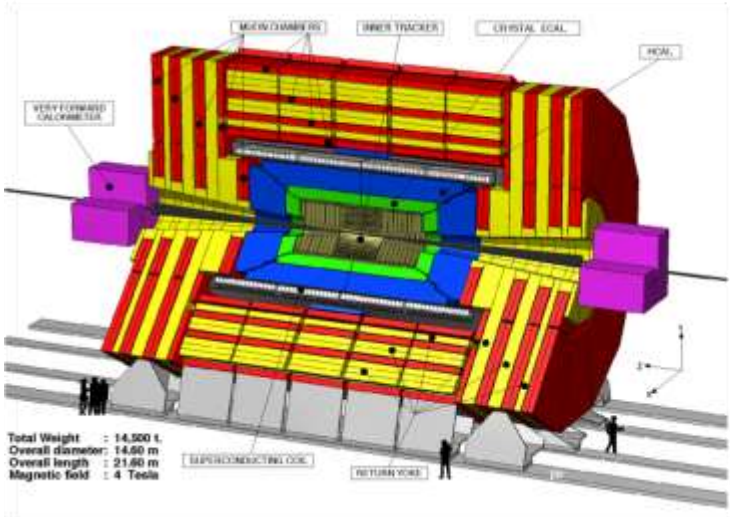
Search for Higgs boson produced in vector-boson fusion (VBF) process and decaying to bottom quarks with CMS (LHC).

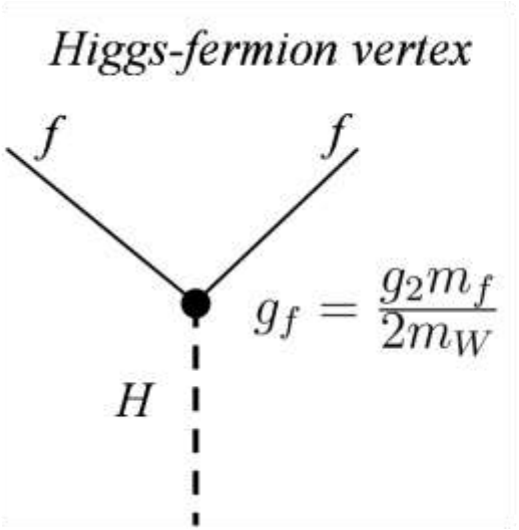
A. Tumasyan

Introduction

While with the LHC Run1 (2010-2012, $\sqrt{s}$ = 7, 8 TeV, L_{tot} ~ 30 fb^{-1}) Higgs boson was discovered (by ATLAS and CMS) in boson-decay channels, the LHC Run2 (2015-2018, $\sqrt{s}$ = 13 TeV, L_{tot} > 145 fb^{-1}) provide discoveries in fermion-decay channels. So far observations ($\geq 5\sigma$) were achieved in two fermion-decay channels: $H \to \tau^{-}\tau^{+}$, $H \to b\bar{b}$

$H \to b\bar{b}$ decay and Higgs-*b* coupling estimation are among the main CMS topics.



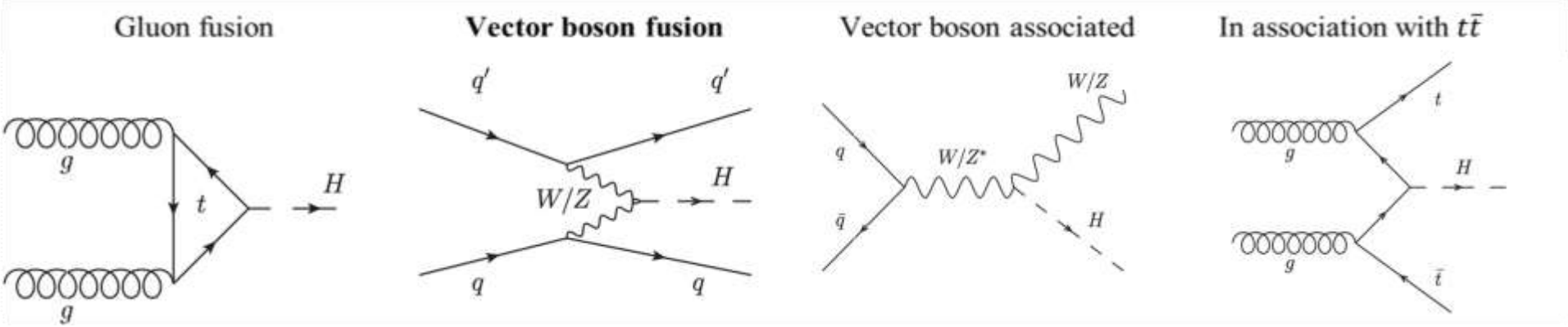


H→bb̄ features:

- branching ratio ~ 58 %
- direct access to *H-b* coupling
- difficulties with event reconstruction

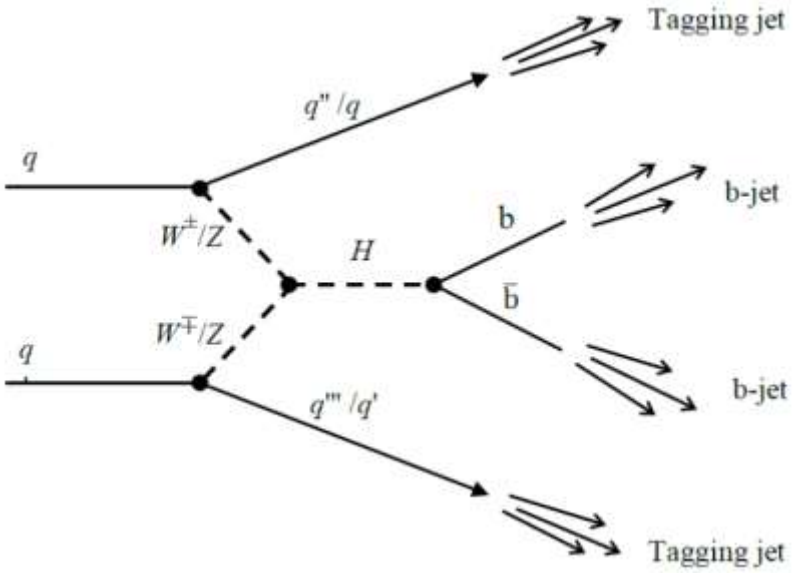
Introduction

Basic processes for Higgs boson production:



Features of VBF $H \to b\bar{b}$: ($\sigma \approx 2.2$ pb at 13 TeV)

- ❑ two *b*-quark jets from Higgs boson decay mainly in central region of CMS
- ❑ two light-quark jets from scattered quarks with large $\Delta\eta$ and inv. mass and forward-backward tend
- ❑ low additional hadron activity
- ❑ well suppressed background



Outline

- VBF $H \to b\bar{b}$ analysis strategy and status
- MC and Data
- Triggers and event preselection
- Pile-Up influence
- Jets identification and Higgs boson mass reconstruction
- Signal vs background discrimination
- Background model validation
- Data/MC comparison
- Upcoming works

Analysis strategy and status

This analysis is using 2017-2018 CMS(LHC) data.

Strategy

Two main categories: ***SingleBTag*** 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 with 2017, statistical analysis, 2018 is not implemented yet

DoubleBTag close to complete implementation with 2017, statistical analysis, 2018 is not implemented yet

MC and DATA samples

MC signal

- VBF_Hbb
- ggF_Hbb
- ttH_Hbb
- VH_Hbb

MC background

- QCD
- TTbar
- SingleTop
- DYJetsToQQ
- WJetsToQQ
- WW, WZ, ZZ

Data

2017F dataset corresponding to ~ 7.7 fb^{-1} integral luminosity.

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

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
- **Among 7 P_T-leading jets which passing Loose Puid**
  - 1 medium (0.82 Trigger, 0.8838 Offline) 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 $|\Delta\eta| > 4.2$ excluding the highest b-tagged jet
- **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)

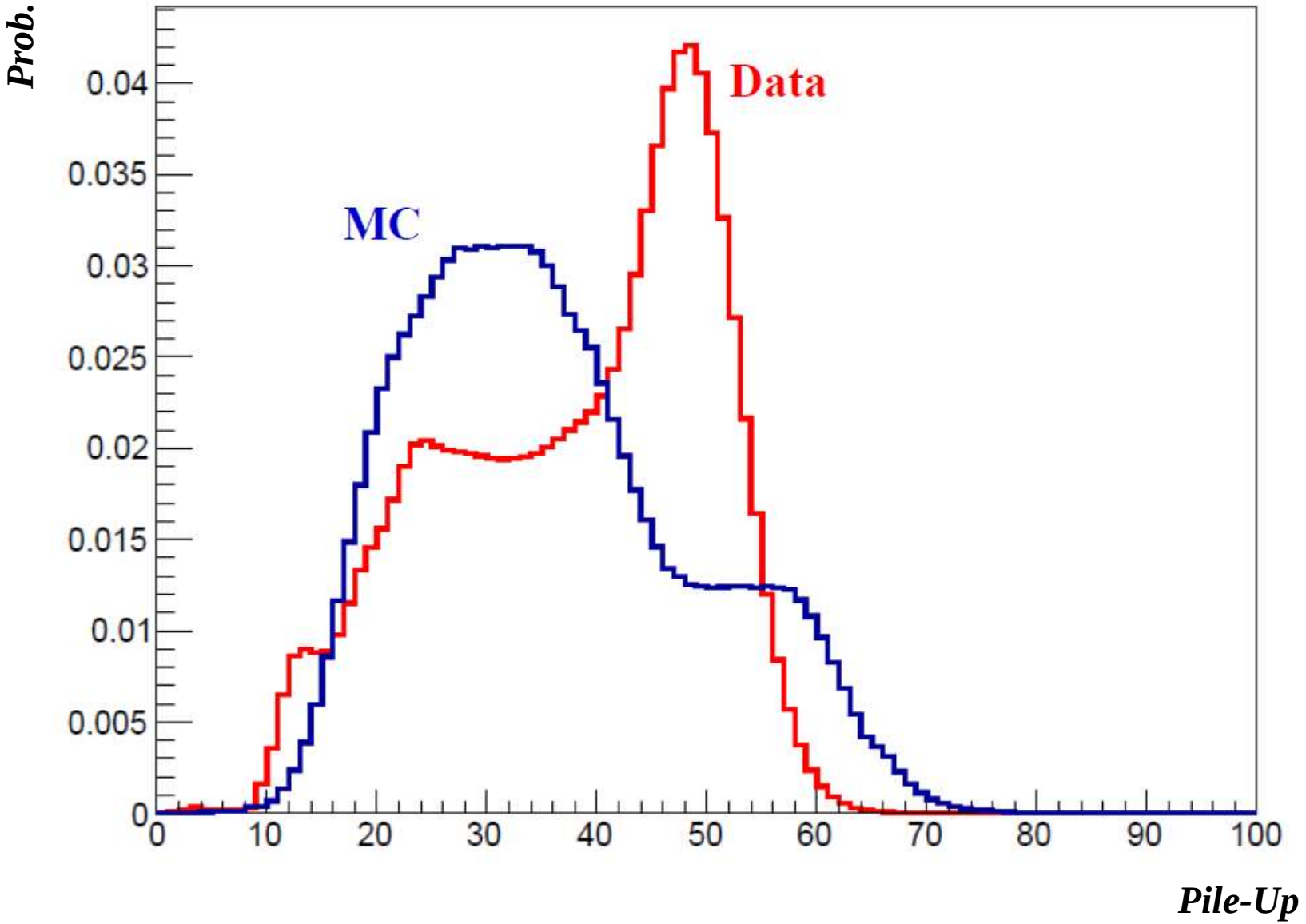
Signal efficiency ~ 1.2 %

DoubleBTag event preselection

- **Failed SingleBTag Preselection ! ! !**
- **HLT_QuadPFJet105_90_76_15_DoubleBTagCSV_p013_p08_VBF1_v5** OR HLT_QuadPFJet111_90_80_15_DoubleBTagCSV_p013_p08_VBF1_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
- **Among 7 P_T-leading jets which passing Loose Puid**
  - 1 medium (0.82 Trigger, 0.8838 Offline) and 1 loose (0.47 Trigger, 0.5803 Offline) b-tagged jets with $P_T > 20$ GeV and $|\Delta\varphi| < 2.1$.
    b-tagging with "pfCombinedInclusiveSecondaryVertexV2BJetTags"
    https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X
  - a pair of jets with Minv > 250 GeV and $|\Delta\eta| > 2.3$ excluding the b-tagged jets
- **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)

Signal efficiency ~ 0.7 %

Pile-Up Reweighting



Jet identification

Some procedure to identify 2 jets from Higgs decay (b-jets) and 2 jets form scattered quarks (tagging jets) among 7 pt-leading preselected jets should be done.

DoubleBTag category:

- 2 jets with highest b-tagging parameter → jets from Higgs decay (**b-jets**)
- 2 jets from the rest with the highest $\Delta\eta$ and $M_{inv} > 250$ GeV → tagging jets (**q-jets**)

SingleBTag category:

First idea (2015 analysis approach)

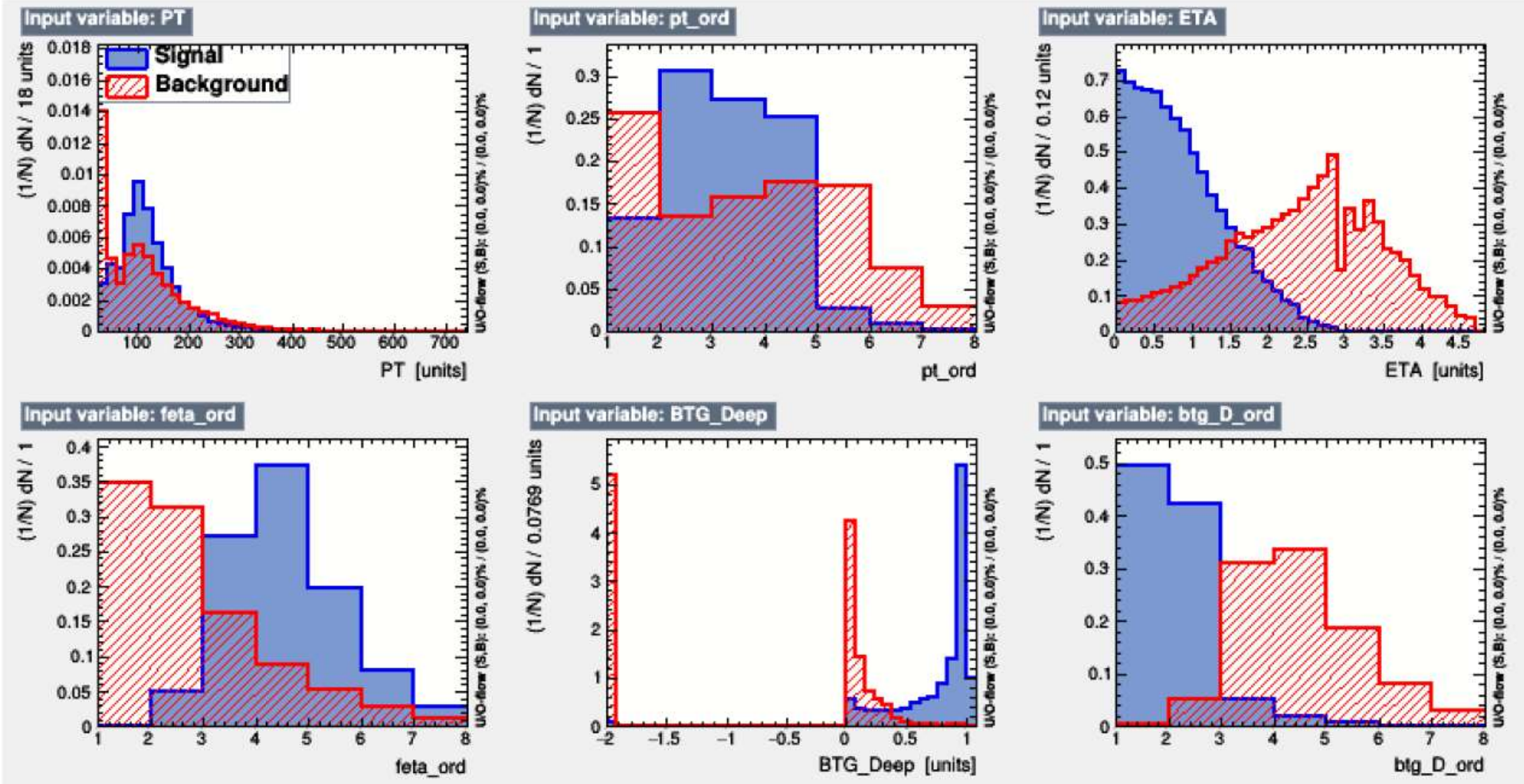
- 2 jets with highest b-tagging parameter → jets from Higgs decay (**b-jets**)
- 2 jets from the rest with the highest $\Delta\eta$ and $M_{inv} > 500$ GeV → tagging jets (**q-jets**)

But efficiency of this procedure can be essentially increased with MultiVariate Analysis (**MVA**). MVA can use not only b-tagging information, but also other parameters like kinematics, q/g-tagging information, etc. (2016, 2017-2018 approach)

Jet identification (b-jets vs q-jets)

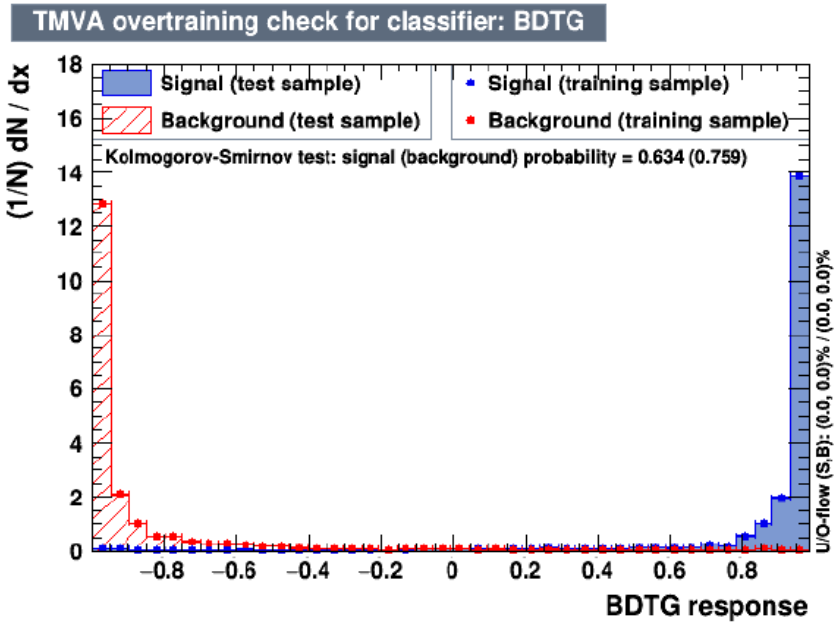
VBF_Hbb process, SingleBTag category:

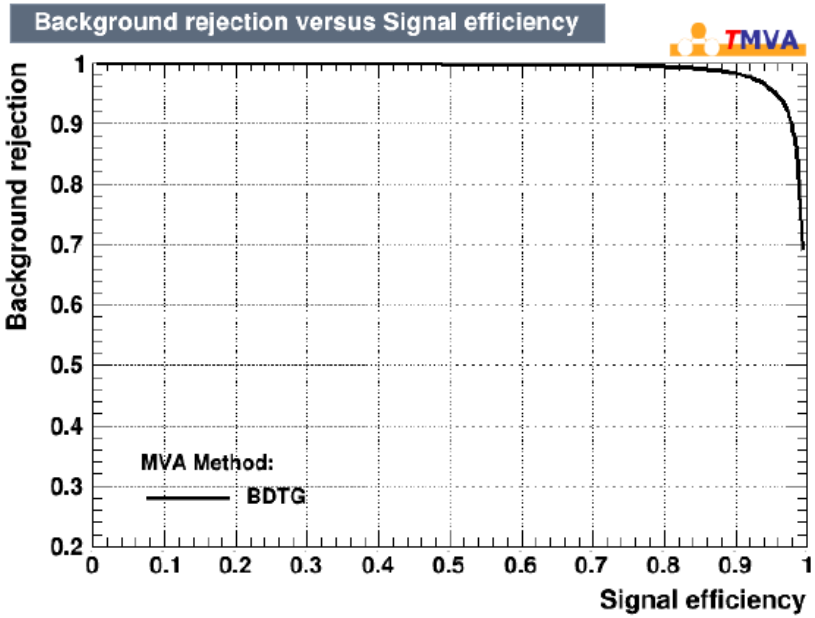
- b-jets from Higgs decay
- q-jets (mainly from scattered quarks)



Jet combination

SingleBTag: (MVA method → *Gradient* BDT)





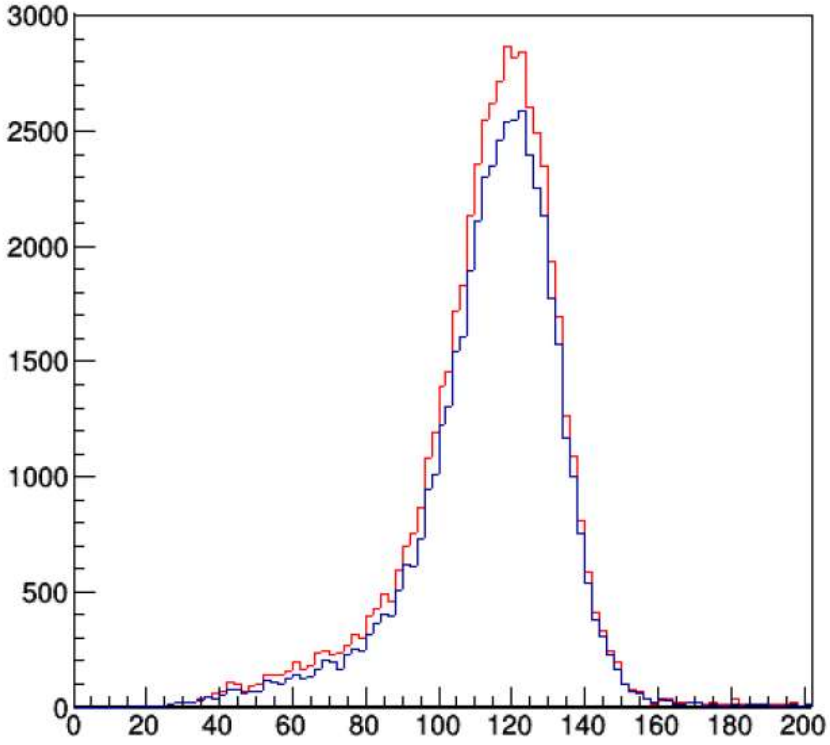
- ❑ 2 jets with the highest BDT-out → **b-jets**
- ❑ 2 jets from the rest with the highest Δη and M_{inv} > 500 GeV → **q-jets**

DoubleBTag:

- ❑ 2 jets with the highest b-tagging discriminator value → **b-jets**
- ❑ 2 jets from the rest with the highest Δη and M_{inv} > 250 GeV → **q-jets**

Signal efficiency (SingleBTag)

b-jets 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.



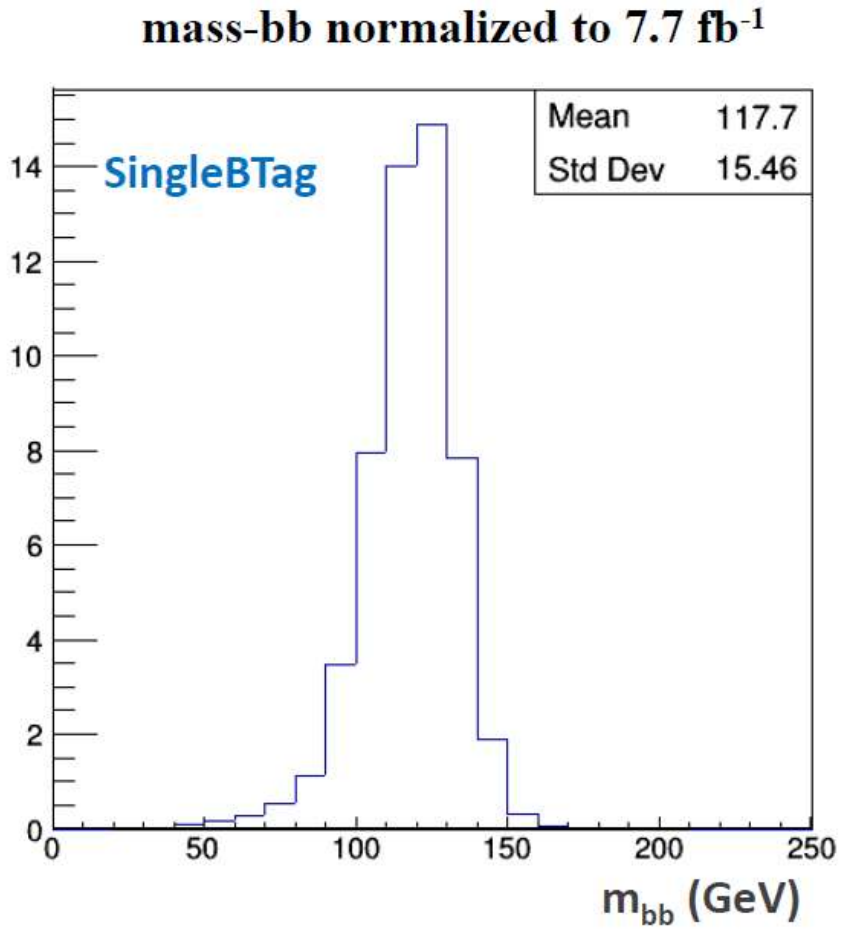
Purity

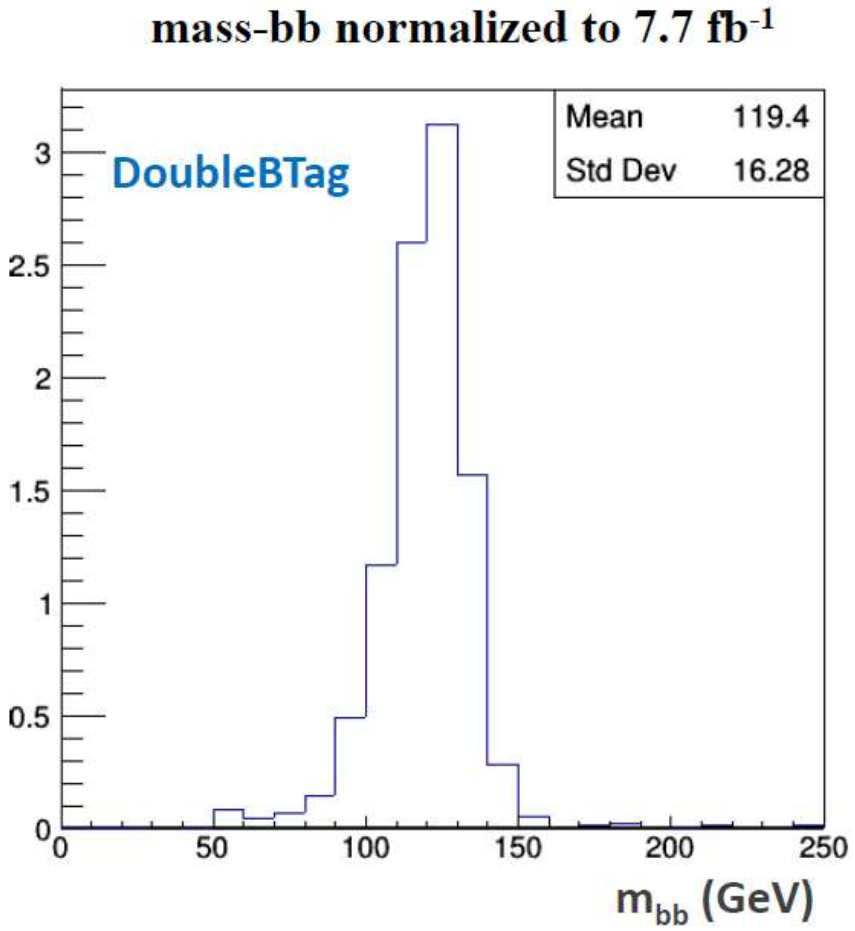
TMVA: ~ 90 %
High_bDisc ~ 91 %

Resulting signal efficiency ~ **1.2 %**

Expected number of events at ~ 7.7 fb^{-1} is ~ **200**

Jet combination (invariant mass)





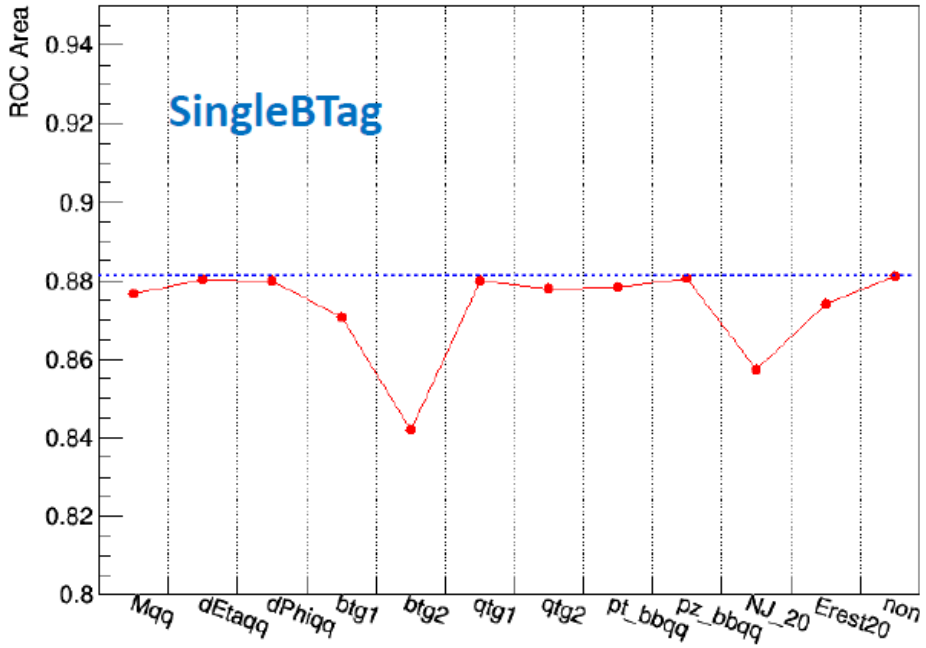
Signal vs Background with MVA

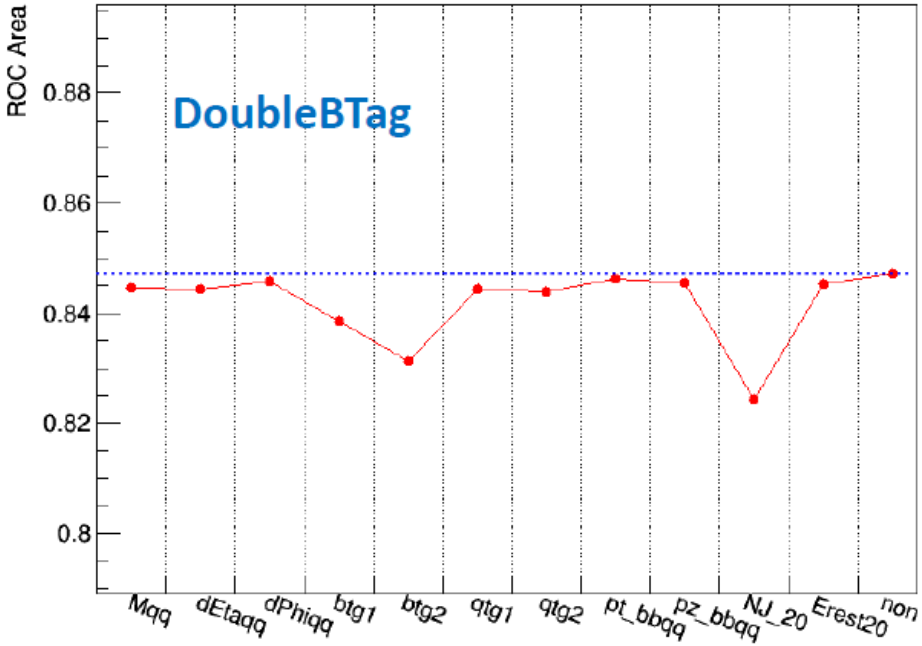
- **As a background ~ 5 % of Data was used**
- **Gradient BDT was used as MVA-method**
- **As the MVA inputs the following components are used:**

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

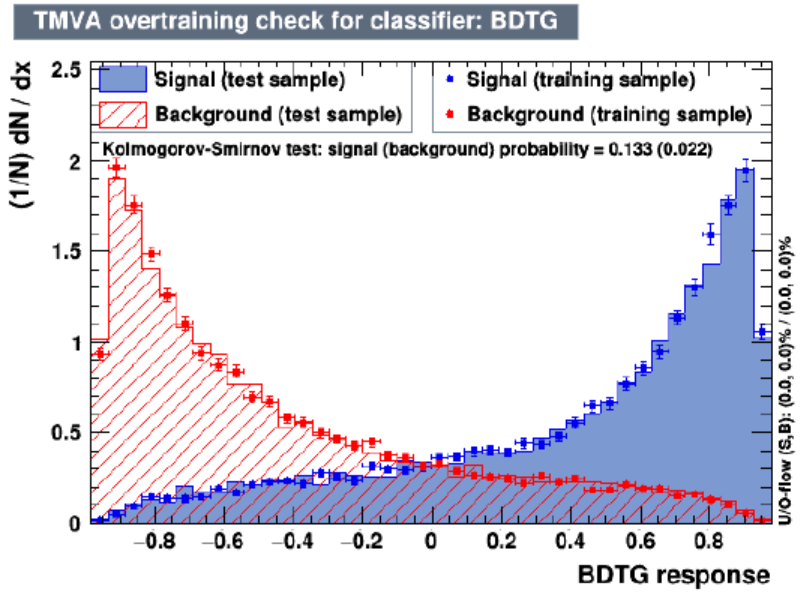
(*optimized with* **N-1 test** *from large variety of possible inputs*)

N-1 test for remaining inputs

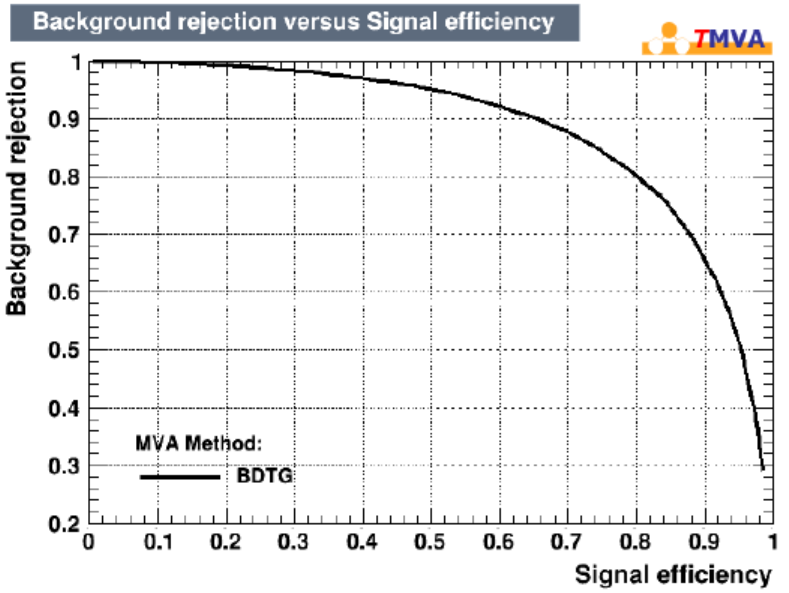


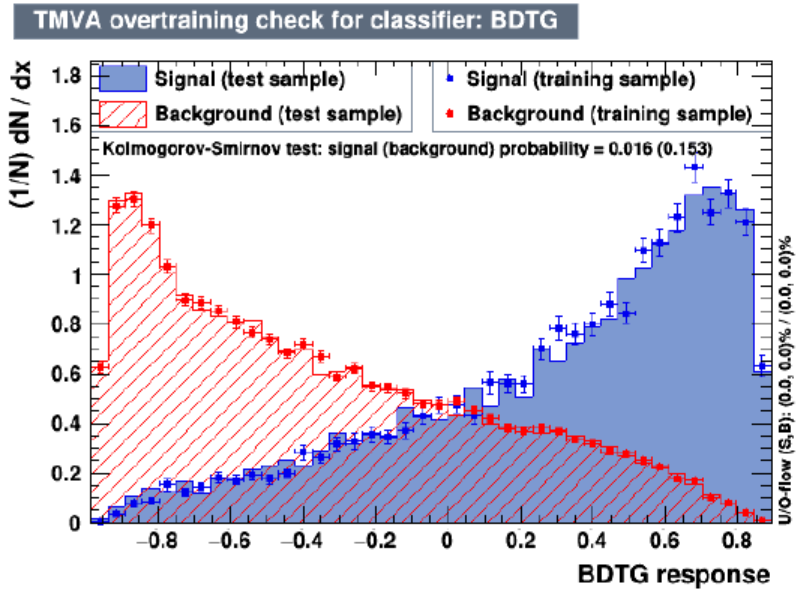


Signal vs Background BDT-responce

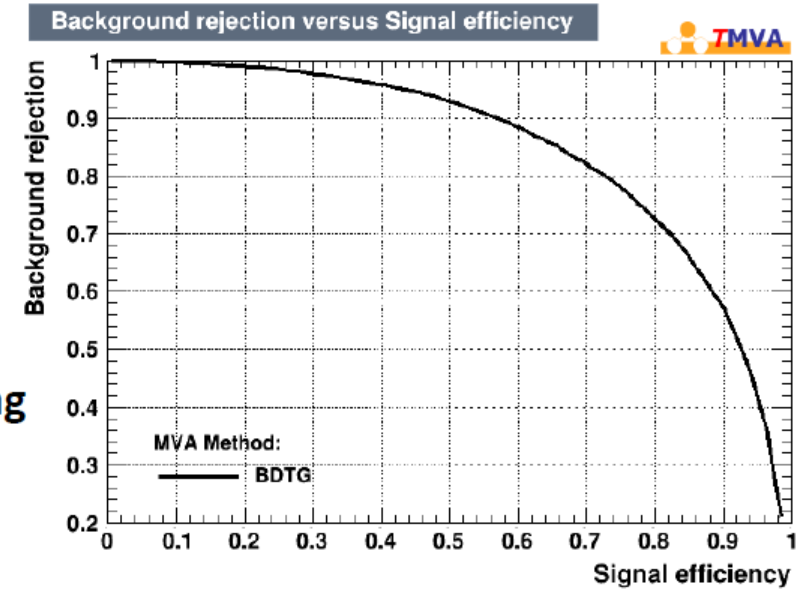


SingleBTag

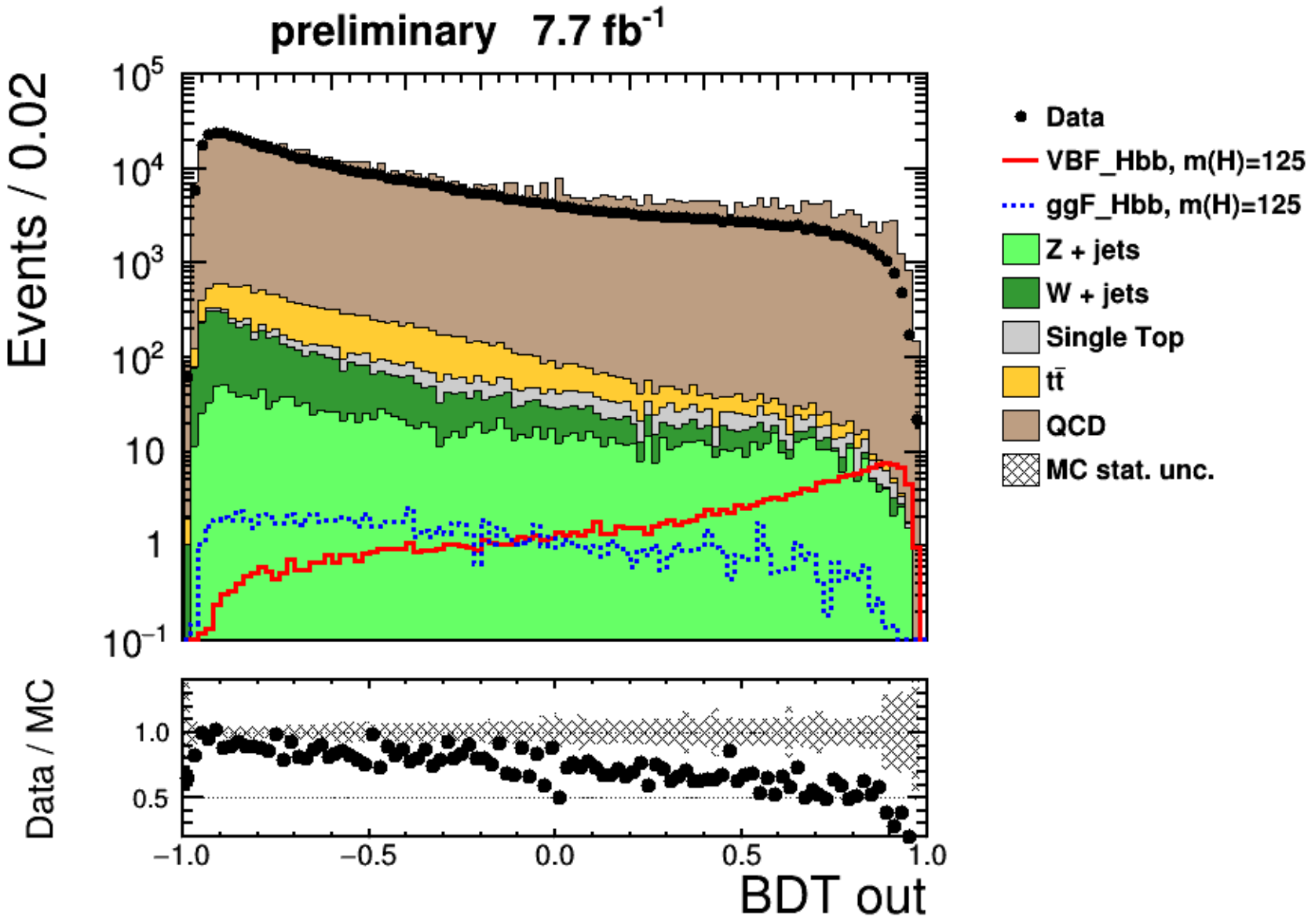




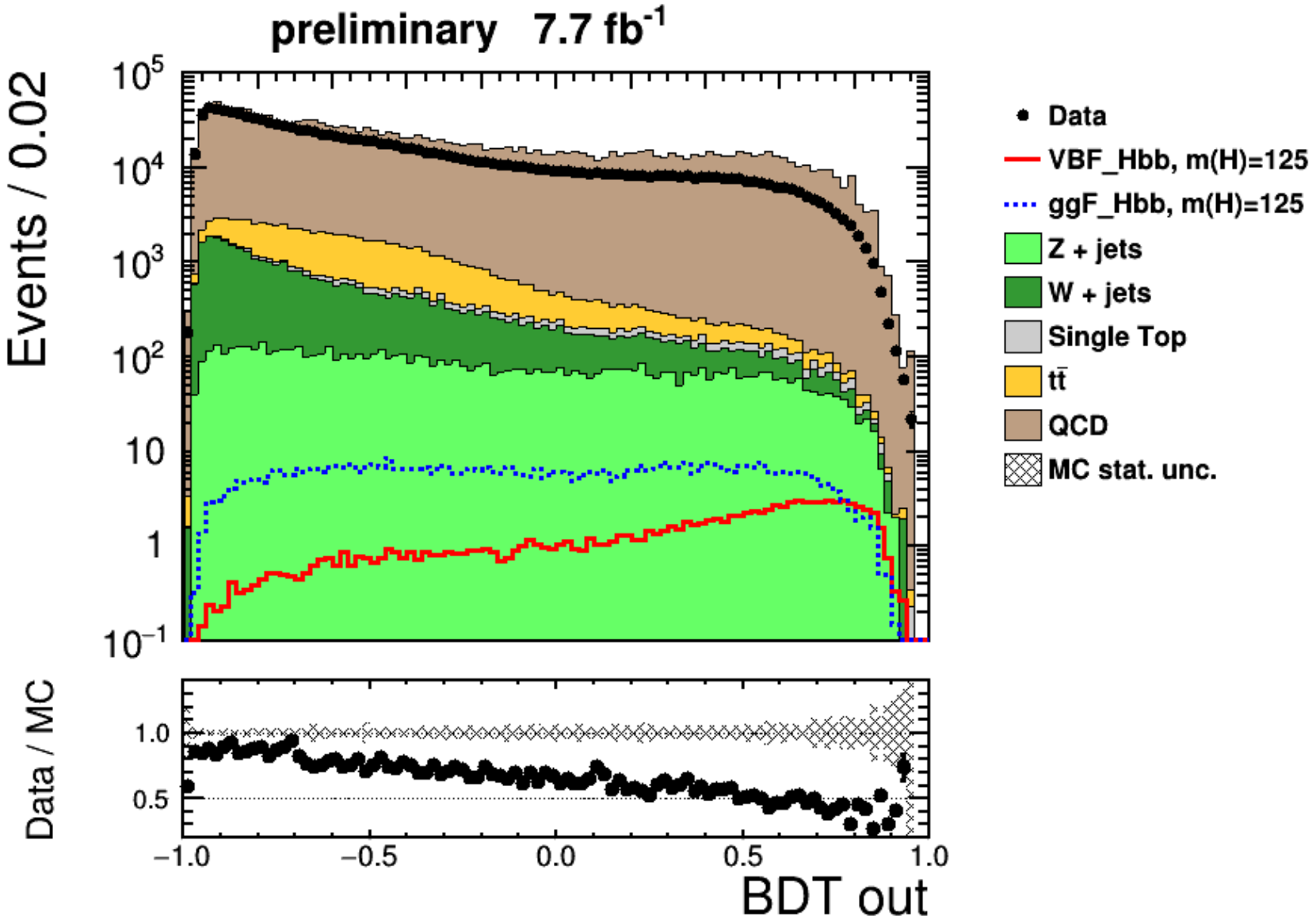
DoubleBTag



BDT-Output (SingleBTag)



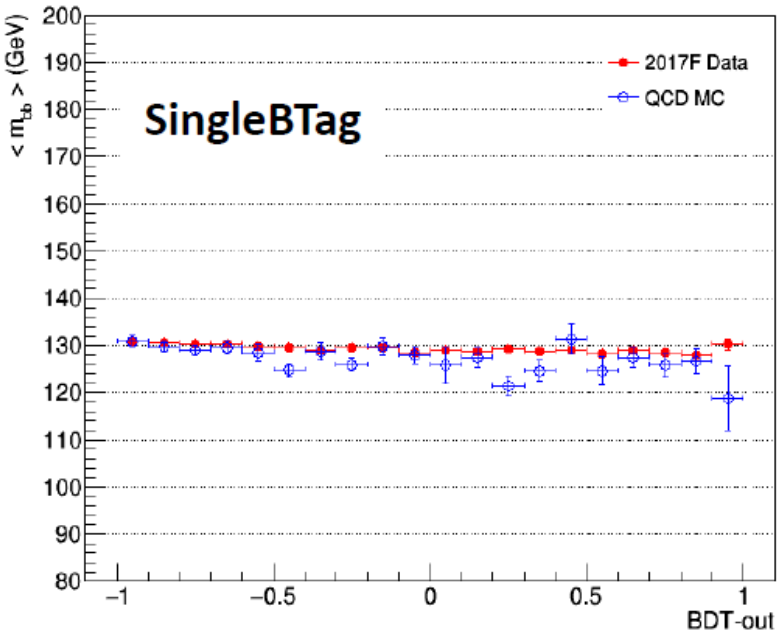
BDT-Output (DoubleBTag)

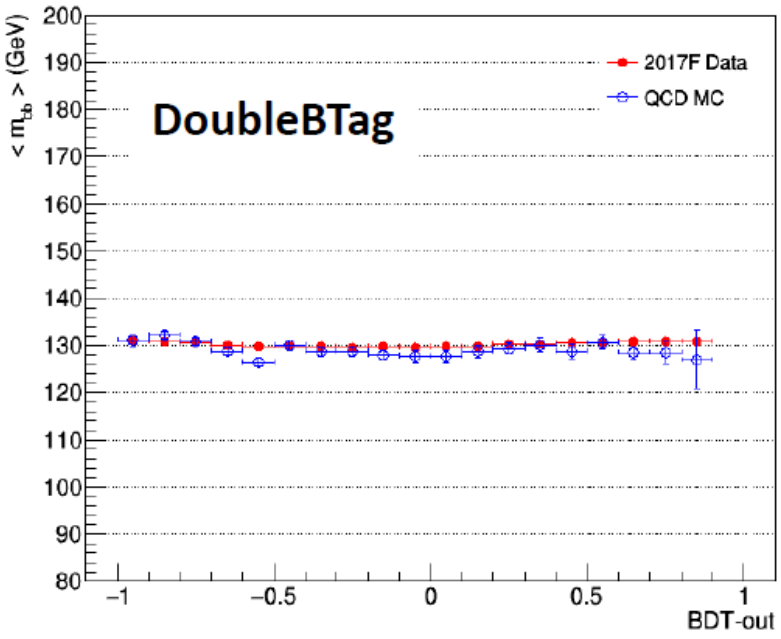


Background model validation-1 ($< m_{bb} >$ - correlation)

Dependence of *b*-jets invariant mass (mean value) from BDT-response in mass window 80-200 GeV.

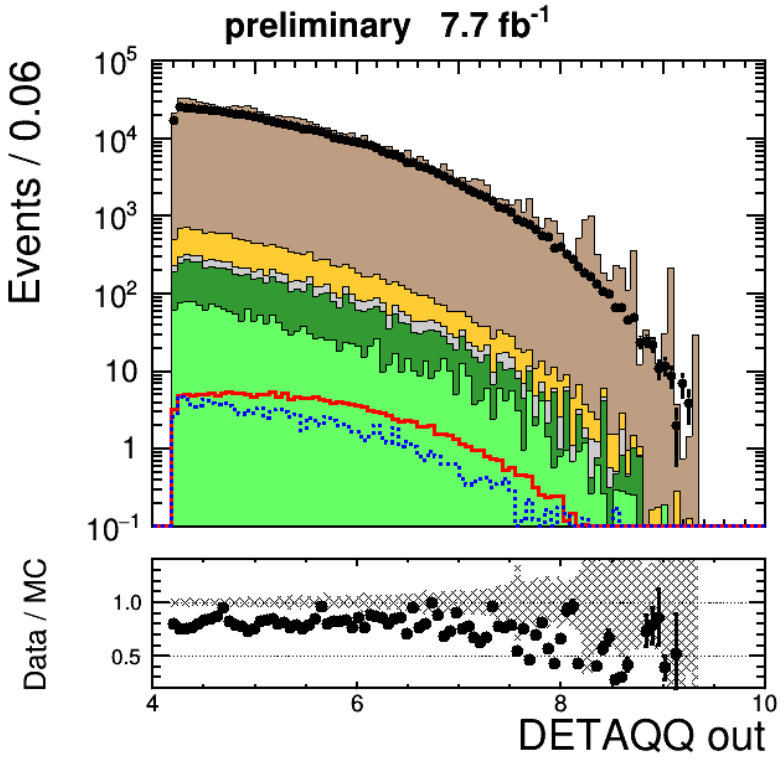
$< m_{bb} >$ vs BDT-out **(2017F Data ~7.7 fb^{-1} / QCD-MC)**

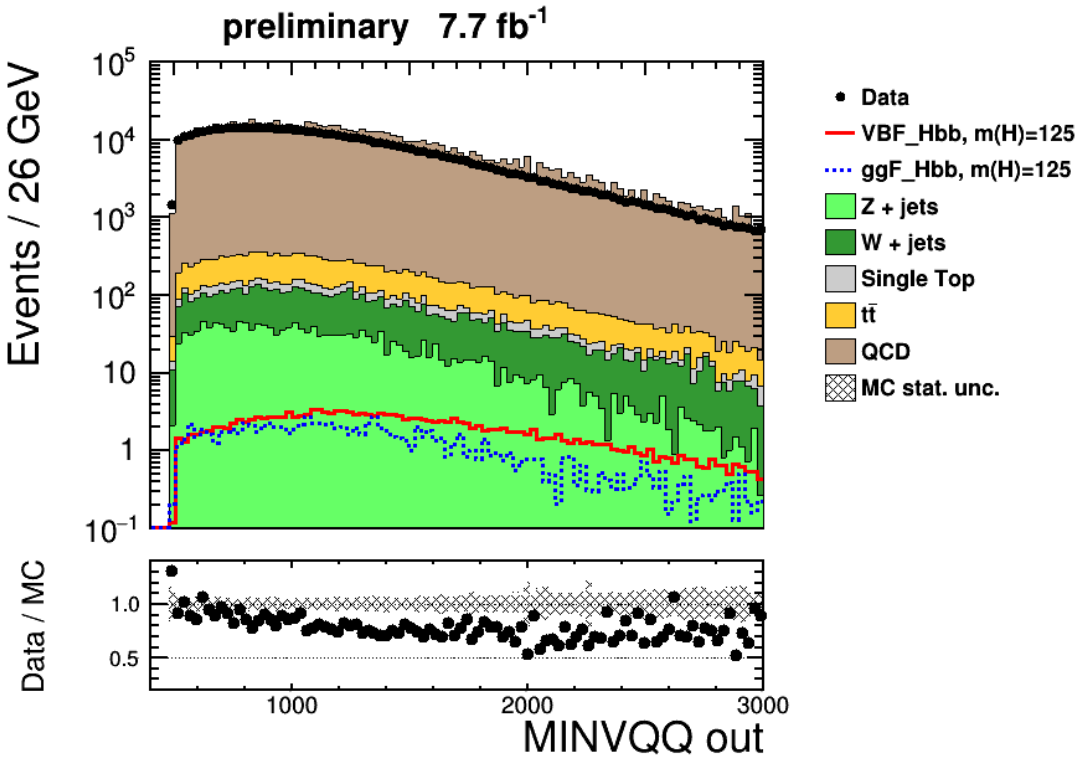




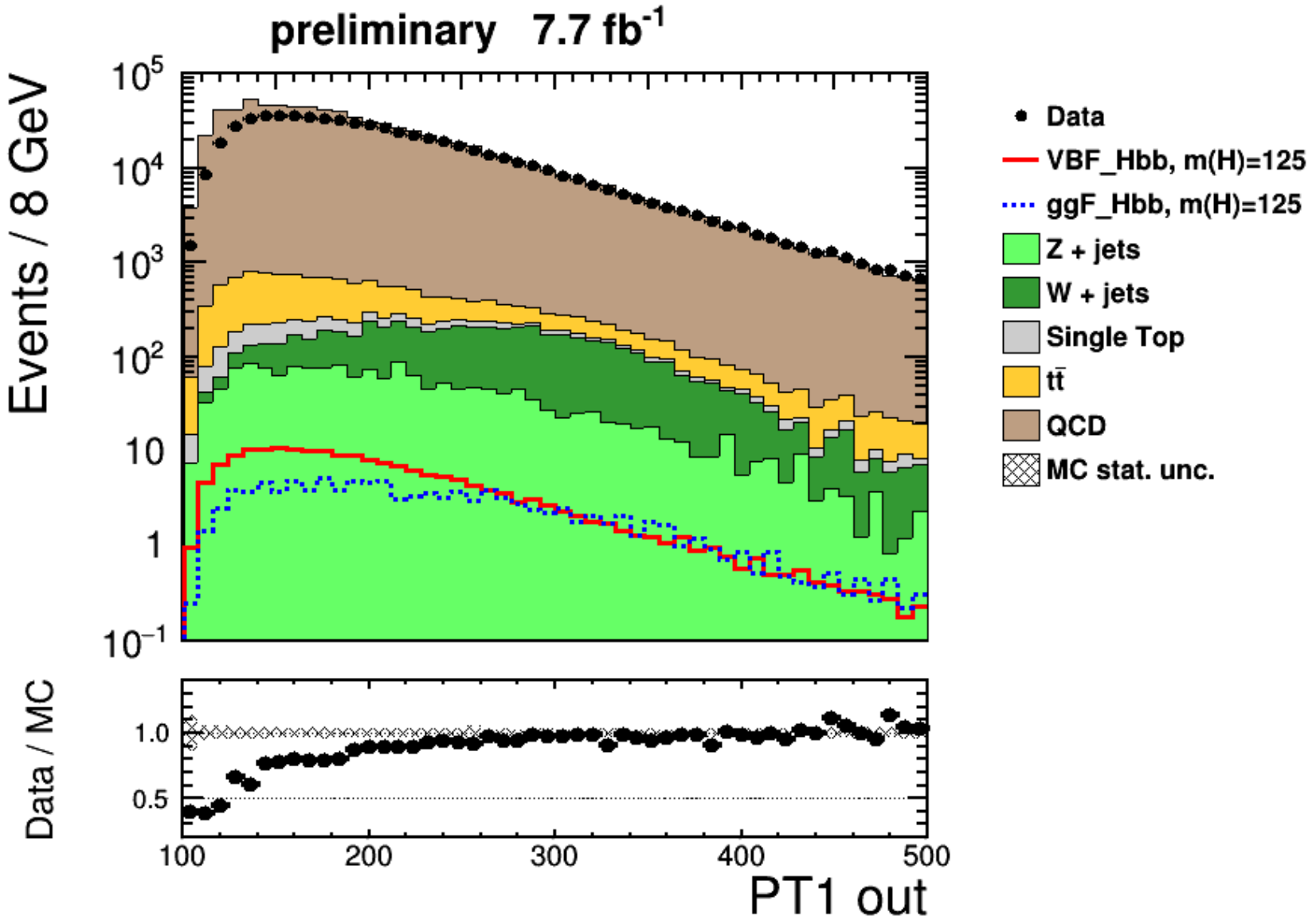
Minor dependence of *b*-jets invariant mass (mean) from BDT-response is observed !!!

Data vs MC

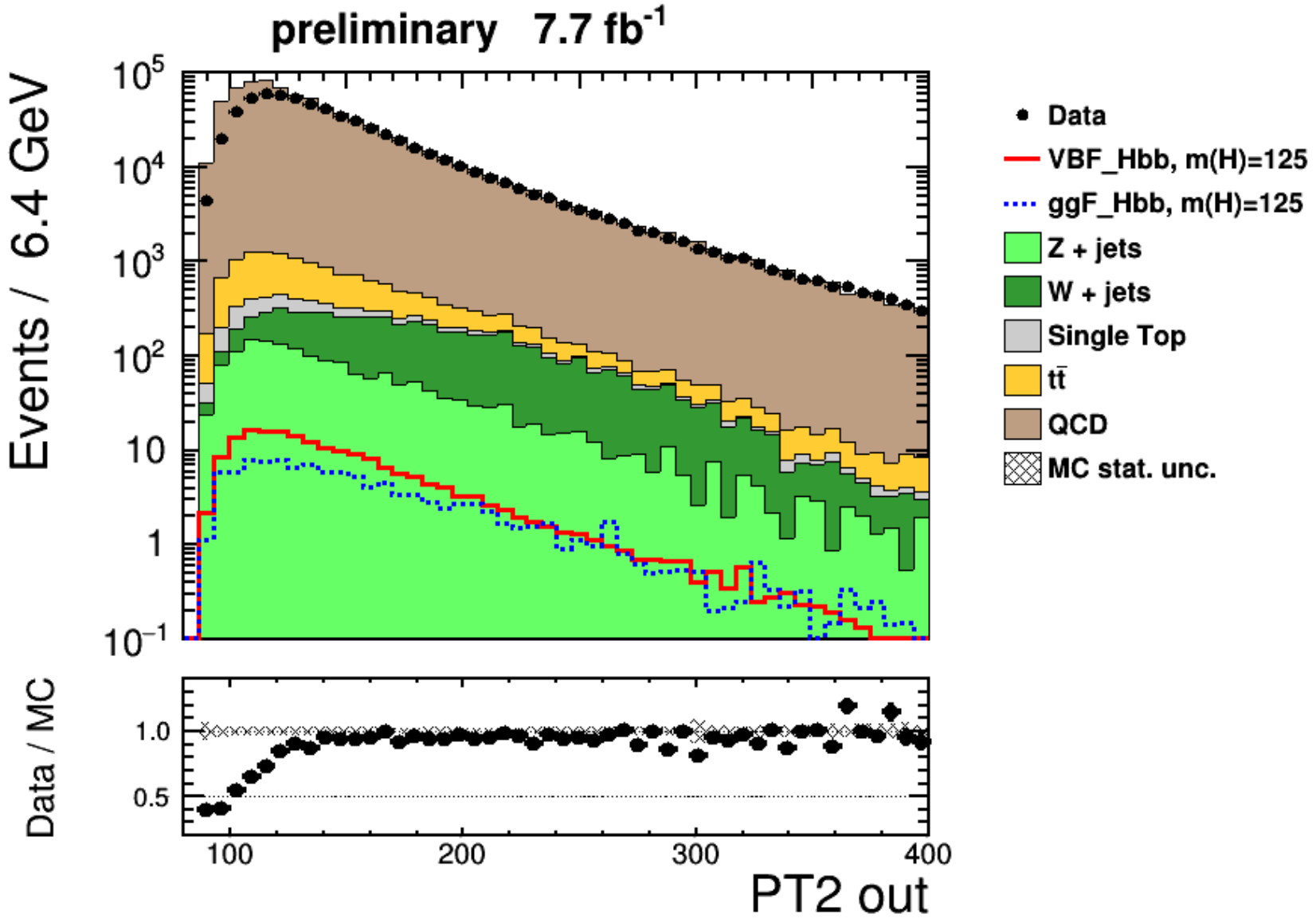

preliminary 7.7 fb-1
Events / 0.06
Data / MC
DETAQQ out
Data
VBF_Hbb, m(H)=125
ggF_Hbb, m(H)=125
Z + jets
W + jets
Single Top
tt̄
QCD
MC stat. unc.


preliminary 7.7 fb-1
Events / 26 GeV
Data / MC
MINVQQ out
Data
VBF_Hbb, m(H)=125
ggF_Hbb, m(H)=125
Z + jets
W + jets
Single Top
tt̄
QCD
MC stat. unc.

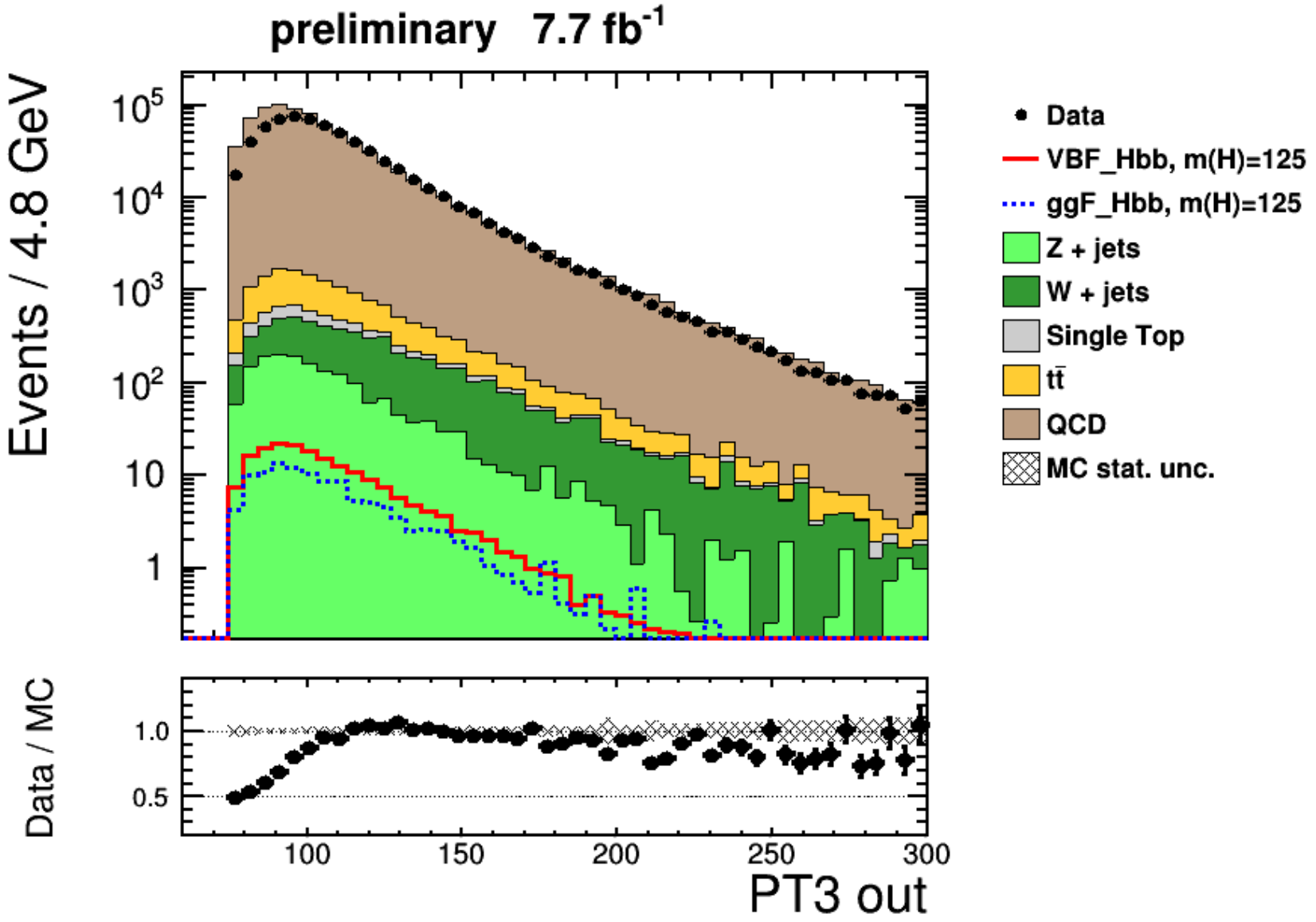
Data vs MC (pt leading jet)



Data vs MC (pt 1st subleading jet)



Data vs MC (pt 2nd subleading jet)



Upcoming works

- Derivation of corrections for MC samples: trigger efficiency/JEC/JER/BTG

Statistical analysis

- Background model validation 2 (shape analysis)
- Estimation of statistical/systematic errors
- Estimation of deviation from pure-background model

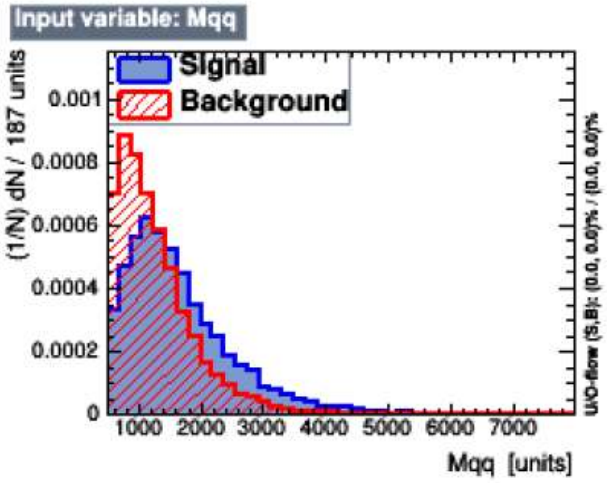
Optimization for 2018 analysis

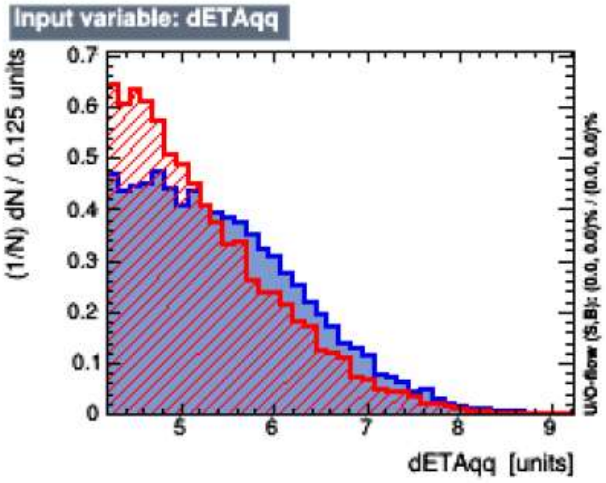
Combine 2017-2018 results

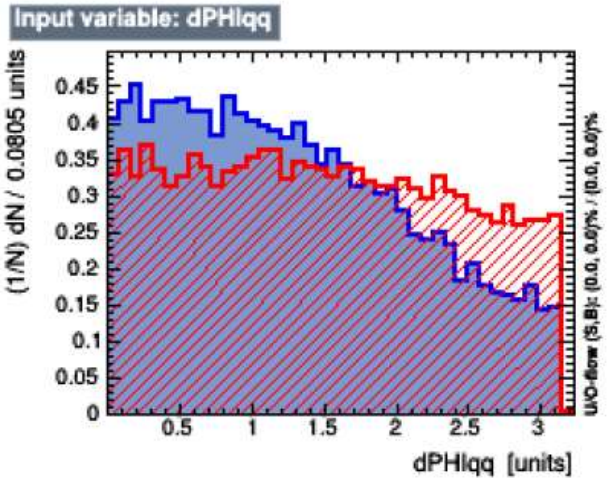
Thank you

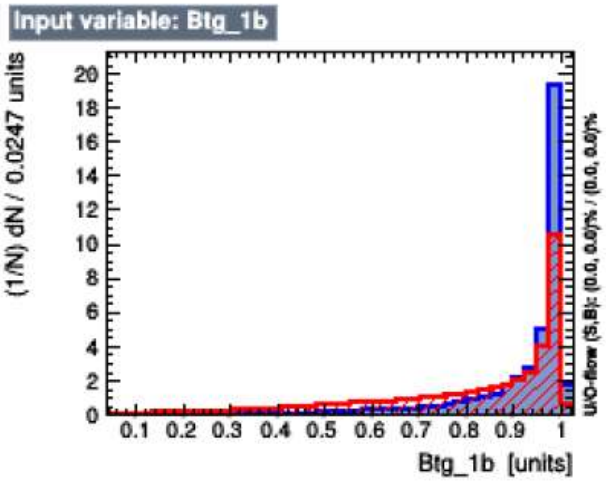
backup

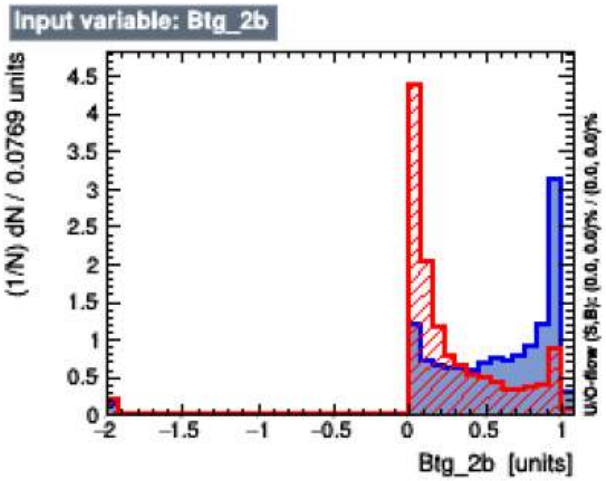
Signal vs Background variables-1 (SingleBTag)

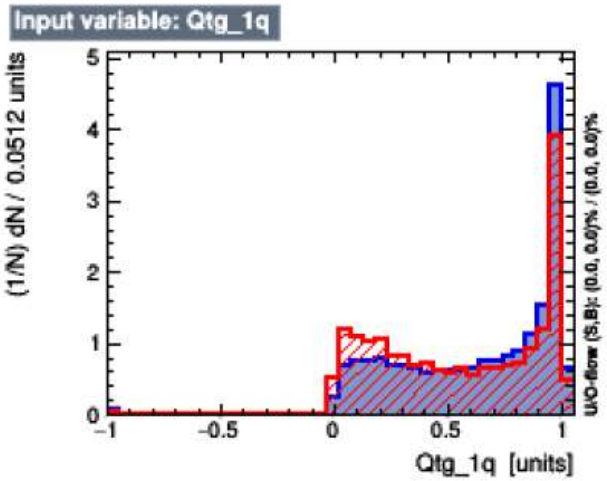




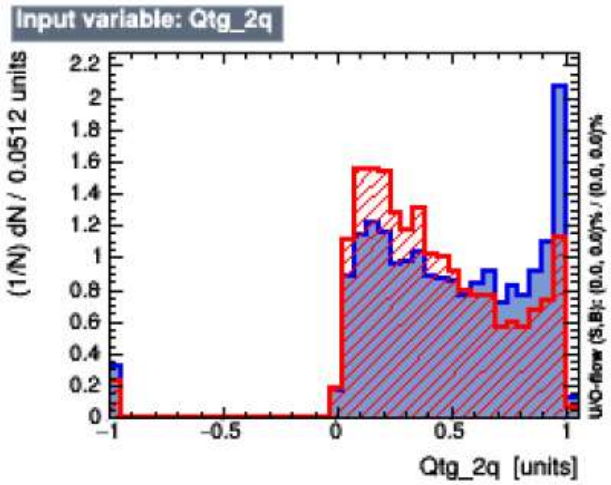


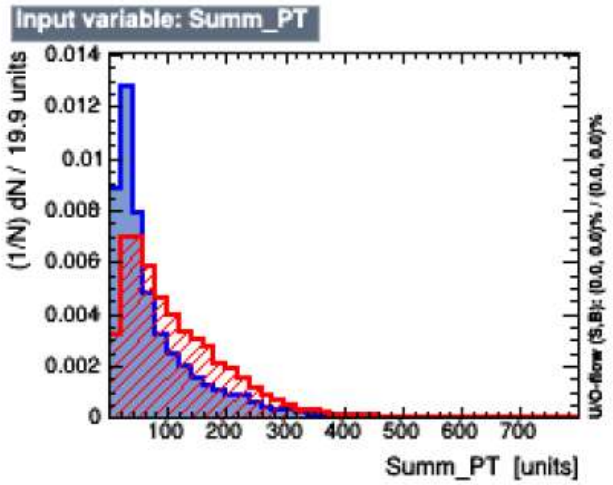


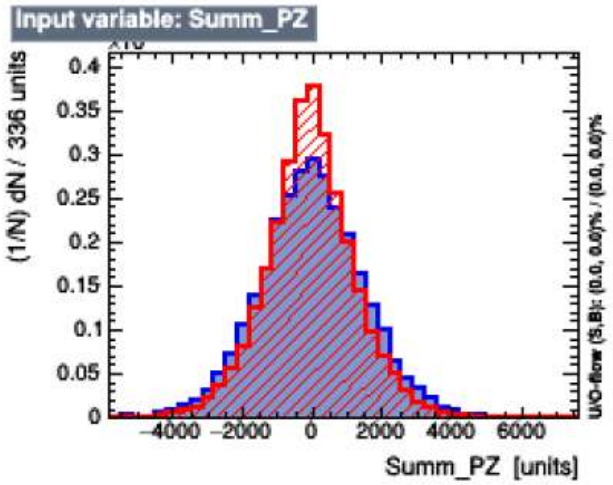


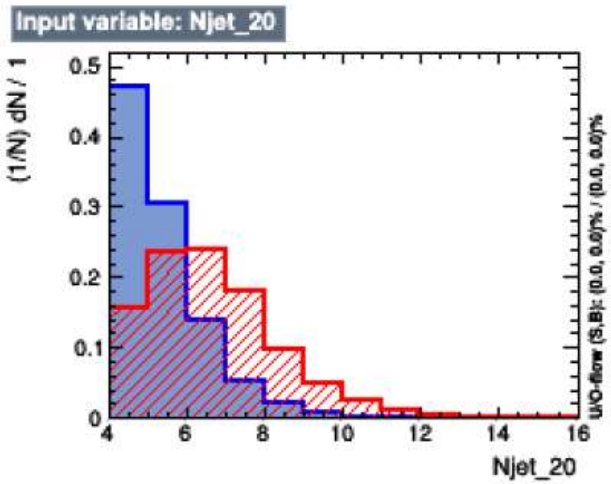


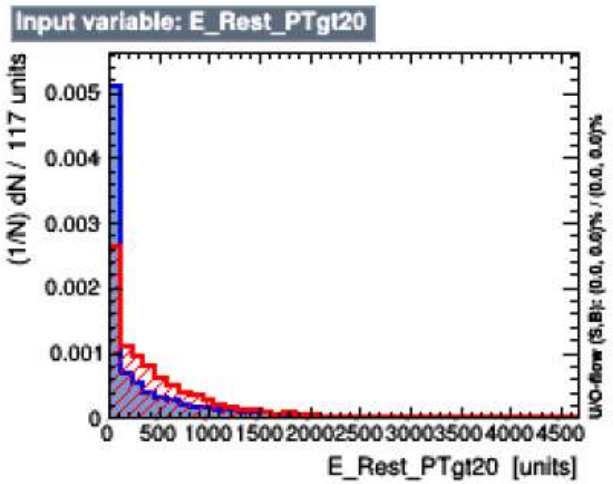
Signal vs Background variables-2 (SingleBTag)





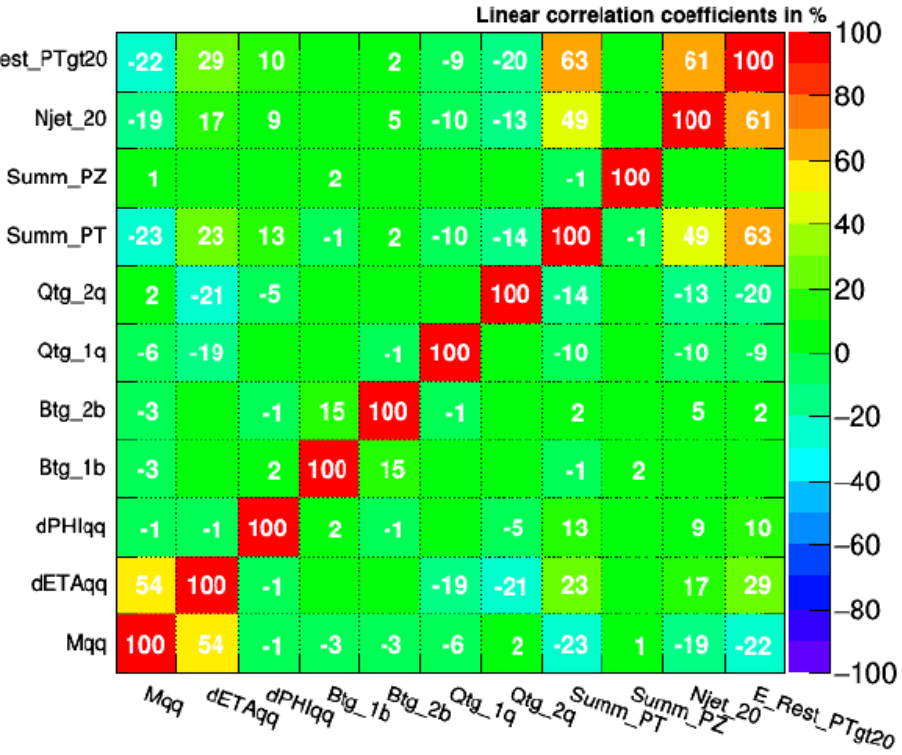




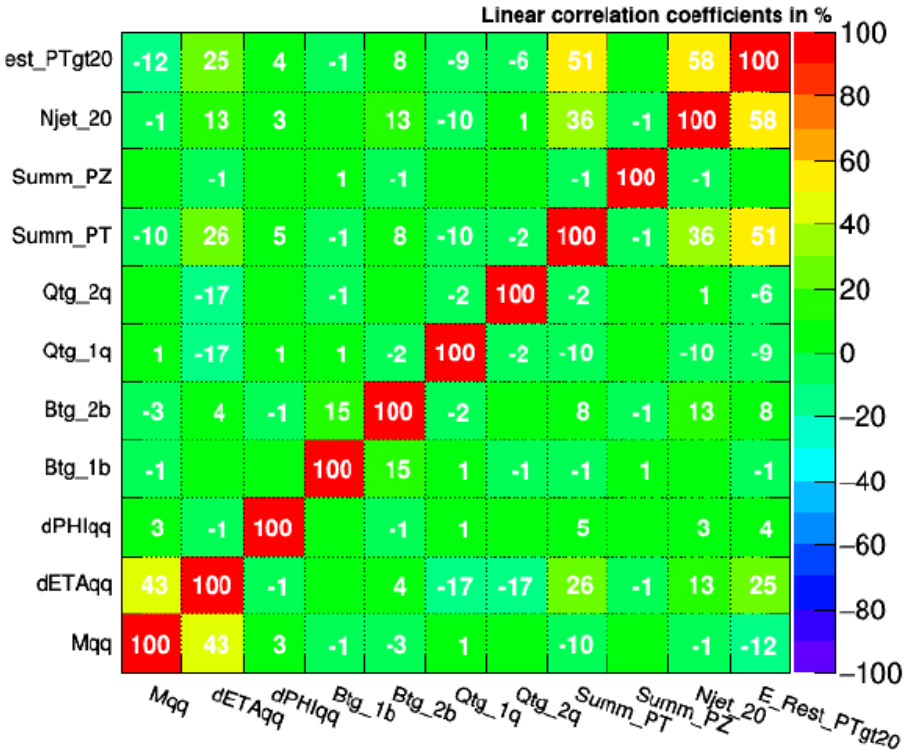


Signal vs Background variables correlation (SingleBTag)

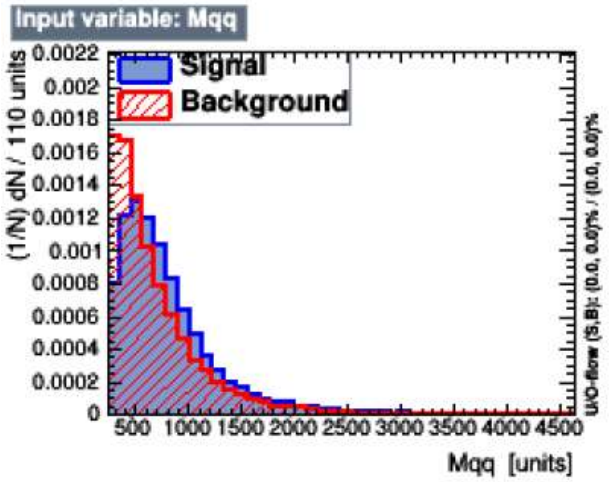
Correlation Matrix (signal)

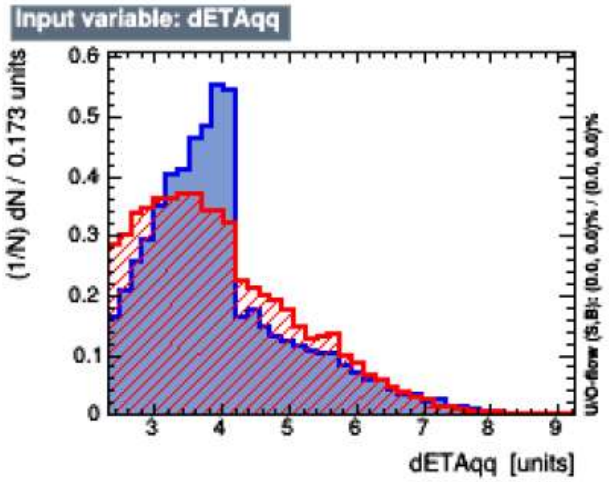


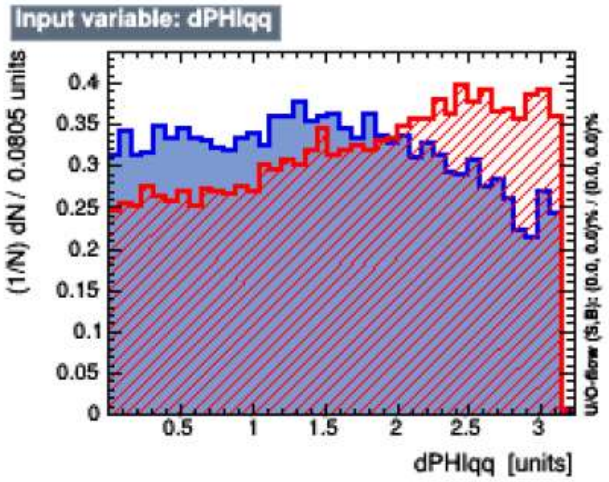
Correlation Matrix (background)

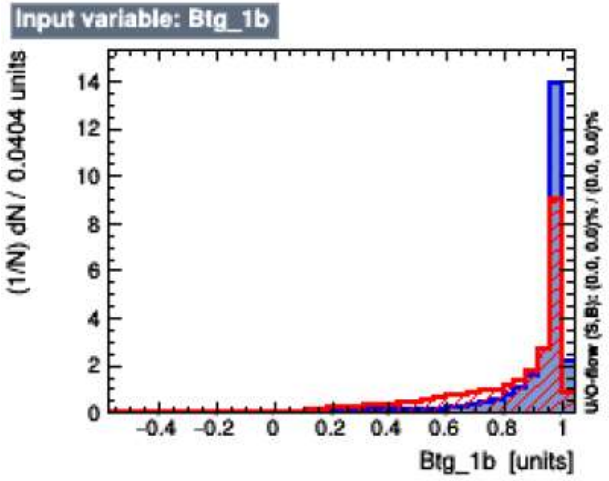


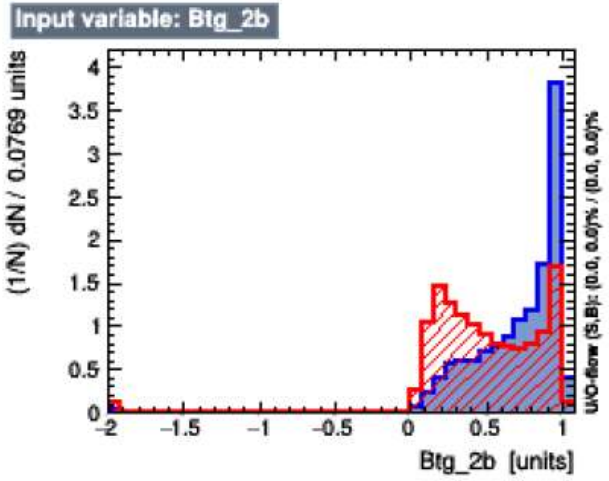
Signal vs Background variables-1 (BoubleBTag)

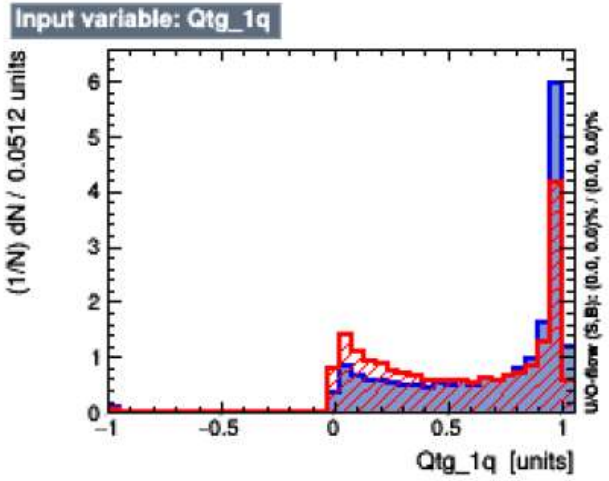




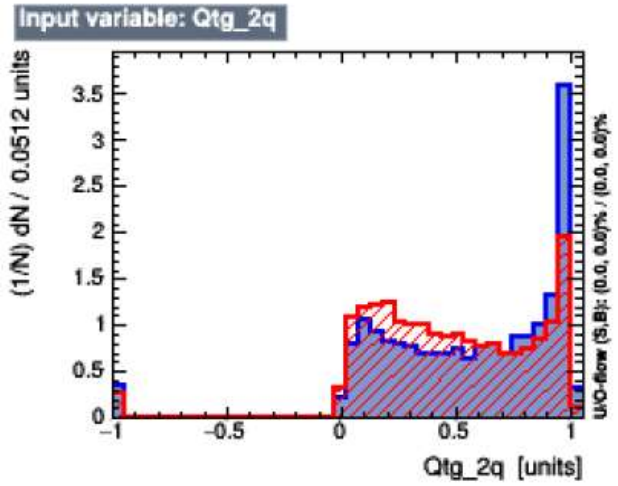


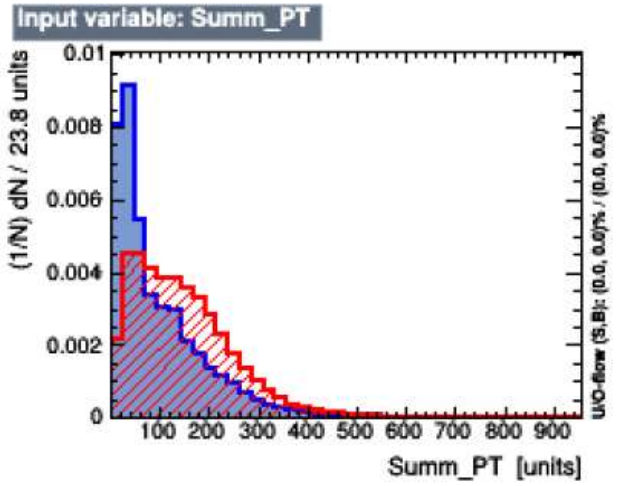


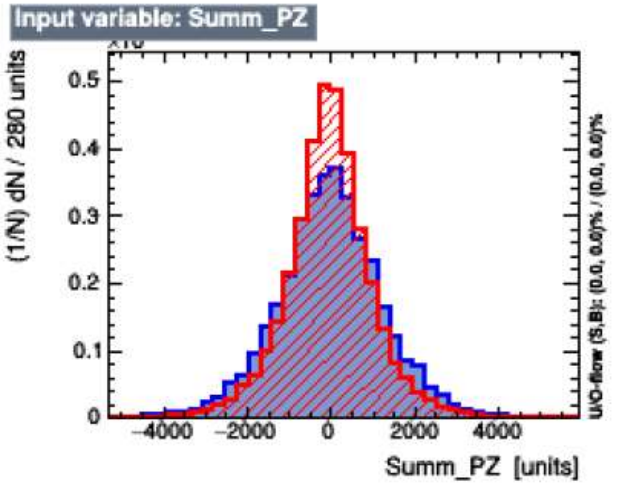


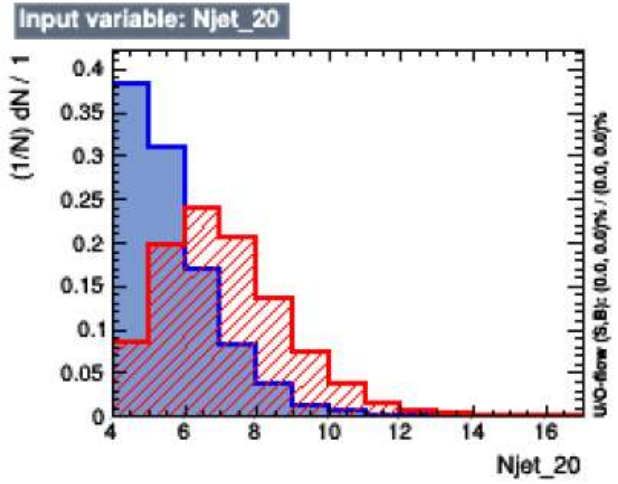


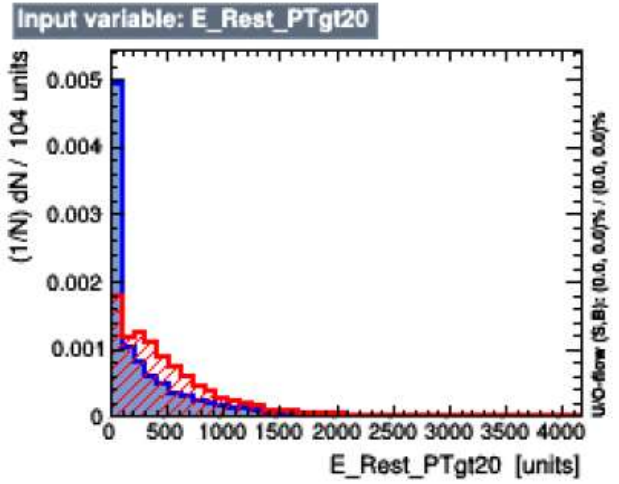
Signal vs Background variables-2 (BoubleBTag)





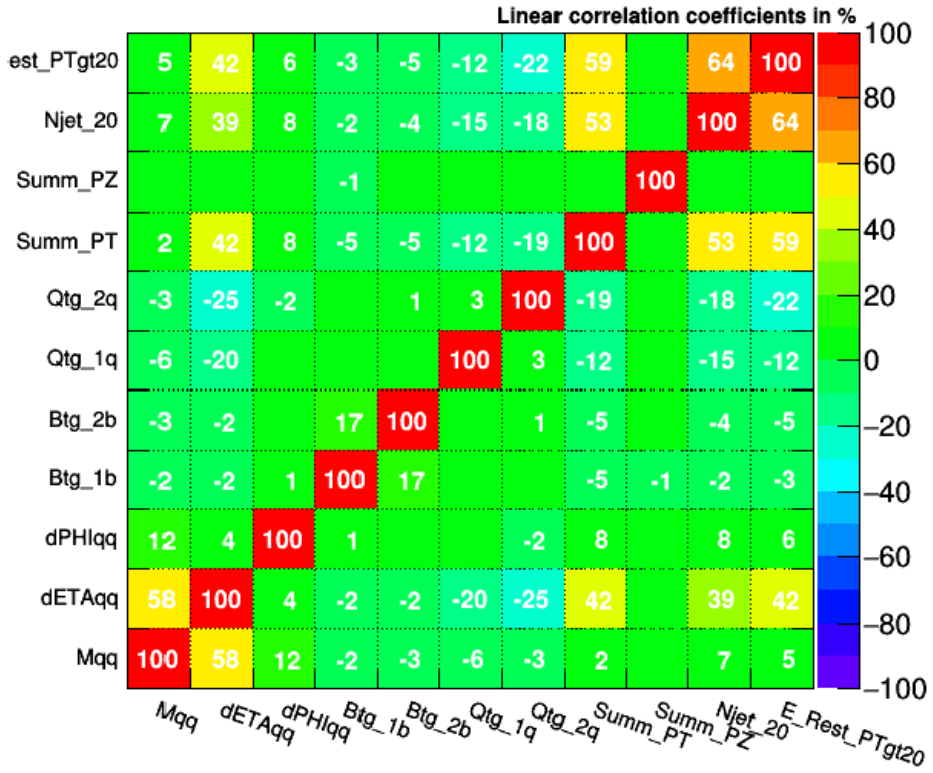




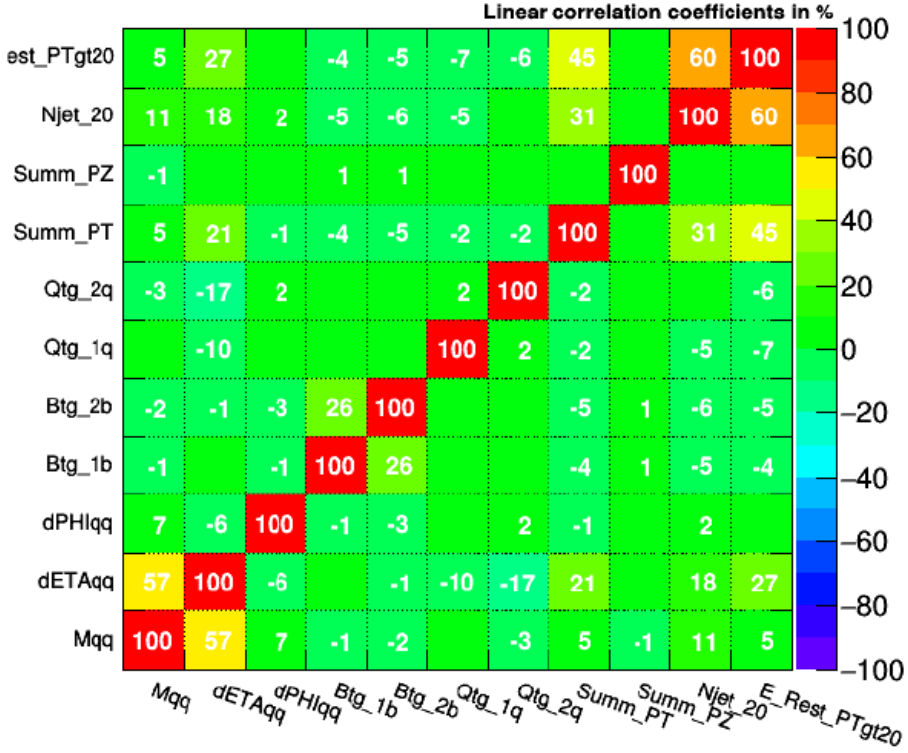


Signal vs Background variables correlation (DoubleBTag)

Correlation Matrix (signal)



Correlation Matrix (background)



Data vs MC (MinvBB)