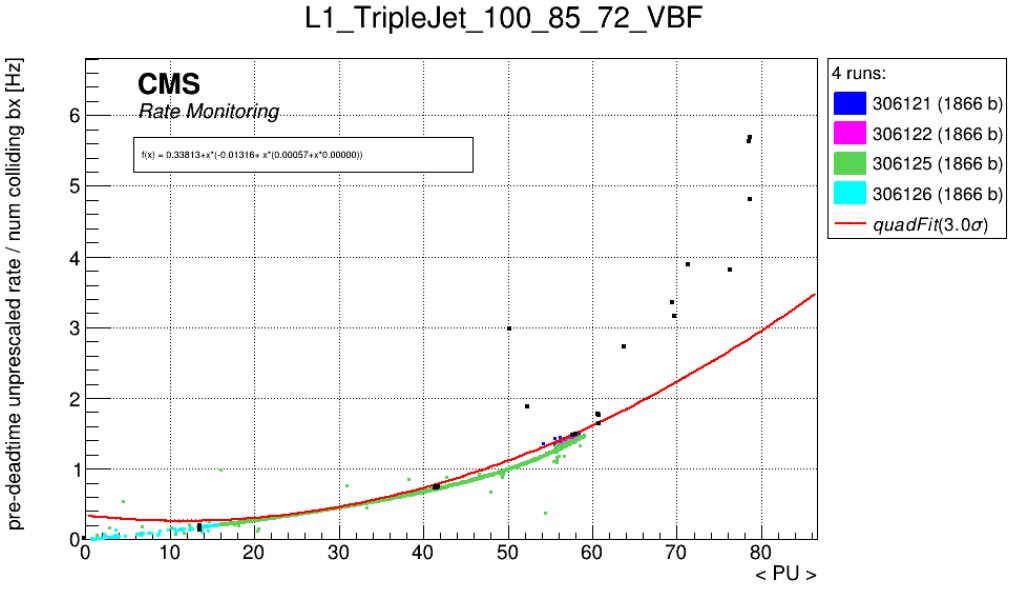
HLT_QuadPFJet105_88_76_15_BTagCSV_p013_VBF2



HLT_QuadPFJet105_90_76_15_DoubleBTagCSV_p013_p08_VBF1

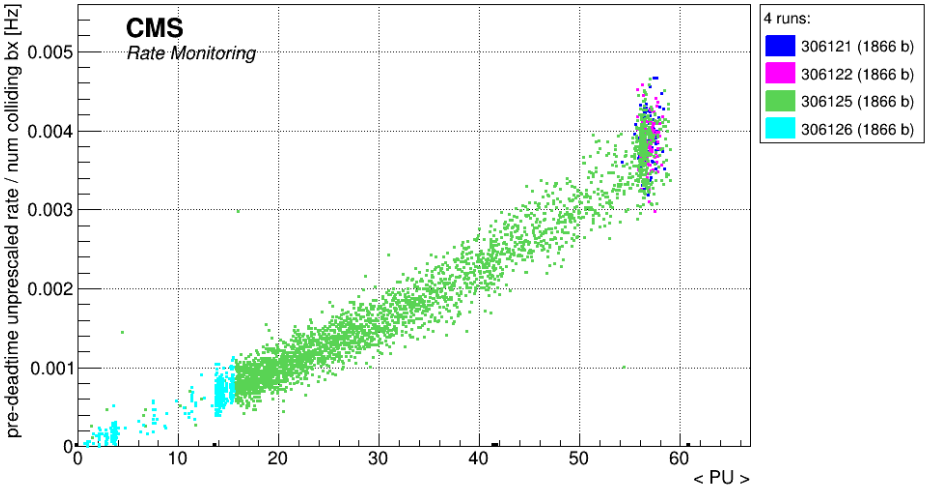


L1_TripleJet_100_85_72_VBF

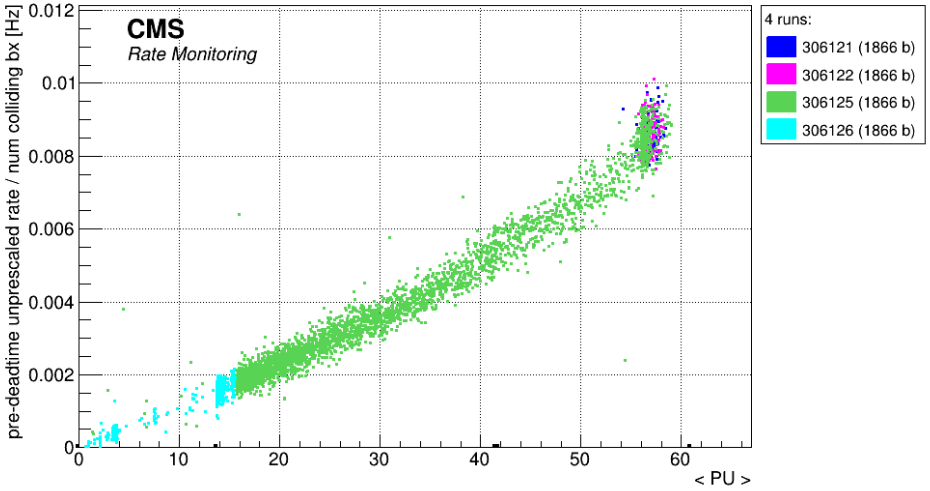


Triggers Performance

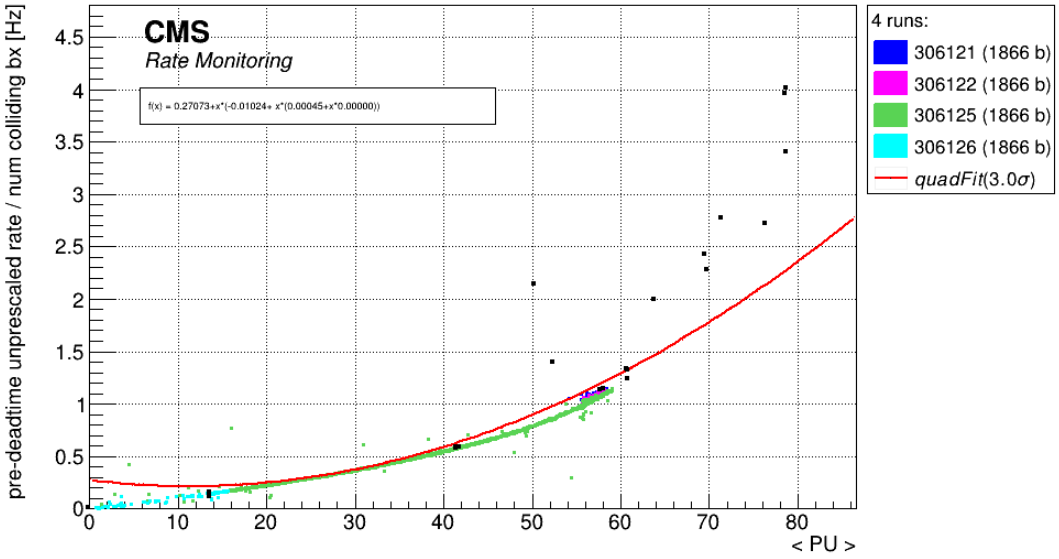
HLT_QuadPFJet111_90_80_15_BTagCSV_p013_VBF2



HLT_QuadPFJet111_90_80_15_DoubleBTagCSV_p013_p08_VBF1



L1_TripleJet_105_85_76_VBF



Analysis strategy and status

Strategy

- Two main categories: **SingelBTag** and **DoubleBTag**
- **SingleBTag** relies on *tight* VBF criteria and presence of two b-likely jets (more sensitive)
- **DoubleBTag** relies on presence of two b-tagged jets and *loose* VBF criteria (less sensitive)

Status

- **SingleBTag** **close to complete implementation going to statistical analysis**
- **DoubleBTag** **implementation ongoing**

SingleBTag: Event Preselection

- **HLT_QuadPFJet105_88_76_15_BTagCSV_p013_VBF2_v5** OR HLT_QuadPFJet111_90_80_15_BTagCSV_p013_VBF2_v5
- Good PV
- 4 jets with $P_T >$ 105, 90, 76, 20 GeV and $|\eta| < 4.7$ and passing Loose PUid
  https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJetID
- 1 medium b-tagged jet with $P_T > 20$ GeV:
  b-tagging with "pfCombinedInclusiveSecondaryVertexV2BJetTags"
  https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X
- a pair of jets with Minv > 500 GeV and $\Delta\eta > 4.2$ among 7 P_T-leading jets excluding the highest b-tagged jet
  (requirement for jets: $|\eta| < 4.7$ and pass loose PUid and)
- MuonVeto: *No isolated muon* with $P_T > 5$ GeV, $|\eta| < 2.4$, $\Delta V_Z < 1$cm, $\Delta V_{XY} < 0.25$cm (*Isolated* if Energy_of_04_cone / Mu_energy < 0.4)
- Electron Veto: *No isolated electron* with $P_T > 7$ GeV, $|\eta| < 2.4$, $\Delta V_Z < 0.2$ cm, $\Delta V_{XY} < 0.05$ cm (*Isolated* if Neutral Energy_of_03_cone / Ele_energy < 0.4)

Jet combination

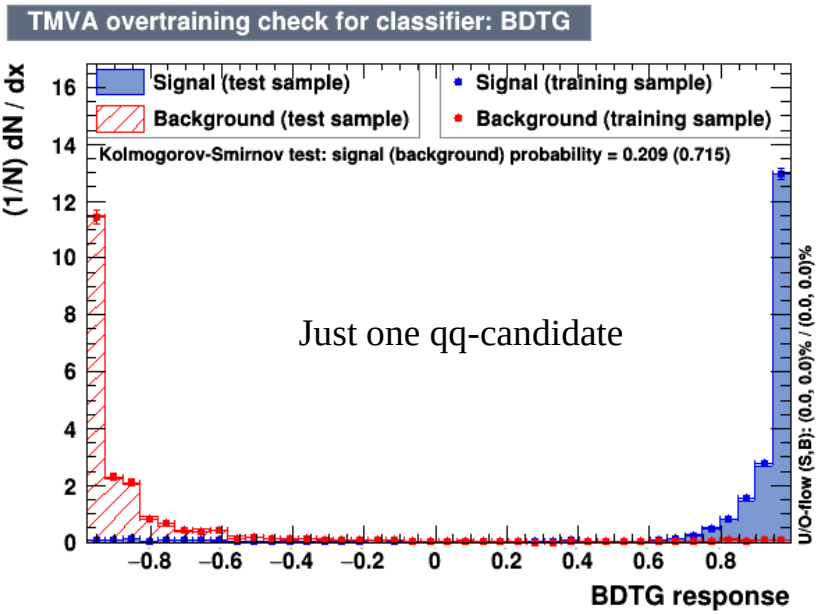
- **The most DeepCSV b-tagged jet in event is taken as 1st b-jet. (~ 99 % purity)**
- **2nd b-jet identification is done using MVA in 2 different categories**
  1. **Just one pair of jets which satisfy the requirement of VBF-topology:** Minv > 500 GeV and $\Delta\eta > 4.2$. This pair is taken as tagging jets. The 2nd b-jet is selected among 7 P_T-leading jets (*excludeing1st b-jet and tagging jets*) based on MVA
  2. **More than one pair of jets satisfying the requirements of VBF-topology:** The 2nd b-jet is selected among 7 P_T-leading jets (*excluding 1st b-jet*) based on MVA. Then as tagging are selecting the pair with the maximum of $\Delta\eta$ and satisfying Minv > 500 GeV.
- **$\Delta\varphi_{bb} < 1.6$ applied**

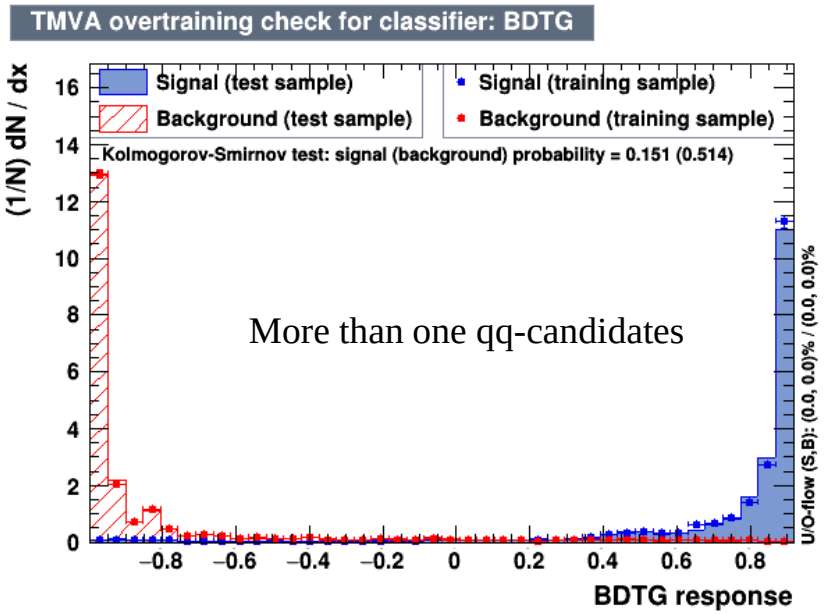
As the MVA inputs the following components are used for both cases:

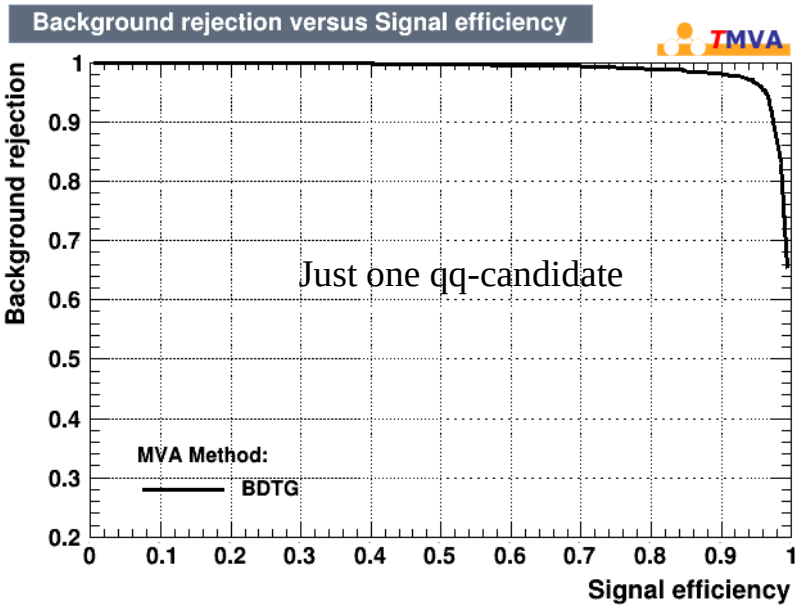
P_T, P_T-order, η, η-order, DeepCSV-out, DeepCSV-out-order, PU-jetID

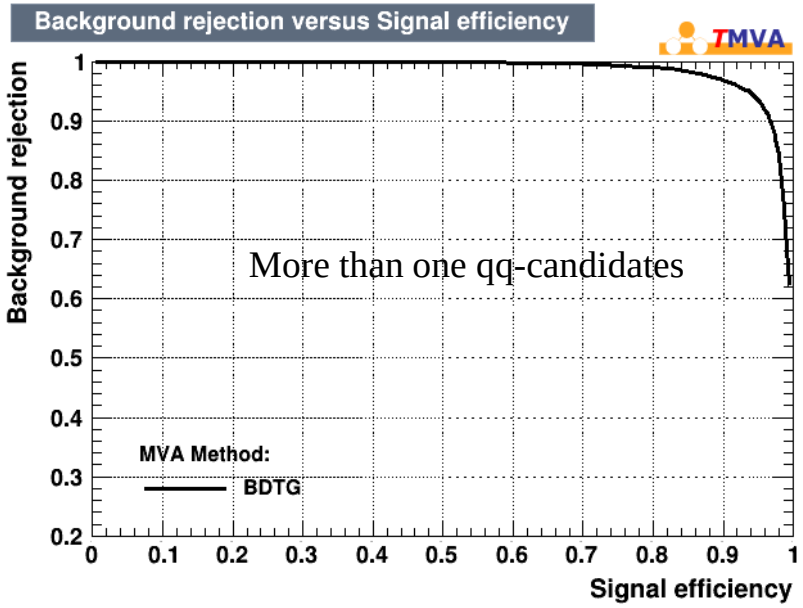
Gradient BDT has been chosen as MVA-method

2nd b-jet identification with TMVA



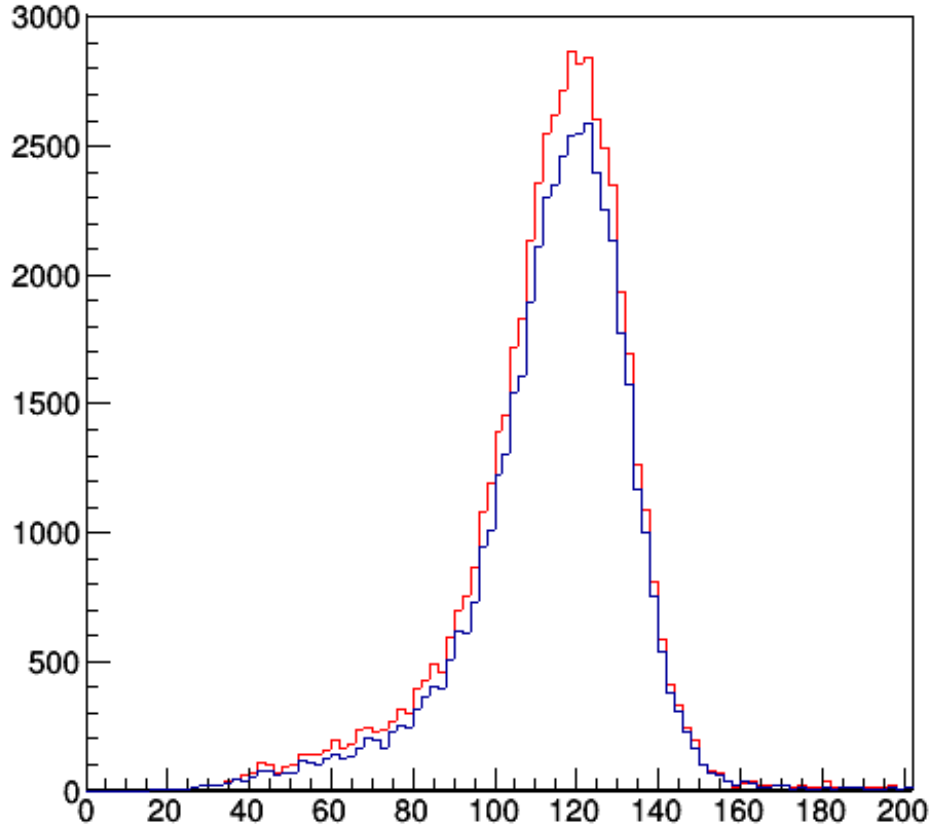






Signal efficiency

2^{nd} b-jet identification with MVA allows extend the acceptance ~ 11 % wrt the case when two most b-tagged jets are selecting as b-jets from Higgs decay.



Purity

TMVA: ~ 90 %
High_bDisc ~ 91 %

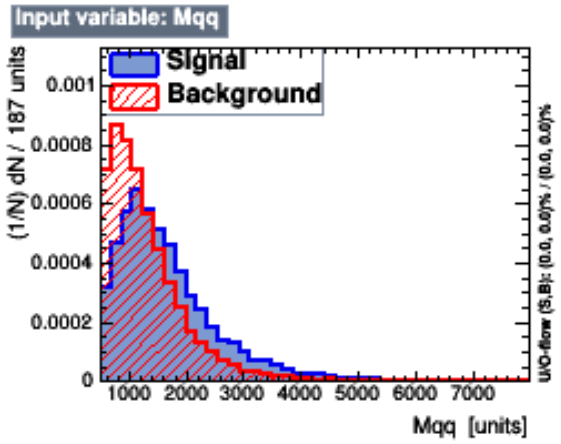
Resulting signal efficiency is ~ **1.1 %**
Expected number of events with ~7.73 fb^{-1} is ~ **188**

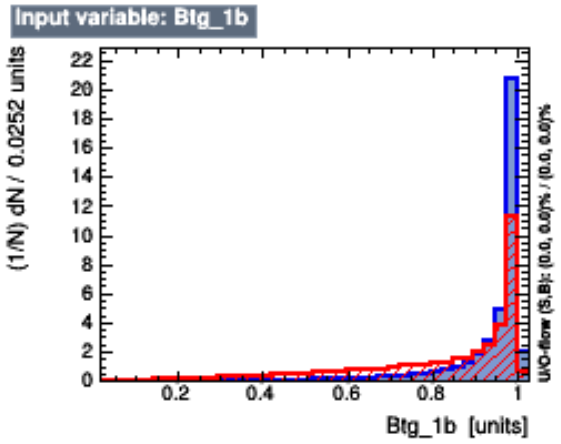
Signal vs Background

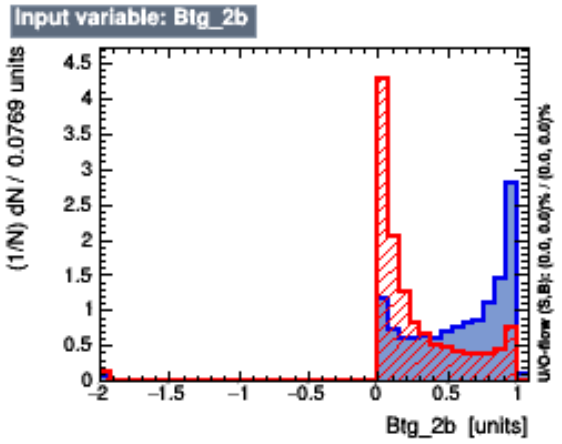
- **As a background ~ 5 % of Data was used**
- **As the MVA inputs the following components are used:**

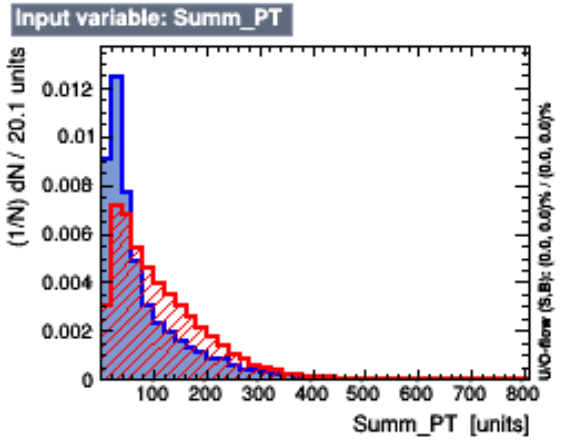
M_{qq}, $\Delta\eta_{qq}$, $\Delta\varphi_{qq}$, BTagDisc of b-jets, QGdisc of q-jets, Njets with $P_T > 20$, Njets in qq η-gap, HTT_Rest, E_Rest, PT-Summ of 4-jets, PZ-Summ of 4-jets

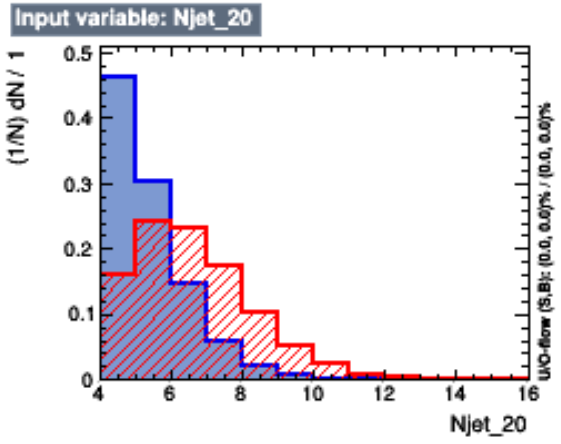
- **Gradient BDT was used as MVA-method**

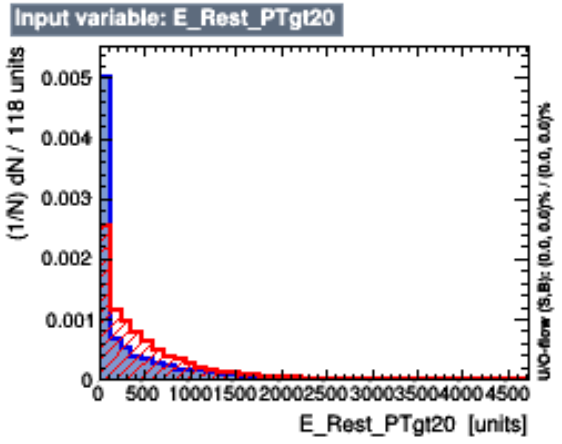




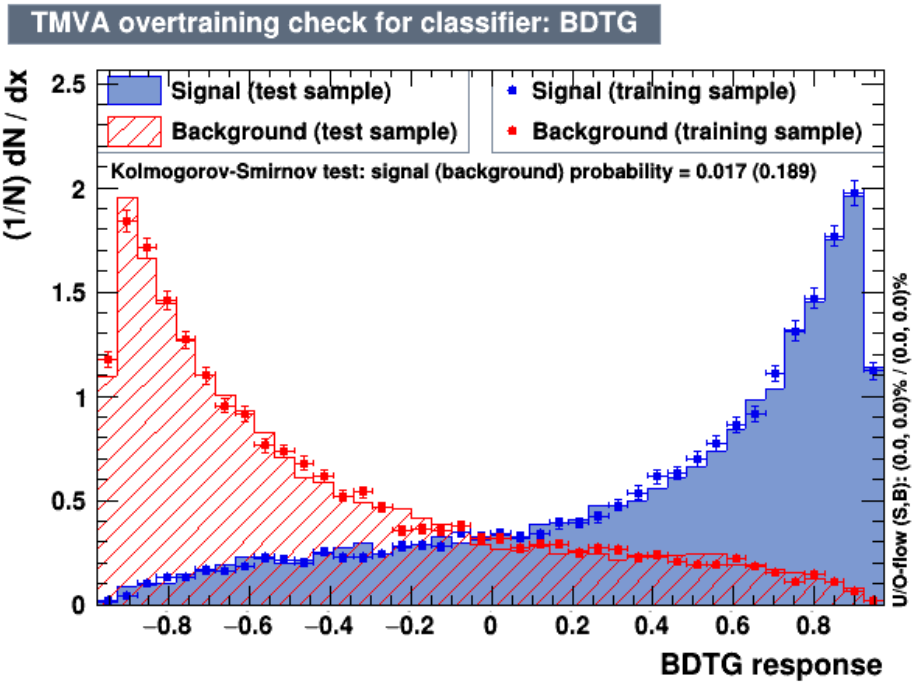


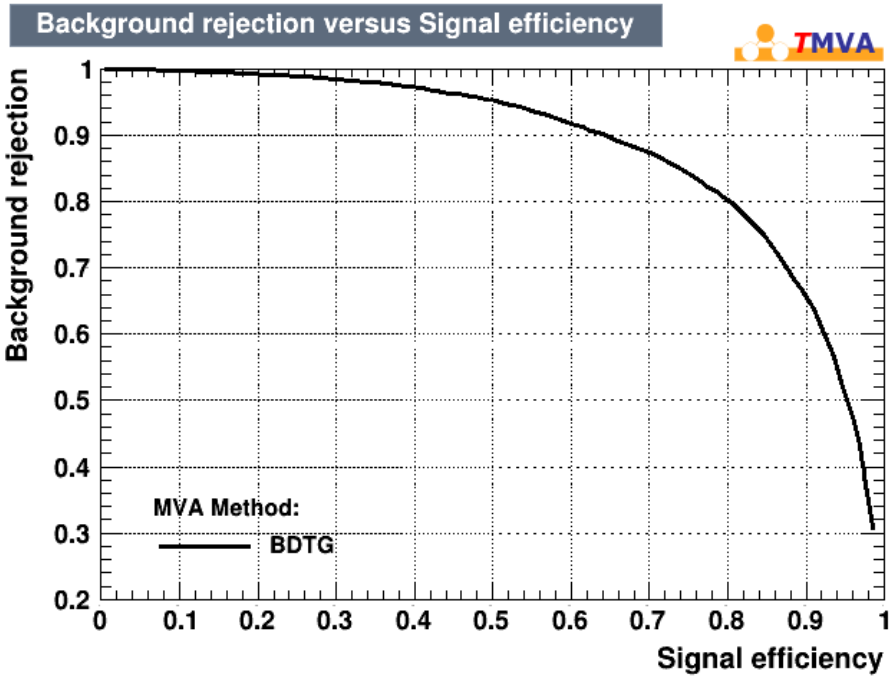




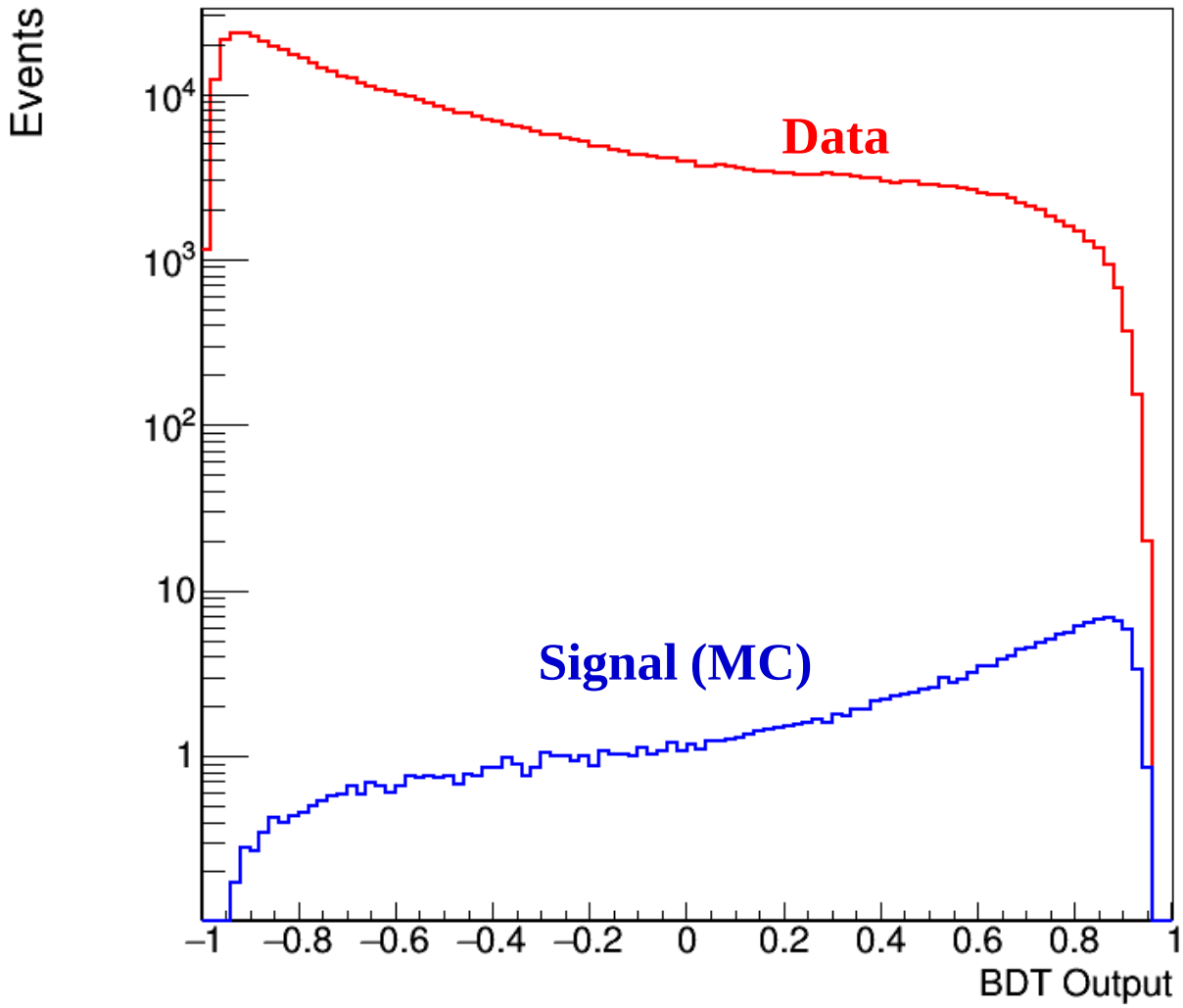


Signal vs Background BDT-response





BDT-Output


Events
10^4
10^3
10^2
10
1
Data
Signal (MC)
–1
–0.8
–0.6
–0.4
–0.2
0
0.2
0.4
0.6
0.8
1
BDT Output

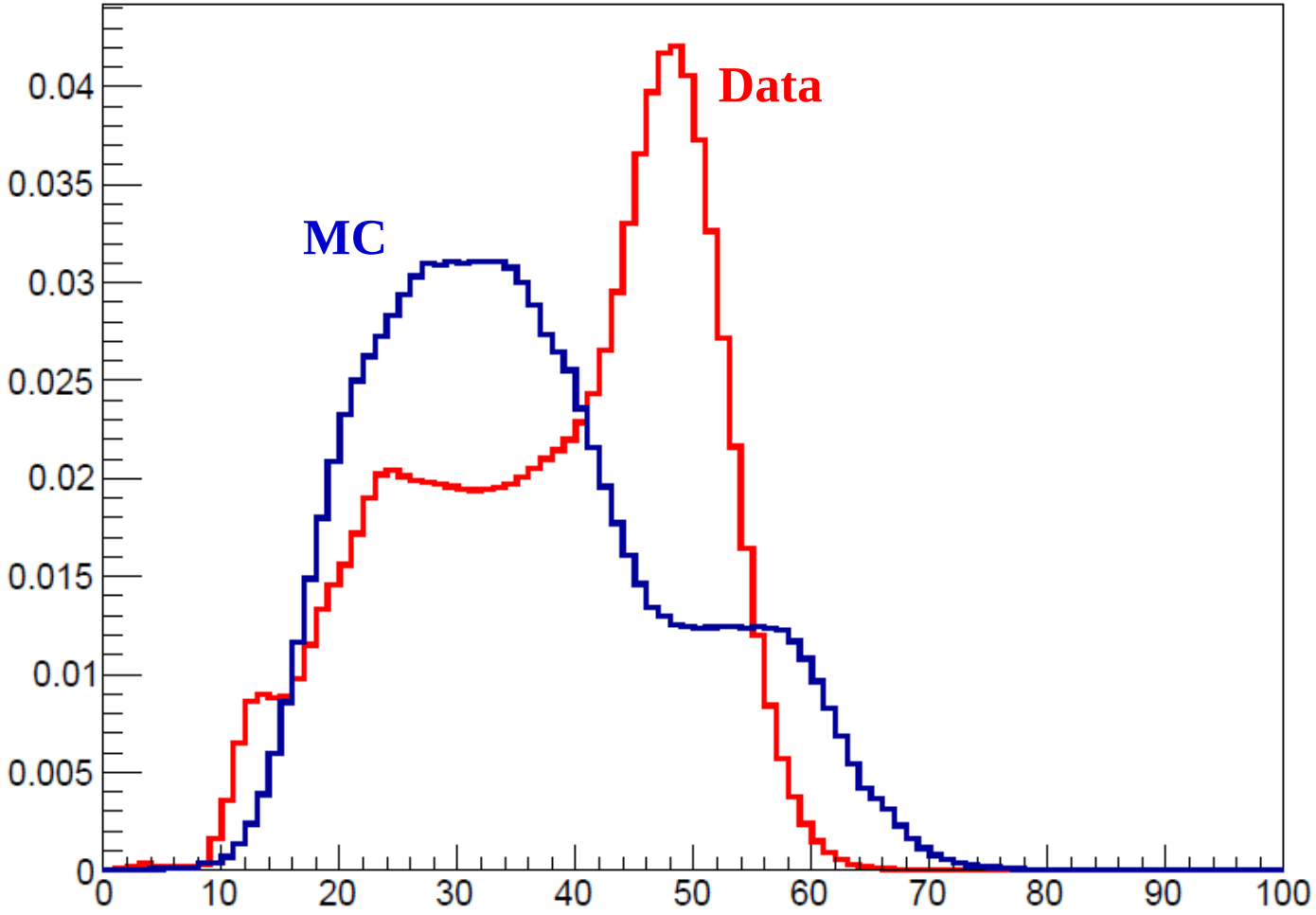
Plans for near future

- Optimization of Signal vs Background MVA algorithm
- MC (background) vs DATA comparison
- Implementation of DoubleBTag category
- Application of b-jet regression

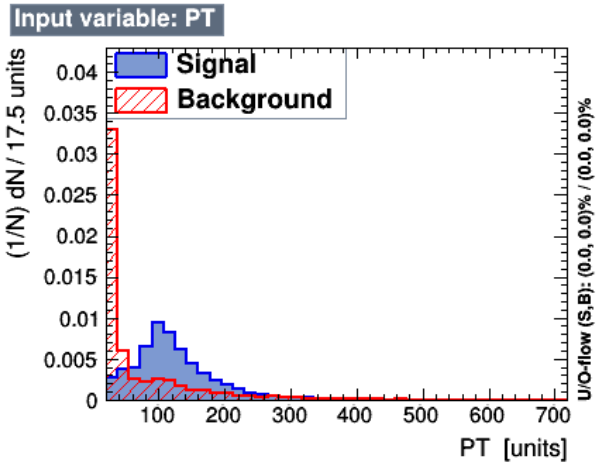
Thank You

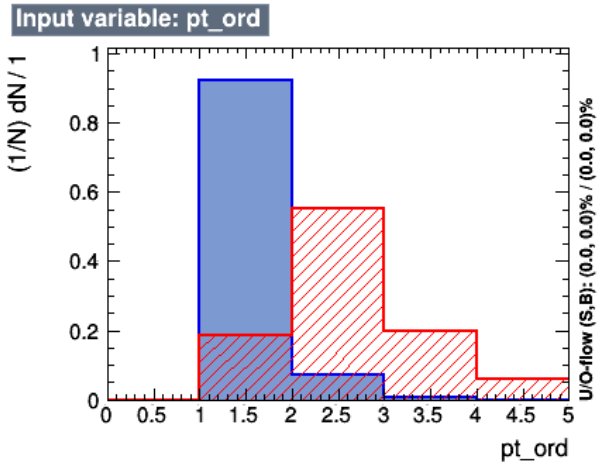
backup

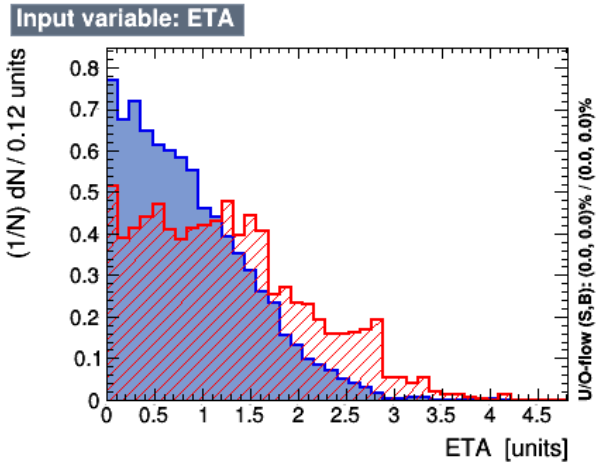
PileUp Reweighting

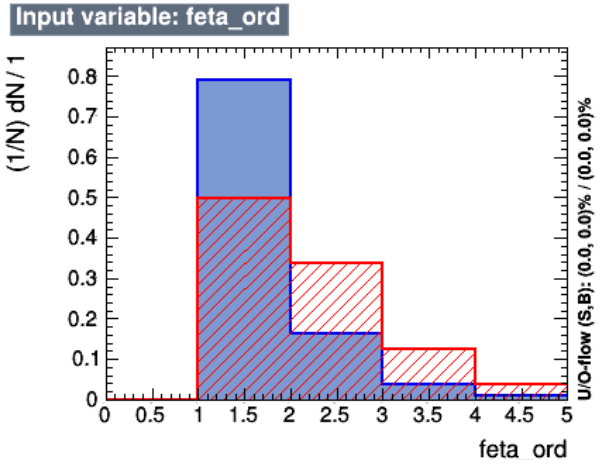


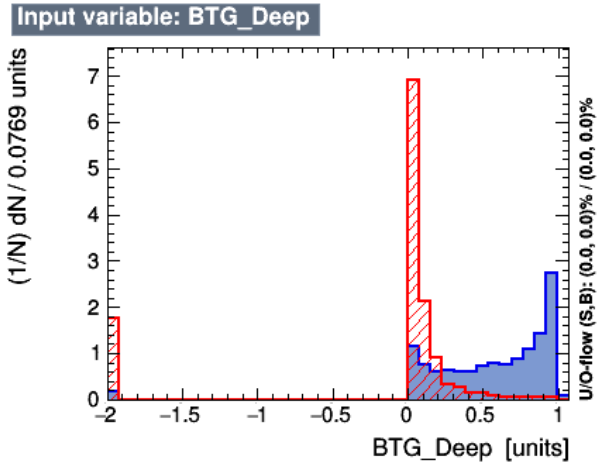
Instruction:
https://github.com/elliot-hughes/fatjet_analysis/wiki/Pileup-reweighting

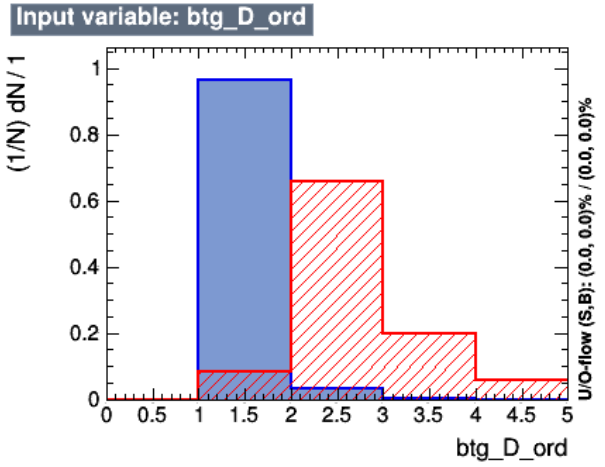


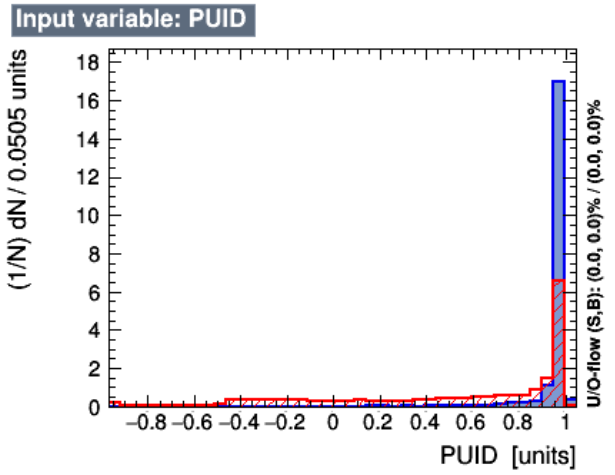






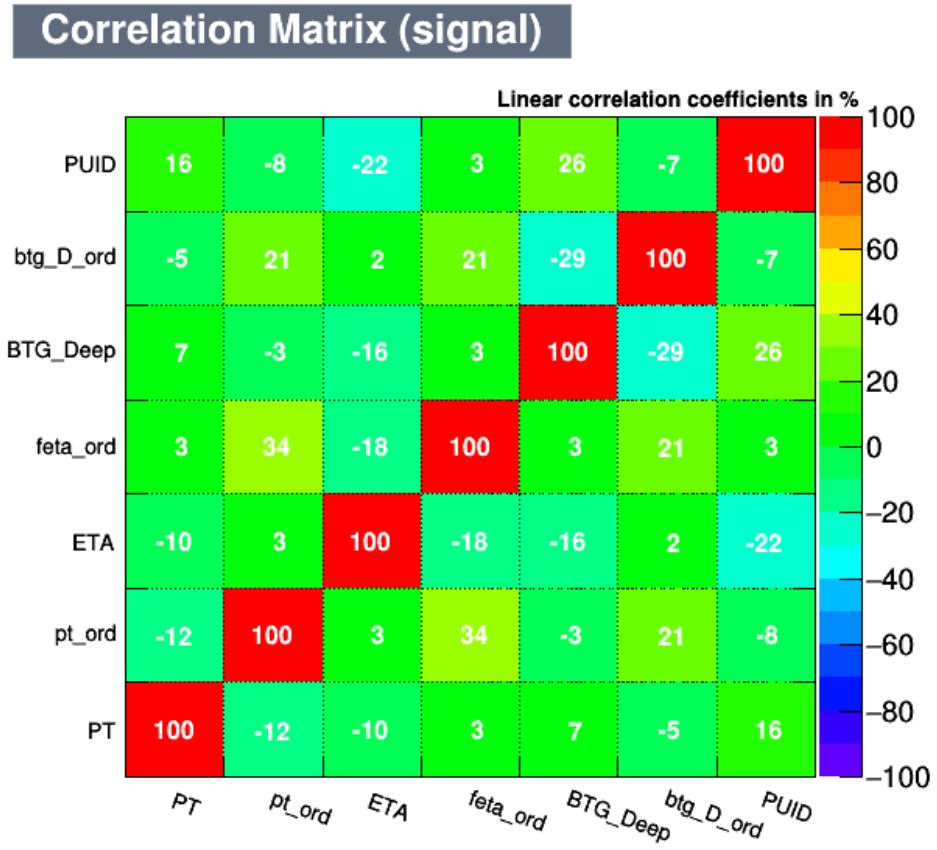


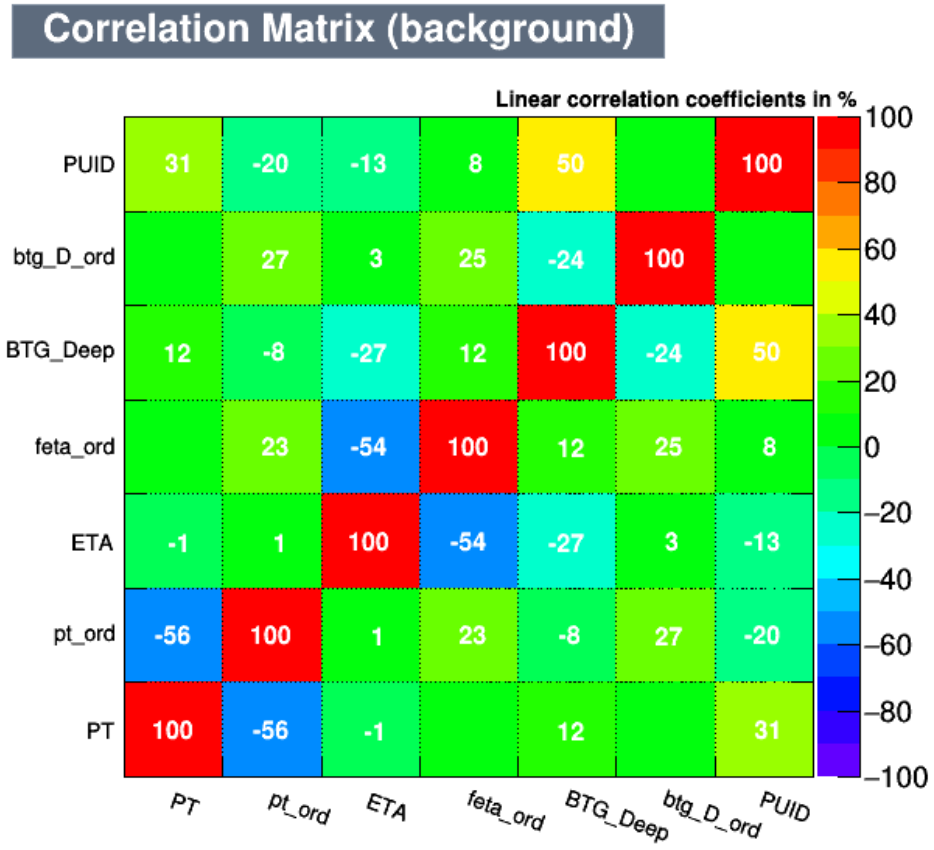


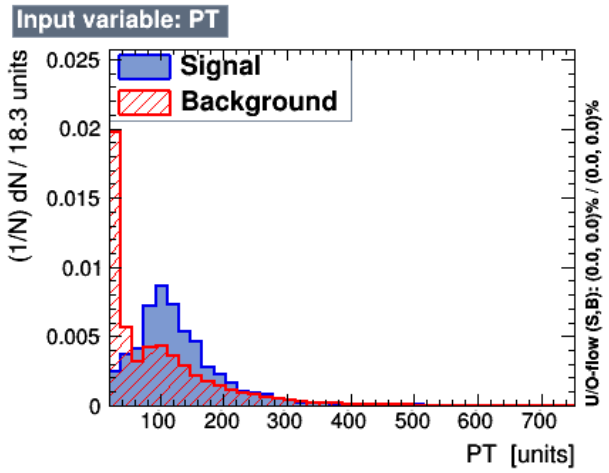


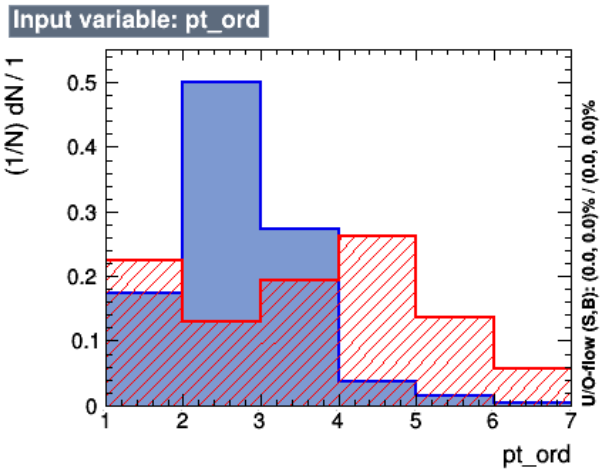
Input variables: 1-qq candidate

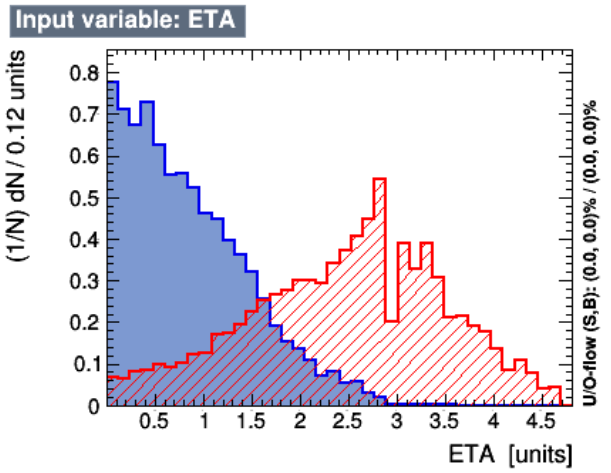
Input variables: 1-qq candidate

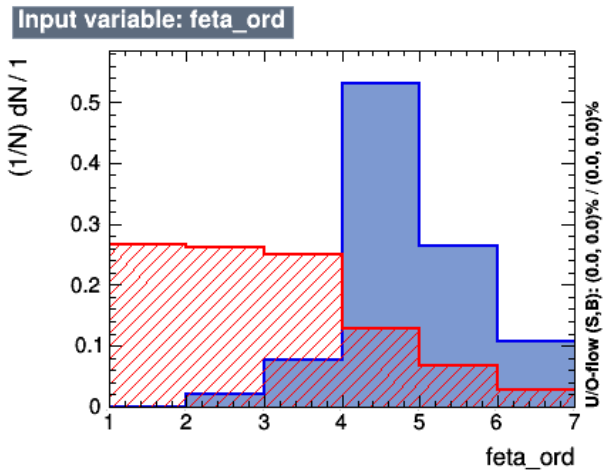


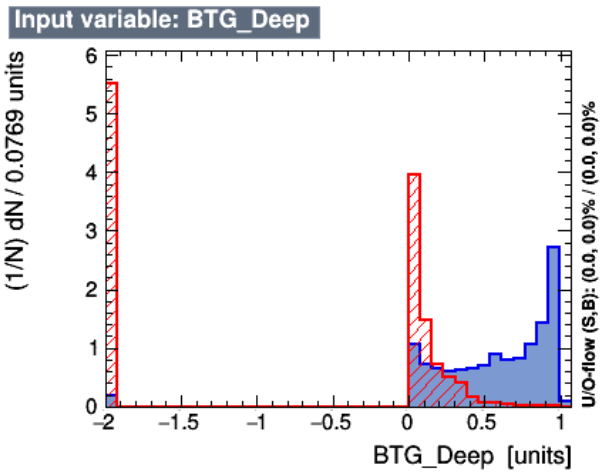


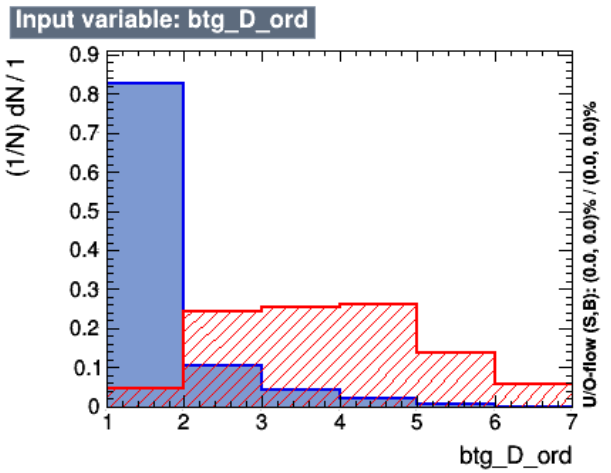


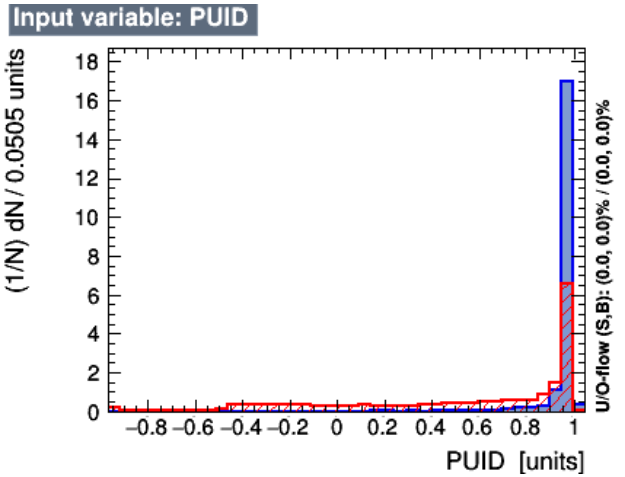






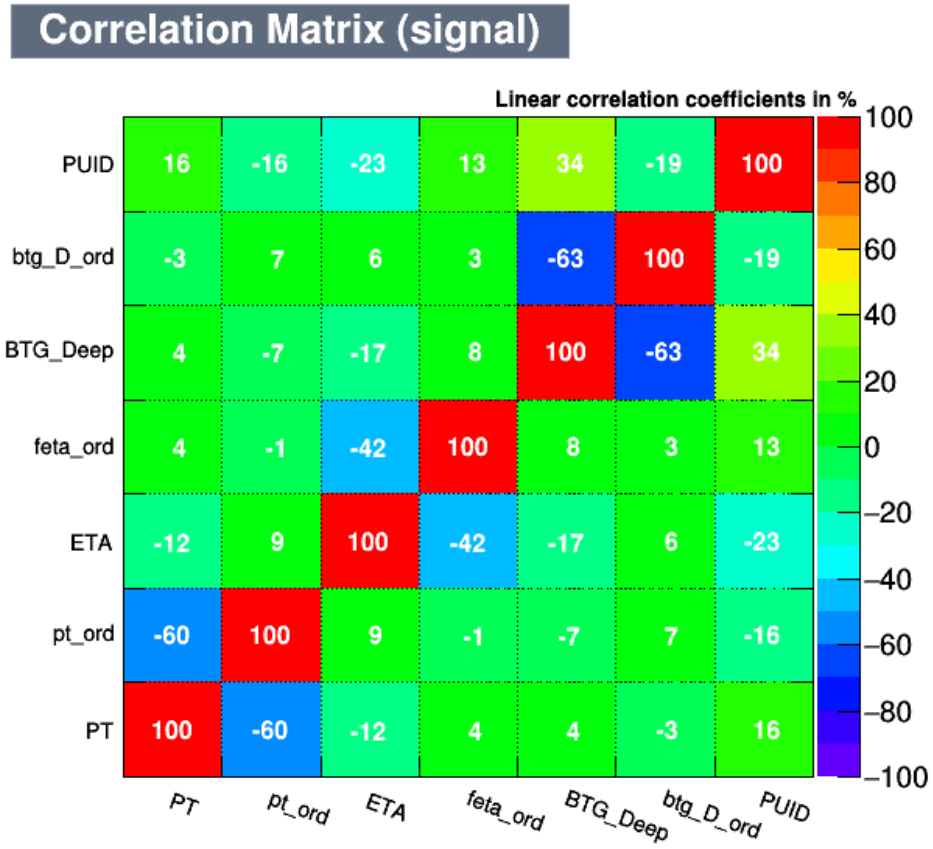


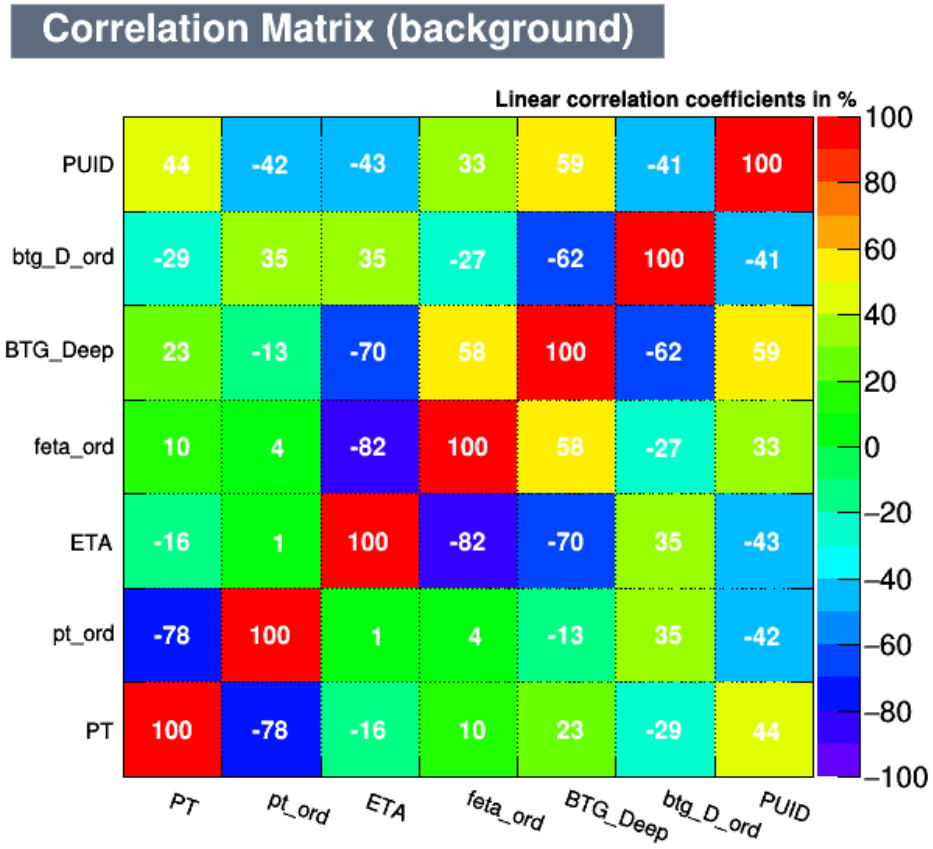




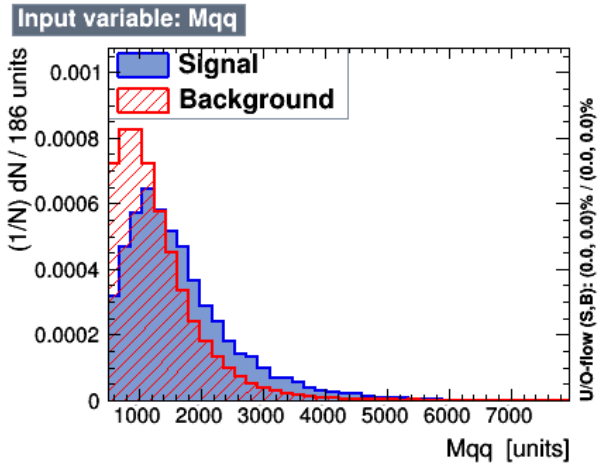
Input variables: > 1-qq candidate

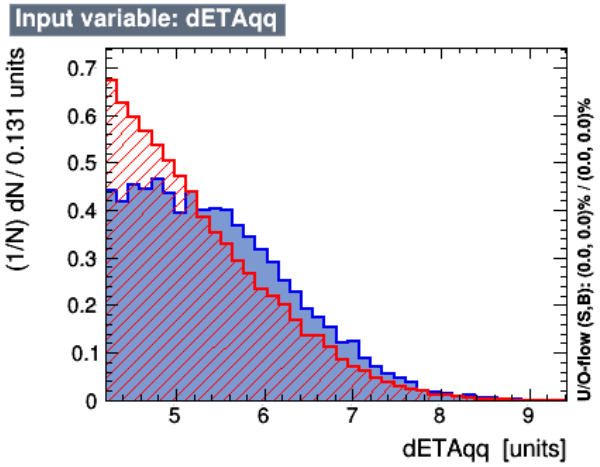
Input variables: > 1-qq candidate

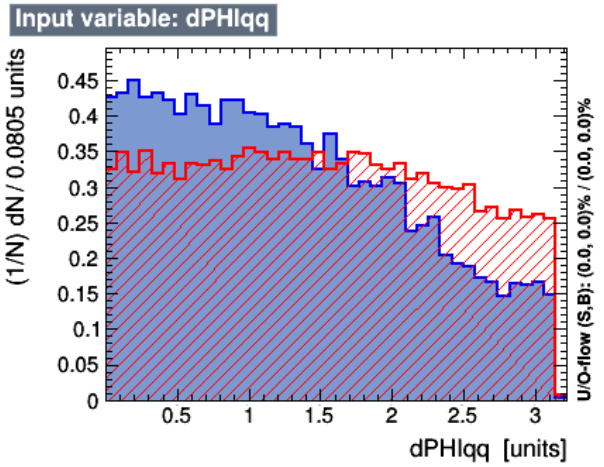


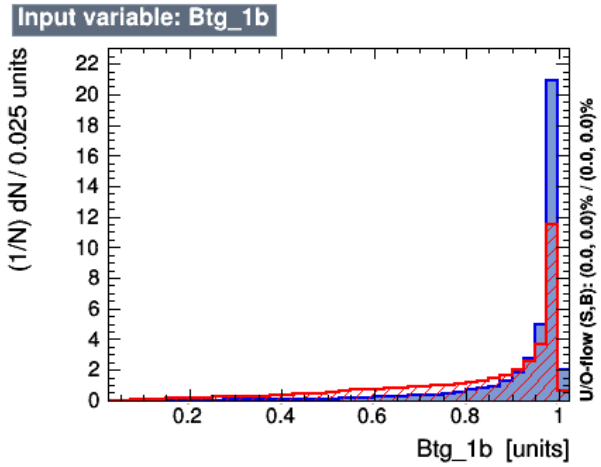


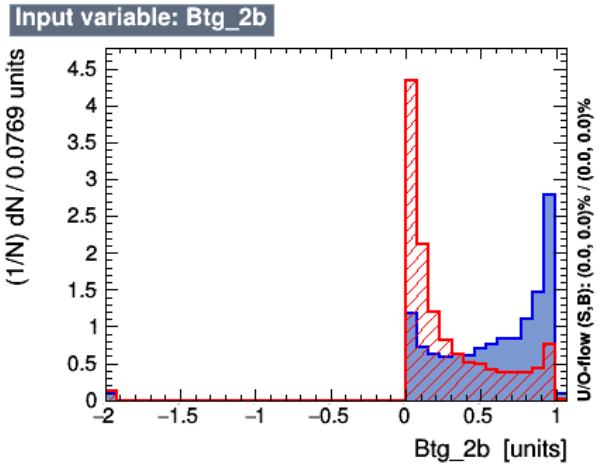
Signal vs Background input variables-1

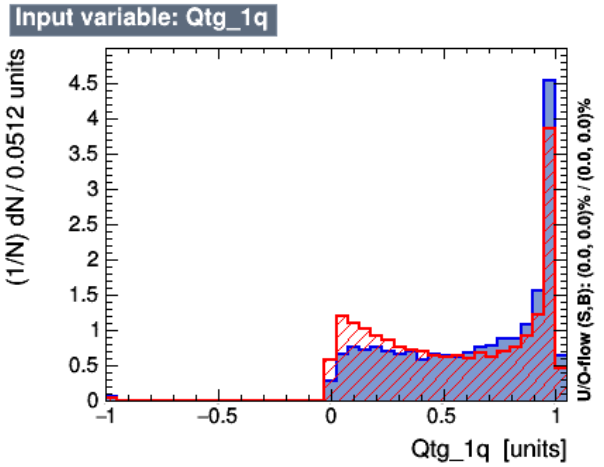




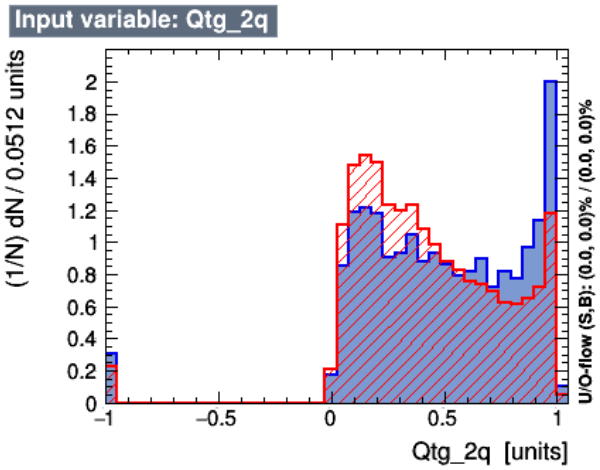


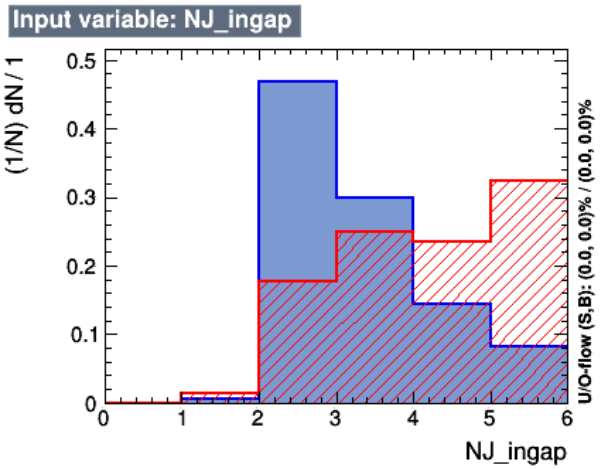


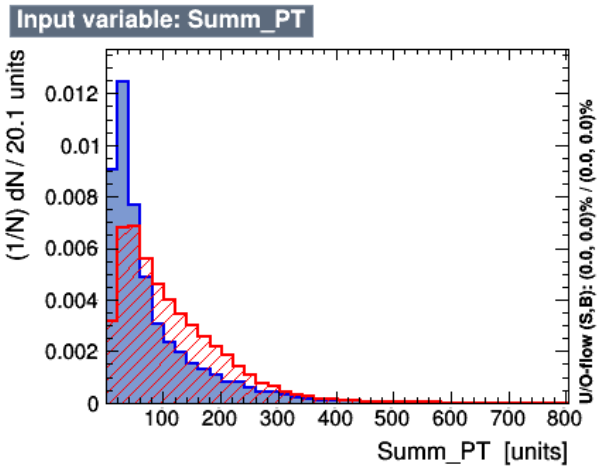


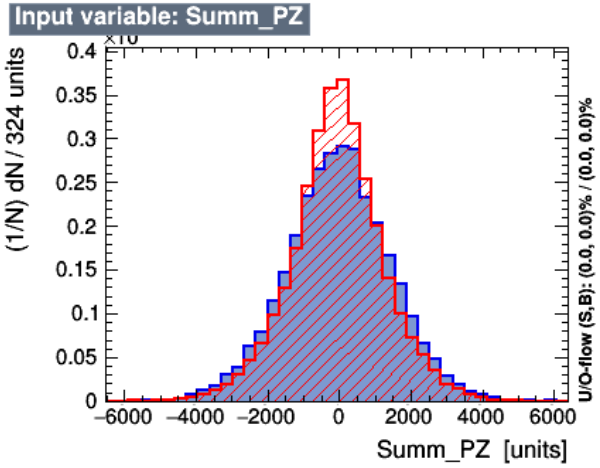


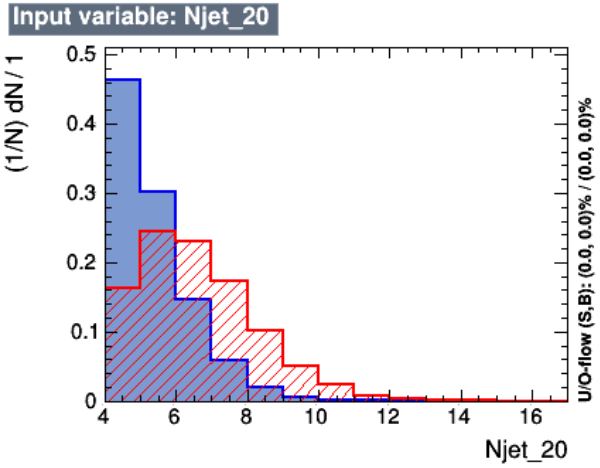
Signal vs Background input variables-2

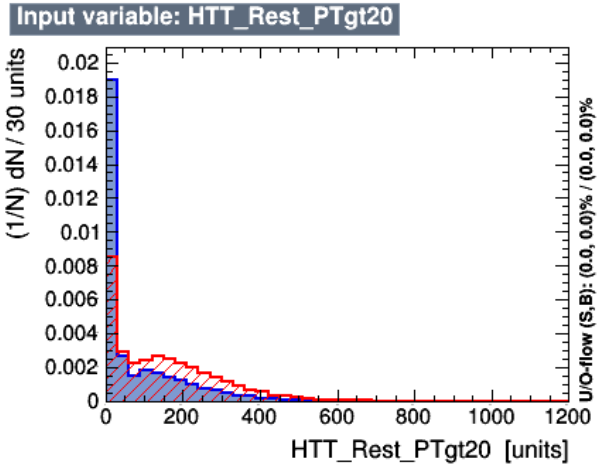




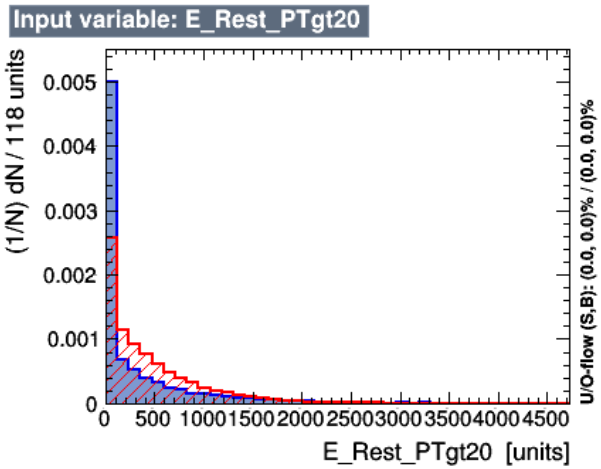




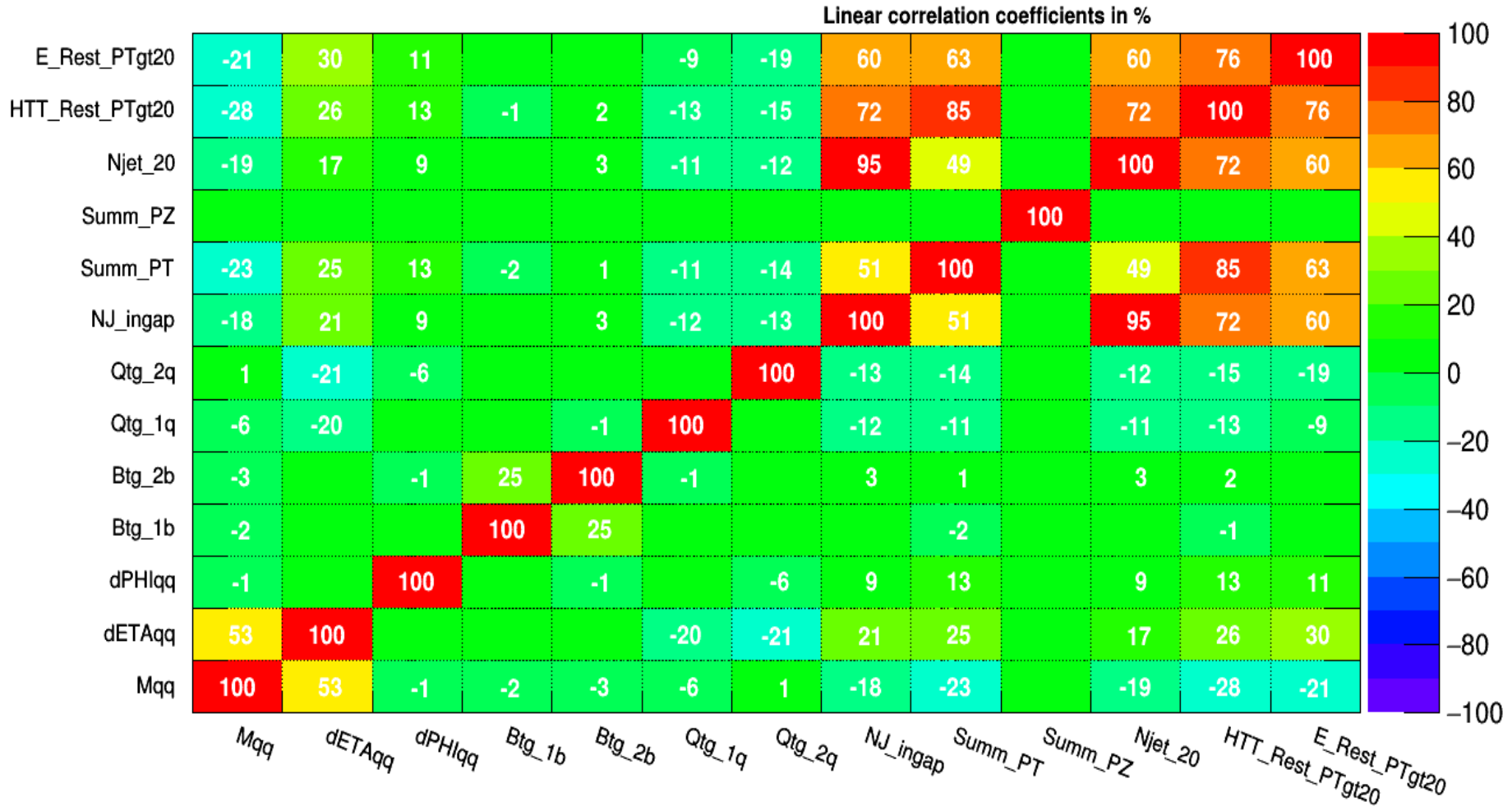




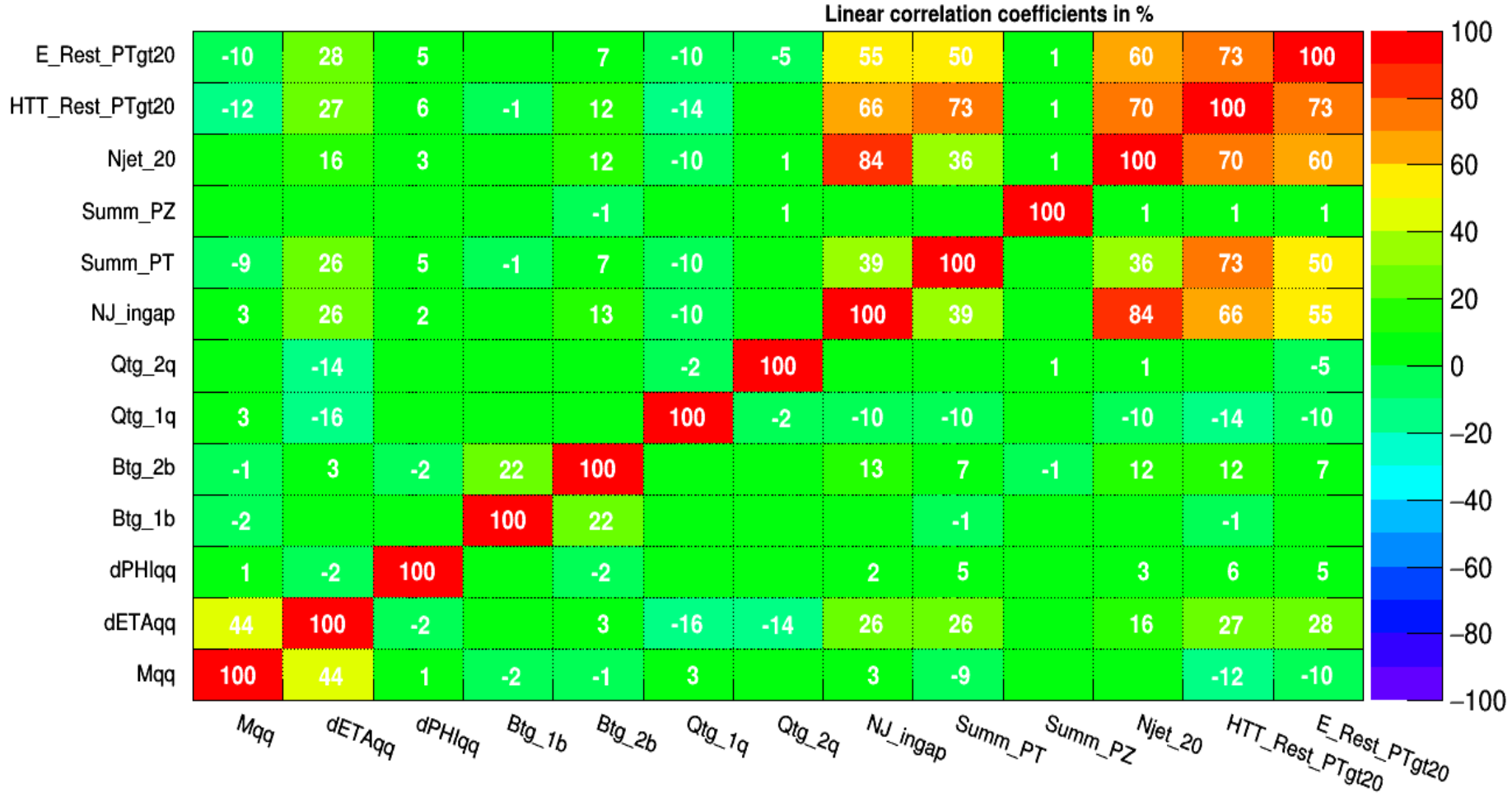
Signal vs Background input variables-3



Correlation Matrix (signal)



Correlation Matrix (background)



Signal_vs_Background

Rank	Variable	Separation	Variable Importance:
1	Btg_2b	2.043e-01	btgb2_D : 1.353e-01
2	Njet_20	1.725e-01	Njet_20 : 1.230e-01
3	NJ_ingap	1.516e-01	E_rest_20 : 9.194e-02
4	Btg_1b	1.279e-01	btgb1_D : 9.093e-02
5	HTT_Rest_PTgt20	1.127e-01	mqq : 9.067e-02
6	E_Rest_PTgt20	8.735e-02	pt_All : 8.804e-02
7	Summ_PT	8.059e-02	qtgq2 : 7.684e-02
8	Mqq	6.136e-02	pz_All : 6.751e-02
9	Qtg_2q	1.844e-02	qtgq1 : 6.610e-02
10	dETAqq	1.816e-02	detaqq : 6.483e-02
11	dPHIqq	1.729e-02	dphiqq : 6.342e-02
12	Qtg_1q	1.605e-02	HTT_rest_20 : 3.626e-02
13	Summ_PZ	9.426e-03	NJ_ingap : 5.118e-03